Evaluation of XML Schema Quality in multimedia content publishing domain

Maja Pušnik, Boštjan Šumak

University of Maribor

Faculty of Electrical Engineering and Computer Science

maja.pusnik@um.si

Motivation 1: combining efforts from projects and research work

- Optimization of business processes (for the publishing domain)
- Integration of existing IT systems with new solutions (supported by XML technologies)
- Measuring quality of IT solutions (adjusted software metrics)

Motivation 2: integration in the learning process

XML schemas in the learning process

- Academic study program (1st year):
  - Basics of web technologies
    - **Basics** of XML and its connection to HTML
- Academic study program (3rd year):
  - System convergence and integration
    - **Specifics** of XML and its **use in Java** web services and Java applications
- Professional study program (3rd year)
  - Development of information services
    - **Specifics** of XML and its **use in C#** web services ands C# applications

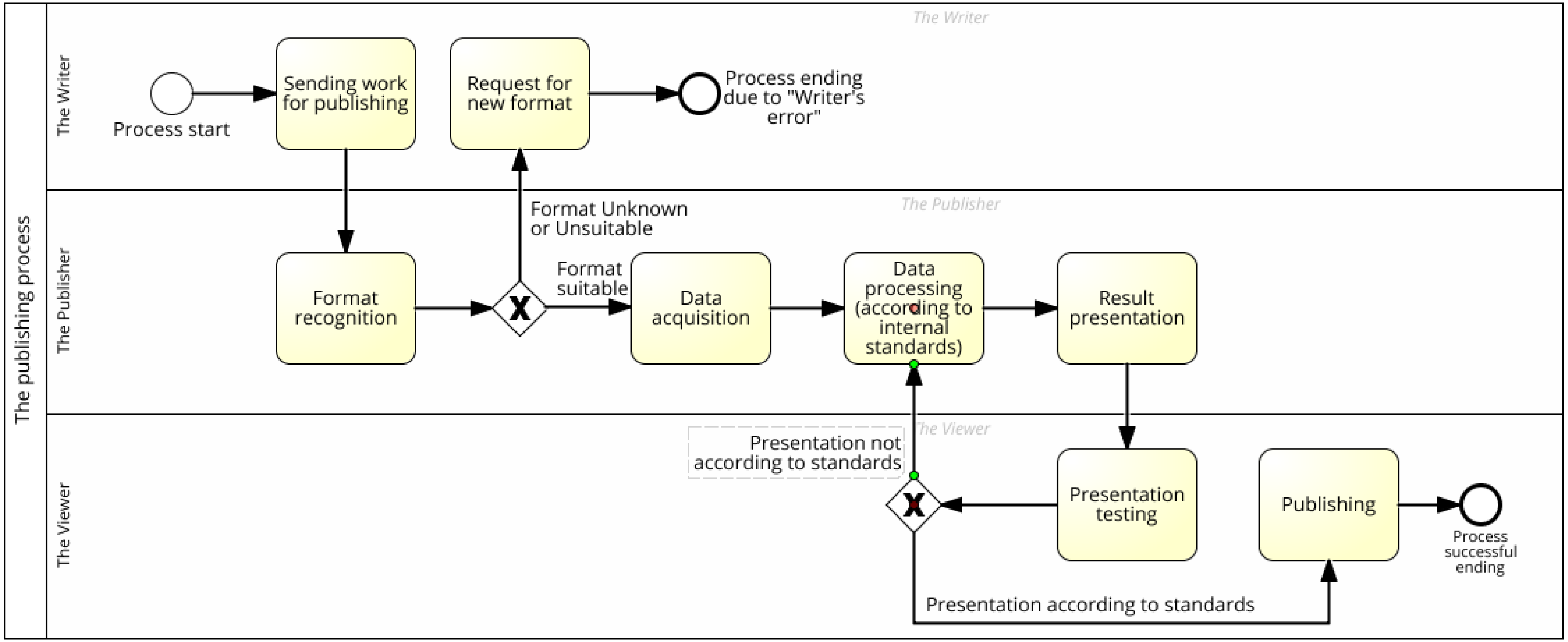
Optimization and quality management

- Second Cycle Bologna Study Programmes (1st year):
  - **Business process optimization**
    - Business process modeling, simulation, optimization and reporting
- Second Cycle Bologna Study Programmes (2nd year):
  - **Operational research**
    - Use of operational research methods for optimization of business process and IT solution optimization

Agenda

- The multimedia content publishing domain
- The role of XML technologies
- XML Schemas and the metric system Quality index
- Evaluation of the publishing domain

Publishing process


The publishing process
The Writer
The Writer
Process start
Sending work for publishing
Request for new format
Process ending due to "Writer's error"
The Publisher
The Publisher
Format Unknown or Unsuitable
Format suitable
Format recognition
X
Data acquisition
Data processing (according to internal standards)
Result presentation
The Viewer
The Viewer
Presentation not according to standards
X
Presentation testing
Publishing
Process successful ending
Presentation according to standards

The problems of the publishing process

Publishing organizations must provide the same content in various formats in order to meet the needs of their clients

- large amounts of data
- poor organization (of knowledge databases)
- Increasing involvement of multimedia contents

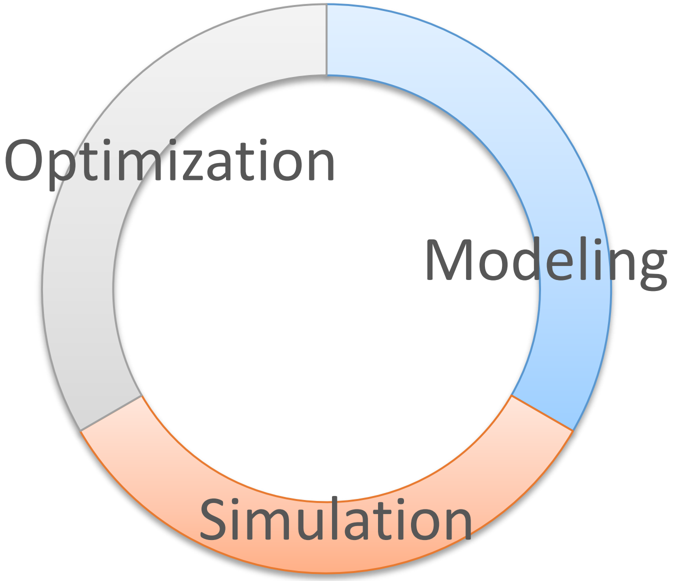
The publishing process is performed in both printed and electronic form, however there is an increasing number eBooks (Shaffer, 2012).

