Universidade do Minho

Topic Maps 2007

2007-03-21

Topic Maps Constraint Specification Languages

Giovani Rubert Librelotto
giovani@unifra.br

José Carlos Ramalho
jcr@di.uminho.pt

Renato Preigschadt
rpa.renato@gmail.com

Pedro Rangel Henriques
prh@di.uminho.pt

Universidade do Minho

Agenda

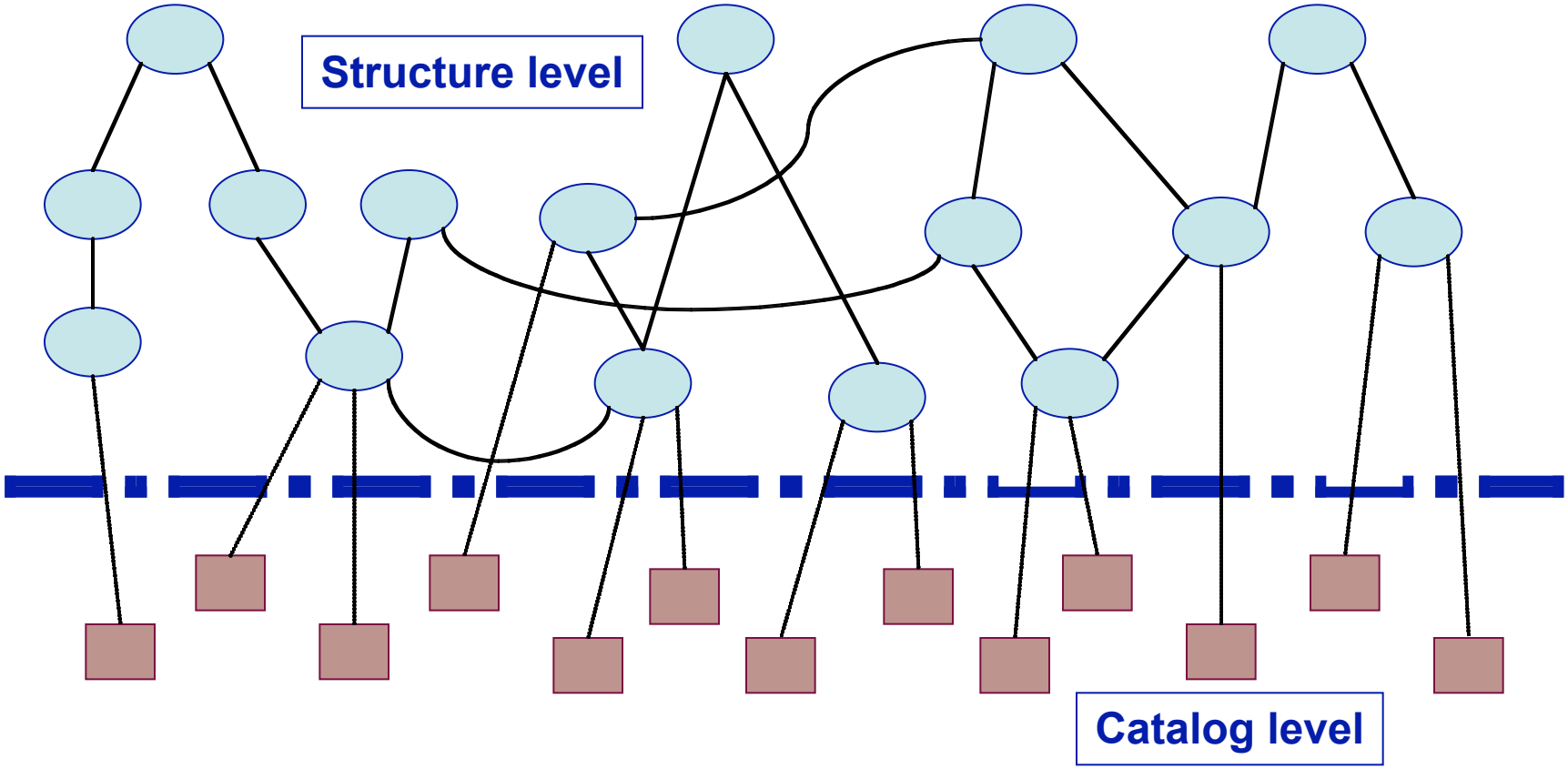
- Motivation
- Our work's background
- What's going on (models)?
- Constraints...
- XTche, ASTMA and OSL
- Case study
- Conclusions