- Little IT support
- Almost no automation

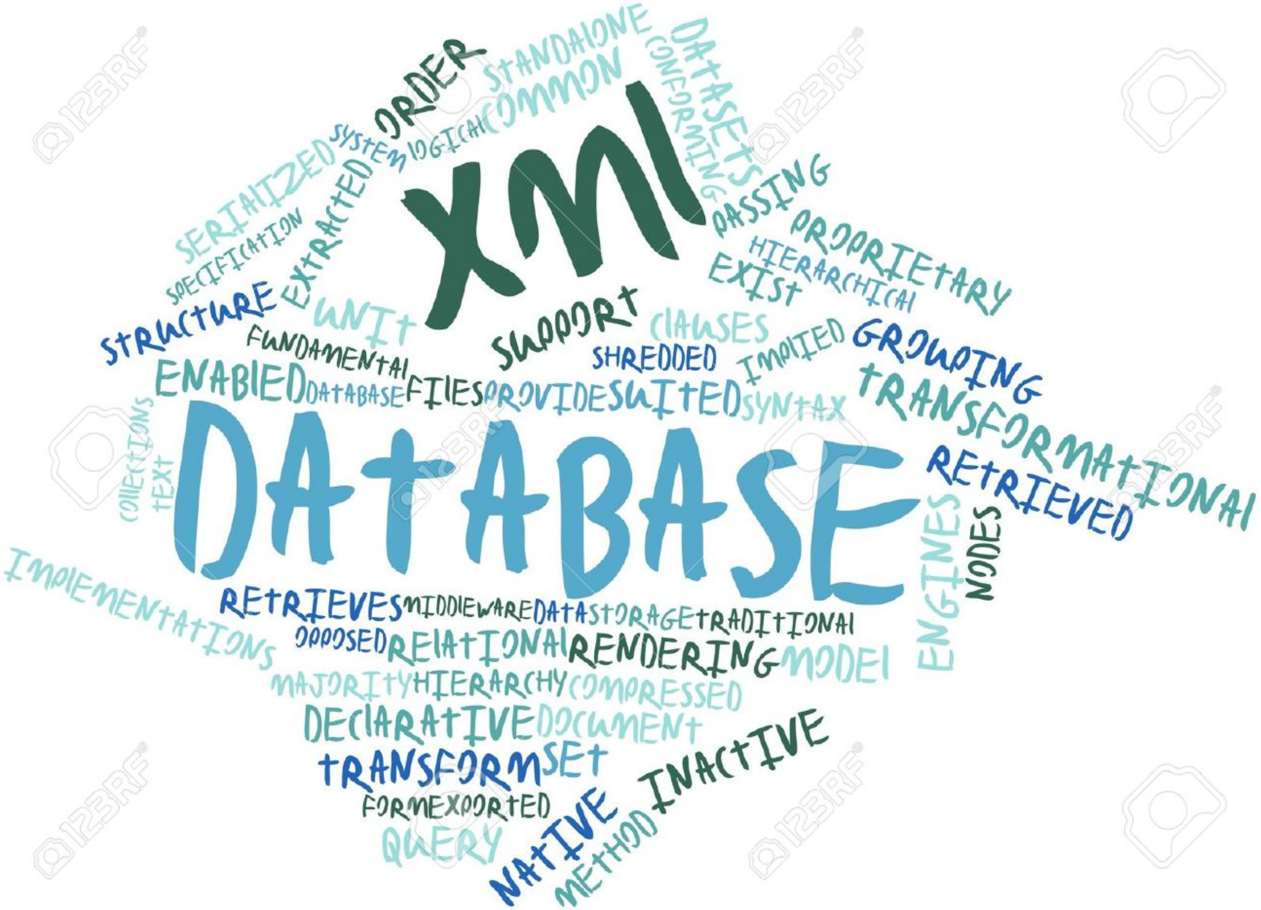
Multimedia content publishing process quality

The publishing process can be optimized, simplified and become more efficient with XML technologies support regardless of the document origin and content

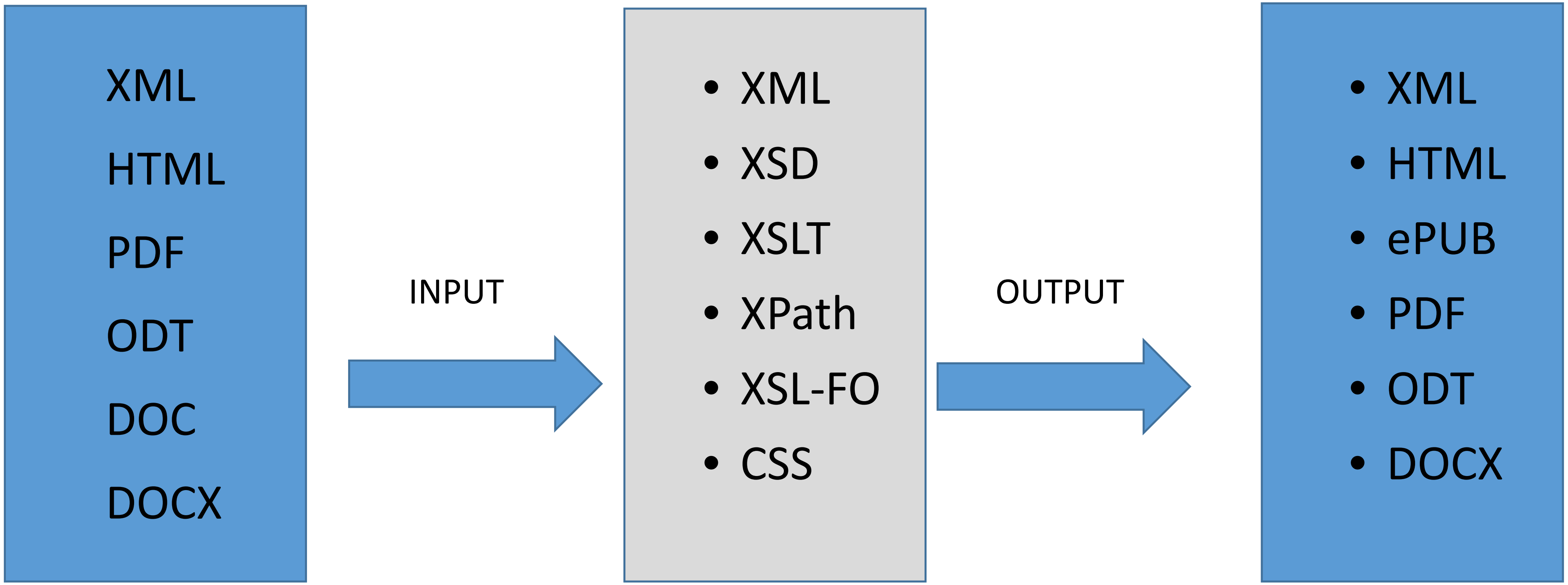
„PUŠNIK, Maja. Using XML technologies for various data format transformations : lecture presented at Workshop in Bohinj (project Software engineering - Computer science education and research cooperation), August 24-29, 2015. 2015“



The role of XML technologies



XML family of technologies connected



XML

HTML

PDF

ODT

DOC

DOCX

INPUT

- XML
- XSD
- XSLT
- XPath
- XSL-FO
- CSS

OUTPUT

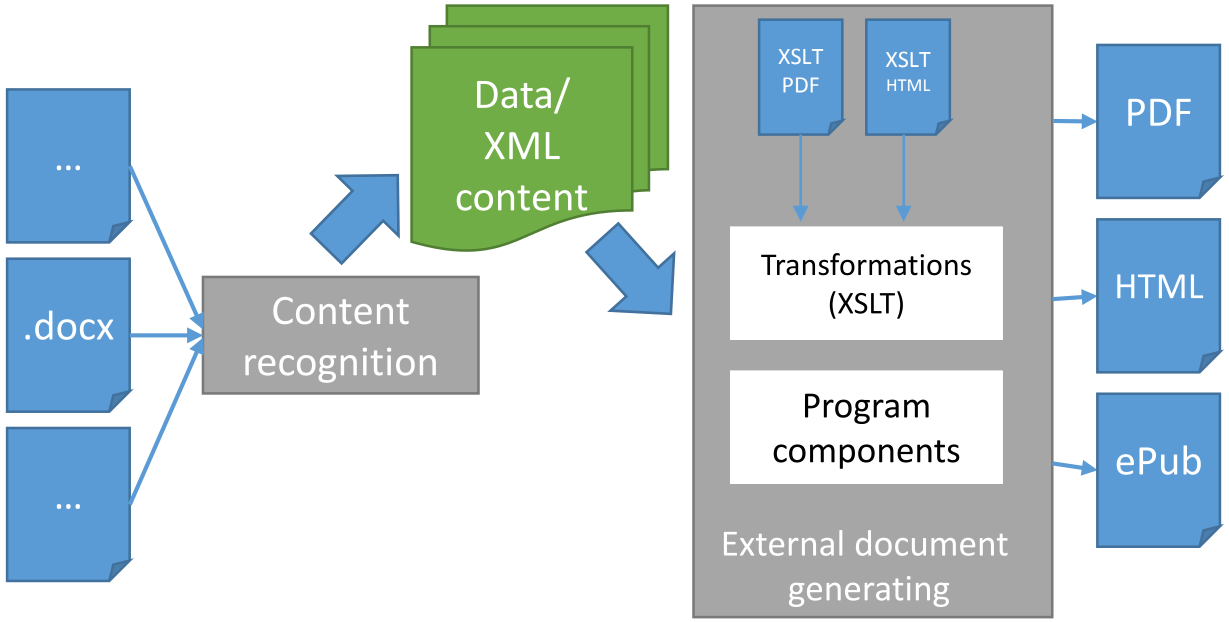
- XML
- HTML
- ePUB
- PDF
- ODT
- DOCX

Why use XML technologies

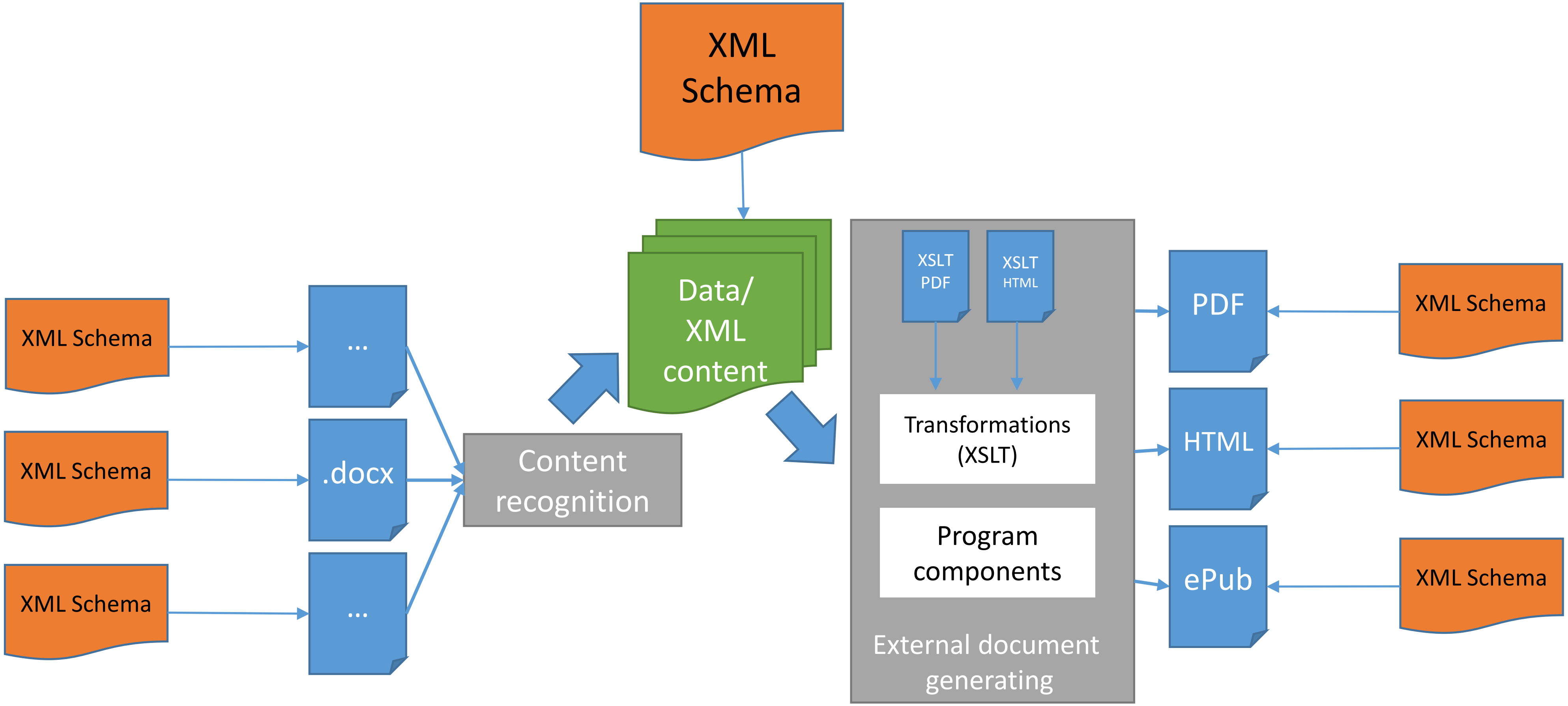
- Enabling a (semi) process automation
- Addressing different challenges:
  - technical
  - organizational
  - financial

- Finding balance between manual and automatic steps

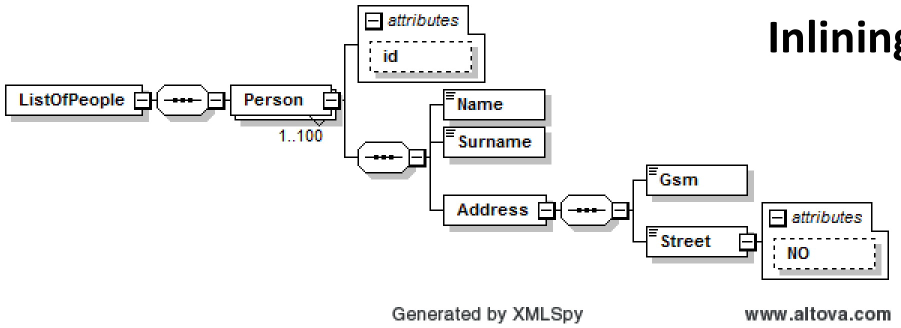
XML technologies in the publishing process (architecture design)

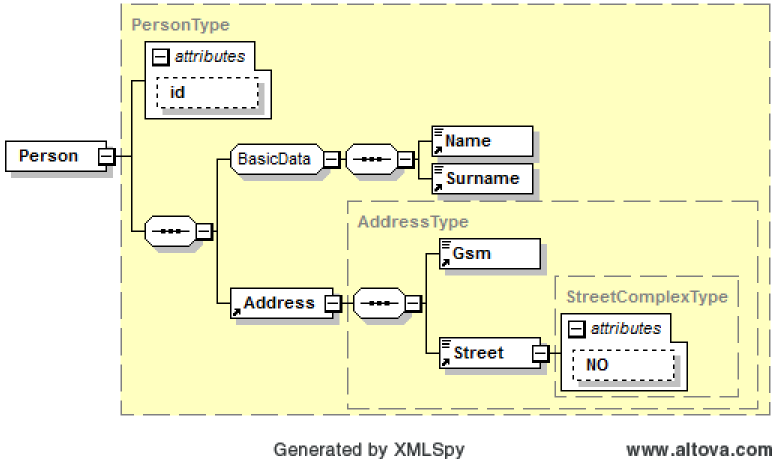


XML schemas in the publishing process



Inlining

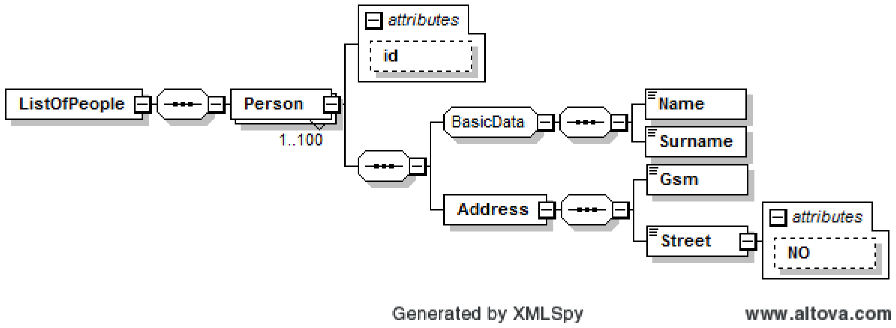




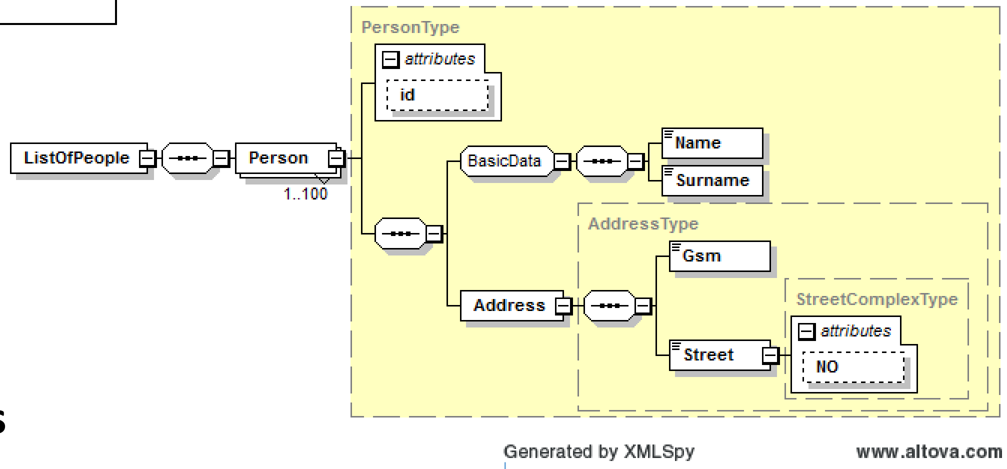
Structured data, within different formats

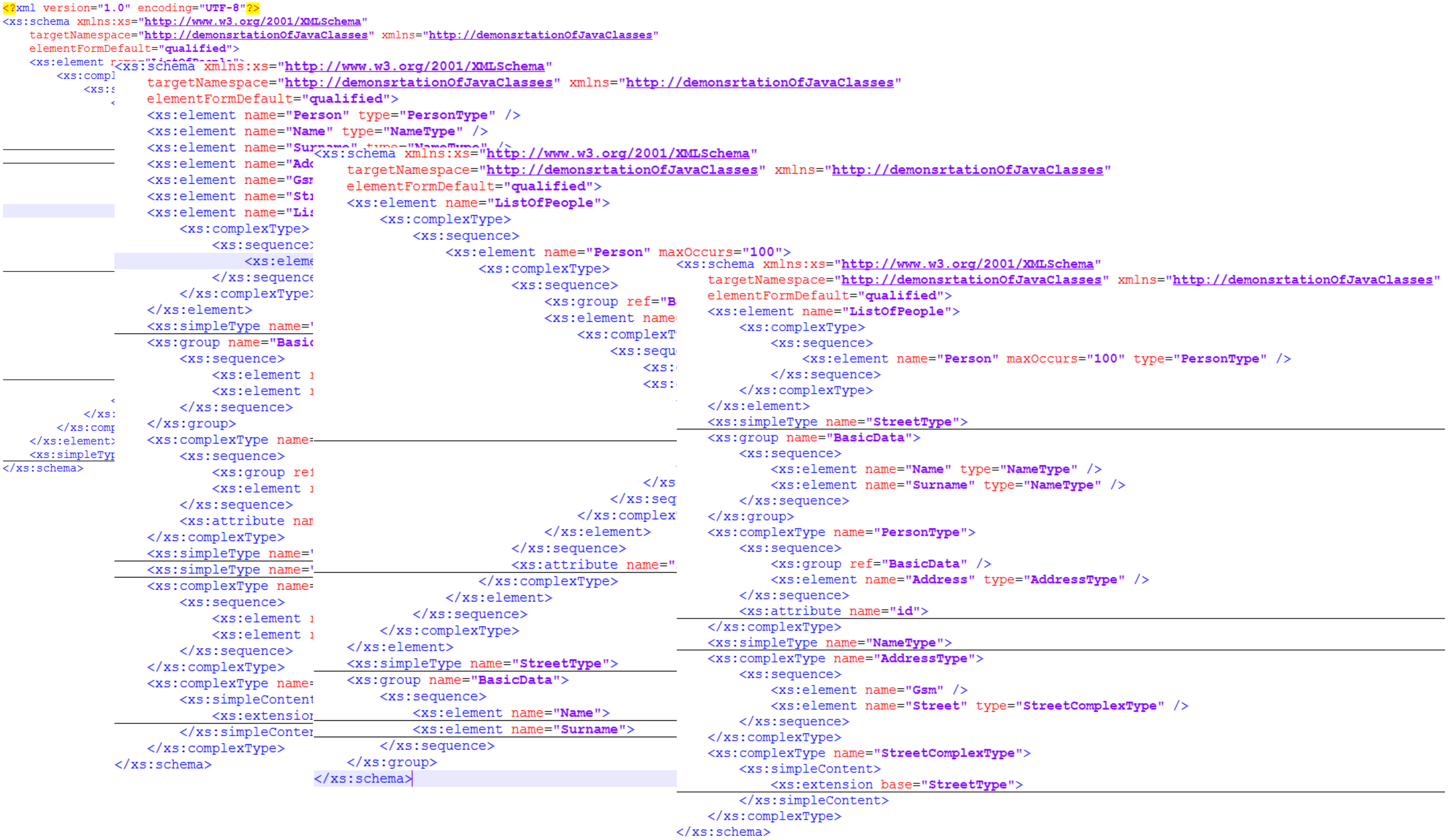
```
<ListOfPeople>
   <Person id="idValue">
      <Name>nameValue</Name>
      <Surname>surnameValue</Surname>
      <Address>
         <Gsm>gsmValue</Gsm>
         <Street NO="noValue">streetValue</Street>
      </Address>
   </Person>
</ListOfPeople>
```

Global data types



Global elements





```
<?xml version="1.0" encoding="UTF-8"?>
<xs:schema xmlns:xs="http://www.w3.org/2001/XMLSchema"
    targetNamespace="http://demonsrtationOfJavaClasses" xmlns="http://demonsrtationOfJavaClasses"
    elementFormDefault="qualified">
    <xs:element name="ListOfPeople">
        <xs:compl
            <xs:s
                <
                <
            </xs:
        </xs:comp
    </xs:element>
    <xs:simpleTyp
</xs:schema>
```

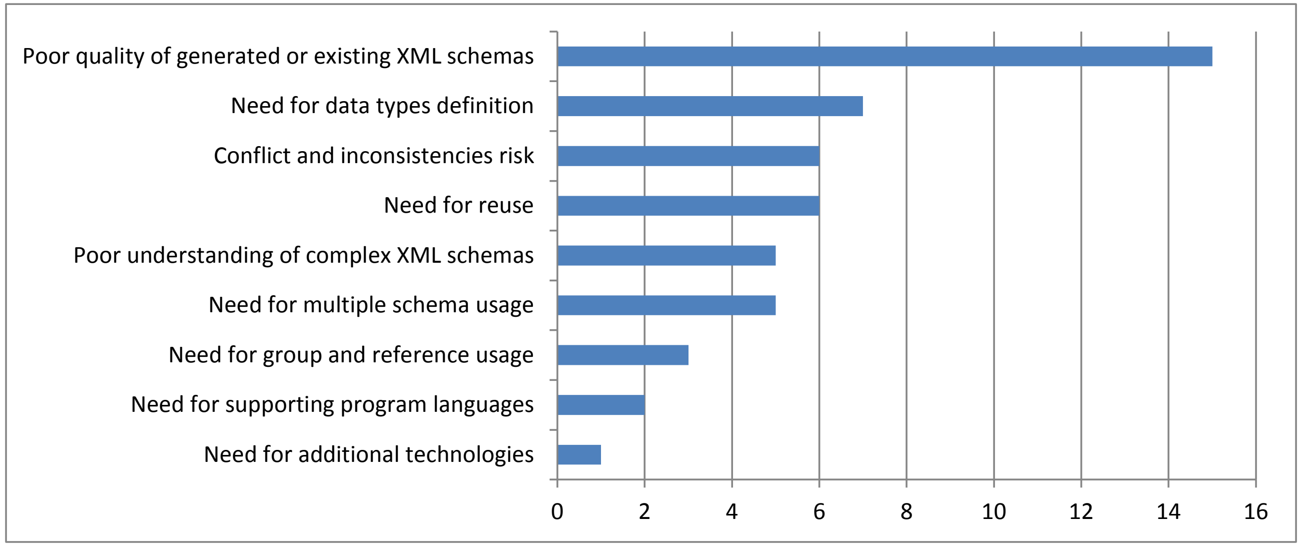
```
<xs:schema xmlns:xs="http://www.w3.org/2001/XMLSchema"
    targetNamespace="http://demonsrtationOfJavaClasses" xmlns="http://demonsrtationOfJavaClasses"
    elementFormDefault="qualified">
    <xs:element name="Person" type="PersonType" />
    <xs:element name="Name" type="NameType" />
    <xs:element name="Surname" type="NameType" />
    <xs:element name="Add
    <xs:element name="Gsm
    <xs:element name="Str
    <xs:element name="Lis
        <xs:complexType>
            <xs:sequence>
                <xs:eleme
            </xs:sequence
        </xs:complexType
    </xs:element>
    <xs:simpleType name='
    <xs:group name="Basic
        <xs:sequence>
            <xs:element n
            <xs:element n
        </xs:sequence>
    </xs:group>
    <xs:complexType name=
        <xs:sequence>
            <xs:group ref
            <xs:element n
        </xs:sequence>
        <xs:attribute nam
    </xs:complexType>
    <xs:simpleType name='
    <xs:simpleType name='
    <xs:complexType name=
        <xs:sequence>
            <xs:element n
            <xs:element n
        </xs:sequence>
    </xs:complexType>
    <xs:complexType name=
        <xs:simpleContent
            <xs:extensio
        </xs:simpleConte
    </xs:complexType>
</xs:schema>
```

```
<xs:schema xmlns:xs="http://www.w3.org/2001/XMLSchema"
    targetNamespace="http://demonsrtationOfJavaClasses" xmlns="http://demonsrtationOfJavaClasses"
    elementFormDefault="qualified">
    <xs:element name="ListOfPeople">
        <xs:complexType>
            <xs:sequence>
                <xs:element name="Person" maxOccurs="100">
                    <xs:complexType>
                        <xs:sequence>
                            <xs:group ref="B
                            <xs:element name
                                <xs:complexT
                                    <xs:sequ
                                        <xs:
                                        <xs:
                                        </xs
                                    </xs:seq
                                </xs:complex
                            </xs:element>
                        </xs:sequence>
                        <xs:attribute name="
                    </xs:complexType>
                </xs:element>
            </xs:sequence>
        </xs:complexType>
    </xs:element>
    <xs:simpleType name="StreetType">
    <xs:group name="BasicData">
        <xs:sequence>
            <xs:element name="Name">
            <xs:element name="Surname">
        </xs:sequence>
    </xs:group>
</xs:schema>
```

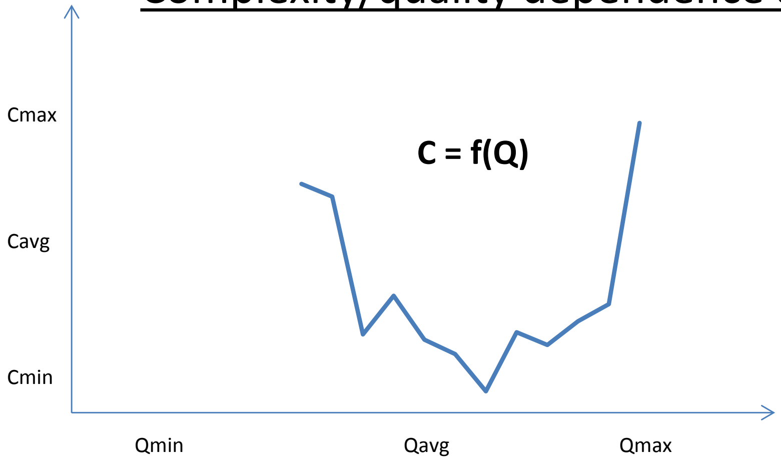
```
<xs:schema xmlns:xs="http://www.w3.org/2001/XMLSchema"
    targetNamespace="http://demonsrtationOfJavaClasses" xmlns="http://demonsrtationOfJavaClasses"
    elementFormDefault="qualified">
    <xs:element name="ListOfPeople">
        <xs:complexType>
            <xs:sequence>
                <xs:element name="Person" maxOccurs="100" type="PersonType" />
            </xs:sequence>
        </xs:complexType>
    </xs:element>
    <xs:simpleType name="StreetType">
    <xs:group name="BasicData">
        <xs:sequence>
            <xs:element name="Name" type="NameType" />
            <xs:element name="Surname" type="NameType" />
        </xs:sequence>
    </xs:group>
    <xs:complexType name="PersonType">
        <xs:sequence>
            <xs:group ref="BasicData" />
            <xs:element name="Address" type="AddressType" />
        </xs:sequence>
        <xs:attribute name="id">
    </xs:complexType>
    <xs:simpleType name="NameType">
    <xs:complexType name="AddressType">
        <xs:sequence>
            <xs:element name="Gsm" />
            <xs:element name="Street" type="StreetComplexType" />
        </xs:sequence>
    </xs:complexType>
    <xs:complexType name="StreetComplexType">
        <xs:simpleContent>
            <xs:extension base="StreetType">
        </xs:simpleContent>
    </xs:complexType>
</xs:schema>
```

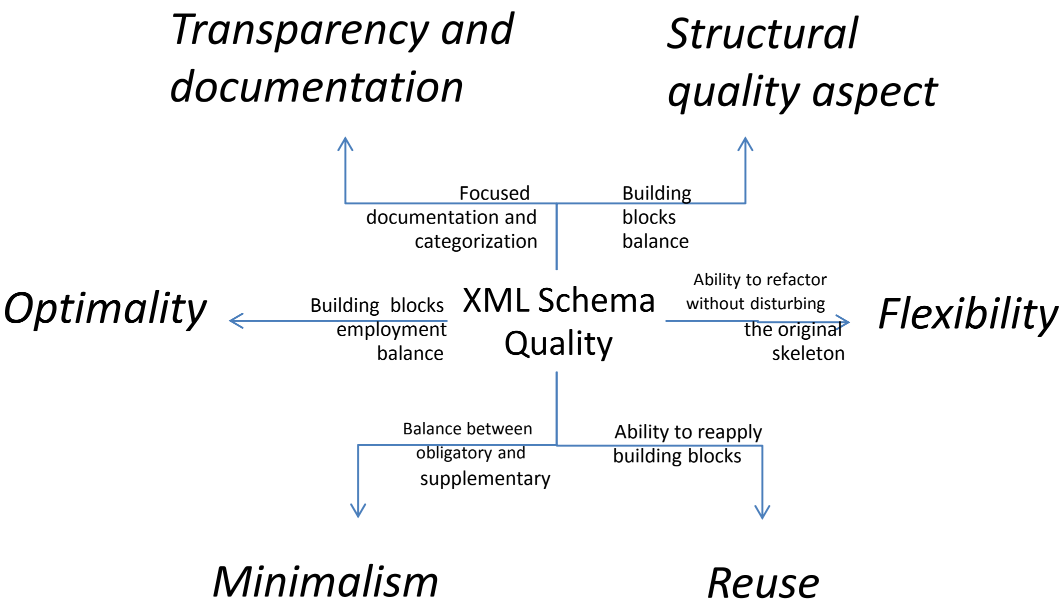
Quality aspects analysis

Survey among developers



Complexity/quality dependence analysis





XML Schema Quality

Transparency and documentation

Focused documentation and categorization

Structural quality aspect

Building blocks balance

Optimality

Building blocks employment balance

Flexibility

Ability to refactor without disturbing the original skeleton

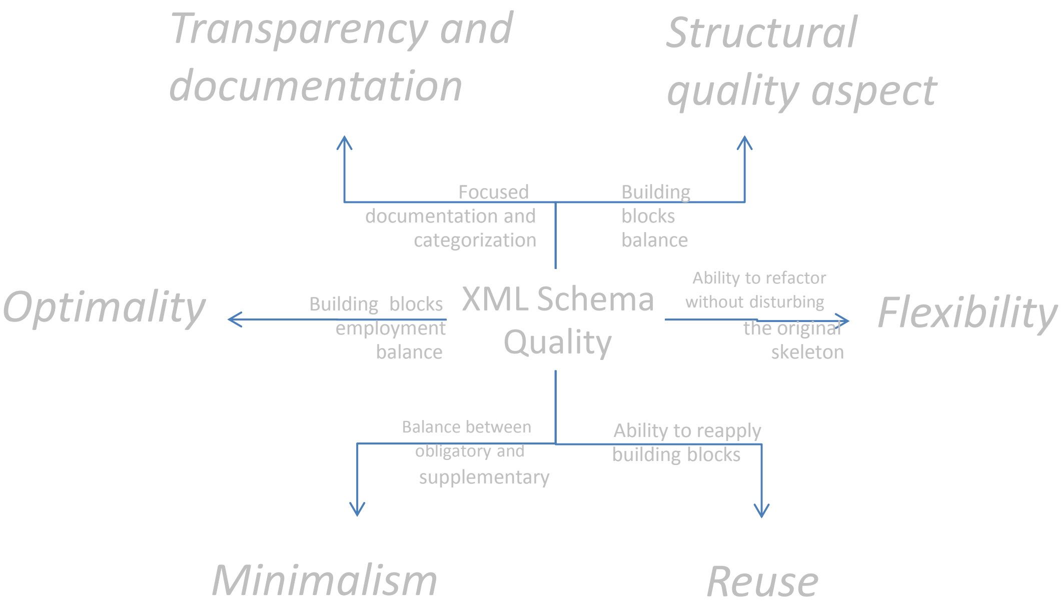
Minimalism

Balance between obligatory and supplementary

Reuse

Ability to reapply building blocks

XML Schemas Quality index parameters



Transparency and documentation

Structural quality aspect

Focused documentation and categorization

Building blocks balance

Optimality

Building blocks employment balance

XML Schema Quality

Ability to refactor without disturbing the original skeleton

Flexibility

Balance between obligatory and supplementary

Ability to reapply building blocks

Minimalism

Reuse

	Aspect 1	Aspect 2	Aspect 3	Aspect 4	Aspect 5	Aspect 6
N_{an} = number of annotations	X	X		X		
N_{ri_all} = number of external XML schemas	X				X	X
N_E = number of elements	X	X	X	X	X	X
N_{E_g} = number of global elements		X				
N_{E_l} = number of local elements			X			
N_{E_s} = number of simple elements			X			
N_{E_gc} = number of global complex elements			X			
N_{E_gs} = number of global simple elements			X			
N_{at} = number of all attributes	X	X	X	X	X	X
N_{at_l} = number of local attributes			X			
LOC = lines of code				X		
N_g = number of all groups		X			X	X
N_{E_group} = number of element groups		X				X
N_{A_group} = number of attribute groups		X				X
N_{re_all} = number of references on elements					X	X
N_{ra_all} = number of attribute references					X	X
N_{rg_all} = number of group references					X	X
N_r = number of restrictions	X					
N_{t_i} = number of derived data types	X				X	
N_t = number of all data types			X	X	X	X
N_{rt_all}= number of all used data types				X	X	X
N_{t_s} = number of simple data types	X					
N_{t_c} = number of complex data types	X					
N_{E_U} = number of unbounded elements	X		X			X

Structural quality aspect (QA1)

$$QA_1 = N_{ri\_all} + \frac{N_E}{N_{at}} + \frac{N_r}{N_{t\_s}} + \frac{N_{t\_s}}{N_{t\_c}} + \frac{N_{an} + N_{t\_i} + N_{E\_U}}{N_E}$$

Transparency and documentation of the XML Schema (QA2)

$$QA_2 = \frac{N_{an}}{N_E + N_{at}} + \frac{N_{E_{group}}}{N_E} + \frac{N_{A_{group}}}{N_{at}} + \frac{N_g}{N_E}$$

XML schema optimality quality aspect (QA3)

$$QA_3 = \frac{1}{7}(\frac{N_{E\_l}}{N_E} + \frac{N_{at\_l}}{N_{at}} + \left(1 - \frac{N_{E\_gc}}{N_E - N_{E\_s}}\right) + \frac{N_{E\_gs}}{N_{E\_s}} + \frac{N_t}{N_E + N_{at}} + \frac{N_g}{N_{E\_gc}} + \left(1 - \frac{N_{E\_U}}{N_E}\right)\Bigg)$$

XML schema minimalism quality aspect (QA4)

$$QA4 = \frac{N_{an} + N_E + N_{at}}{LOC} + \frac{N_{rt\_all}}{N_t}$$

XML schema reuse quality aspect (QA5)

$$QA5 = \frac{N_{re\_all} + N_{ra\_all} + N_{rg\_all} + N_{ri\_all} + N_{rt\_all} + N_{t\_i}}{N_E + N_{at} + N_g + N_t}$$

XML schema flexibility quality aspect (QA6)

$$\text{QA6} = \frac{\begin{gathered} N_{E_{group}} + N_{A_{group}} + N_g + N_{re_{all}} + N_{ra_{all}} + N_{rg_{all}} + N_{ri_{all}} \\ + N_{An} + N_{rt\_all} - N_{E\_U} \end{gathered}}{N_E + N_{at} + N_t + N_g}$$

XML Schemas Quality aspects and quality index

$$\mathrm{Q_i} = 1/6(\mathrm{QA_1} + \mathrm{QA_2} + \mathrm{QA_3} + \mathrm{QA_4} + \mathrm{QA_5} + \mathrm{QA_6})$$

Set of domains, using XML schemas

D1 - Mathematics and Physics
D2 - Materials Science
D3 - Telecommunications
D4 - Manufacturing
D5 - Energy and Electronics
D6 - Engineering
D7 - IT architecture and design
D8 - Traffic
D9 - Communications
D10 –Computer Science

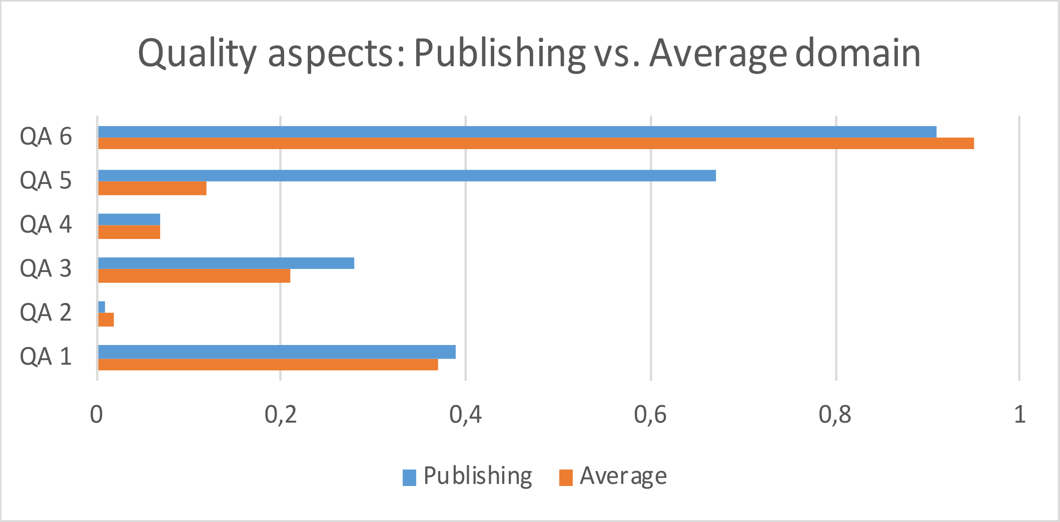
D11 –Decision Science
D12 –Medicine
D13 - Economics and finance
D14 - Law
D15 - Social science
D16 –Health and sport
D17 –Construction
D18 - Librarianship (Library)
D19 - Landscape and geography
D20 –Media, journalism, newspapers

D21 - The publishing domain

Comparing publishing domain with average results (of existing set of domains)

PARAMETERS		
	Publishing domain - average	All domain - average
Number of imports	1,400	0,795
number of all elements	86,600	77,727
number of global elements	48,000	26,755
number of local elements	38,600	50,973
number of simple elements	31,800	27,691
number of complex elements	54,800	50,036
number of global complex elements	35,700	19,250
number of global simple elements	12,700	7,514
number of all attributes	26,600	47,655
number of local attributes	26,600	47,091
number of global attributes	0,000	0,564
Lines of code	14969,200	3188,618
number of element groups	1,000	4,364
number of attribute groups	1,300	1,377
number of element references	65,400	69,118
number of references on simple elements	9,000	3,177
number of references on complex elements	56,400	65,941
number of references on attributes	0,000	1,927
number of references on element groups	0,200	7,114
number of references on attribute groups	3,700	11,664
Number of annotations	0,900	0,977
Number of restrictions	30,100	75,118
Number of derived (extended) types	41,800	35,309

Evaluation of the publishing domain



QUALITY APSPECTS		
	Publishing domain - average	All domain - average
Structural quality aspect (QA1)	0,387	0,374
Transparency and documentation of the XML Schema (QA2)	0,005	0,022
XML schema optimality quality aspect (QA3)	0,280	0,214
XML schema minimalism quality aspect (QA4)	0,071	0,072
XML schema reuse quality aspect (QA5)	0,667	0,117
XML schema flexibility quality aspect (QA6)	0,908	0,950

Evaluation of the publishing domain

- XML schemas from the publishing field are above average:
  - the publishing domain does use XML schemas,
  - the quality of them is above average however they still need to be improved mostly in the quality aspect of:
    - transparency, documentation
    - flexibility

Research questions

Does the publishing domain use XML documents and what standard XML schemas are being used?	Several XML schemas were found, connected to the publishing field (respectively publishing process) through active research.
What is the quality level of XML schemas in the publishing domain?	Average quality of XML schema in the publishing field is 39%.
How are they compared to XML schemas in other domains such as computer science and other?	The quality index of 39% is higher than by the average quality index (of all 20 domains, where XML schemas are most common) which is 29% based on an experiment in 2014.
How can the level of quality be improved?	Comparing to average XML schemas, the publishing field had lower results only at transparency and documentation quality aspect, all other quality aspects were above average.

Work in progress…

Publishied work

*„PUŠNIK, Maja, HERIČKO, Marjan, BUDIMAC, Zoran, ŠUMAK, Boštjan. **XML Schema metrics for quality evaluation**. Computer science and information systems, 2014, vol. 11, no. 4, str. 1271-1289 “*

*„PUŠNIK, Maja, RAKIĆ, Gordana, BUDIMAC, Zoran, HERIČKO, Marjan. **Different approaches for measuring XML Schemas**. Collaboration, software and services in information society : proceedings of the 18th International Multiconference Information Society - IS 2015, October 12th, 2015, Ljubljana, Slovenia : volume D. Ljubljana: Institut Jožef Stefan, 2015, str. 17-20. “*

Sent work

*„Maja Pušnik, Marjan Heričko And Boštjan Šumak, Gordana Rakić, X**ML Schema Quality index in the multimedia content publishing domain**, SQAMIA 2016“*

*„Gordana Rakić, Zoran Budimac, Marjan Heričko, Maja Pušnik . **Towards the XML Schema Measurement Based on Mapping Between XML and OO Domain**. SCLIT 2016“*

Discussion

- What are the (unpremeditated) problems in the publishing process?
- What is the existing level of IT support in similar processes?
- What is the quality of the publishing process?
- What is the user experience of all involved?
- How often errors occur and how critical they are?

- Are XML Technologies the only solution?
- Are XML Technologies the best solution?
- Are XML Technologies the most suitable solution for the publishing domain?

Thank you for listening!

QUESTIONS?

Maja.Pusnik@um.si

Bostjan.Sumak@um.si