Universidade do Minho

Motivation

- Why do we need constraints?
  - Topic Maps are too abstract!
    - Many ontologies have a rigid structure
    - Many ontologies have semantic constraints
  - we do not control the process of creation
    - uploads
    - merges
  - manual editing of the Topic Map
- As a side effect we can have a Reporting Service

Ontology



Universidade do Minho

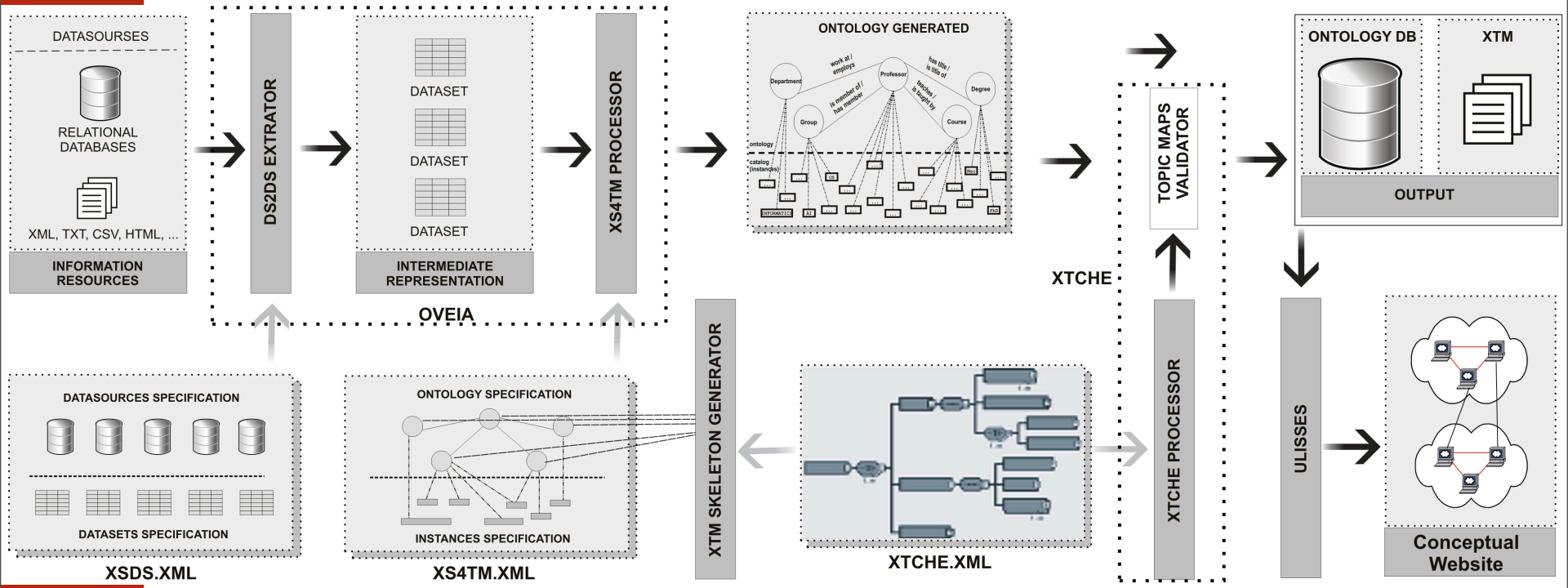
More questions

- Is a Schema a Constraint Language?
- Creating a new language ... Why?
  - We are always creating new DSL arguing that we will hide complexity
  - Aren´t we introducing complexity?
- How should we implement this new language?

Universidade do Minho

Metamorphosis

Back in 2001 we had great expectations towards Topic Maps



Universidade do Minho

TMDiscovery

★ When the number of applications grows you start to worry about the model and storage...

- In 2005 an Msc Work delivered a Relational Model for TM and a browser working over it
- In 2006 TMDiscovery was tested:
  - Worked well with small Topic Maps
  - It was used in the classroom and in small projects
  - Failed to scale with large Topic Maps (>50.000 topics): AGROVOC, Emigration Museum DB.

RODA: 2006/2007

- Digital Preservation
- Digital Objects: text documents, still images, relational databases.
- Model specification:
  - Descriptive Metadata (EAD)
  - Preservation Metadata (PREMIS)
  - Technical Metadata (NISO MIX, DBML, ...)
  - Wrapping metadata (METS)

http://roda.iantt.pt

Universidade do Minho

Topic Maps 2007

Universidade do Minho

RODA: the model

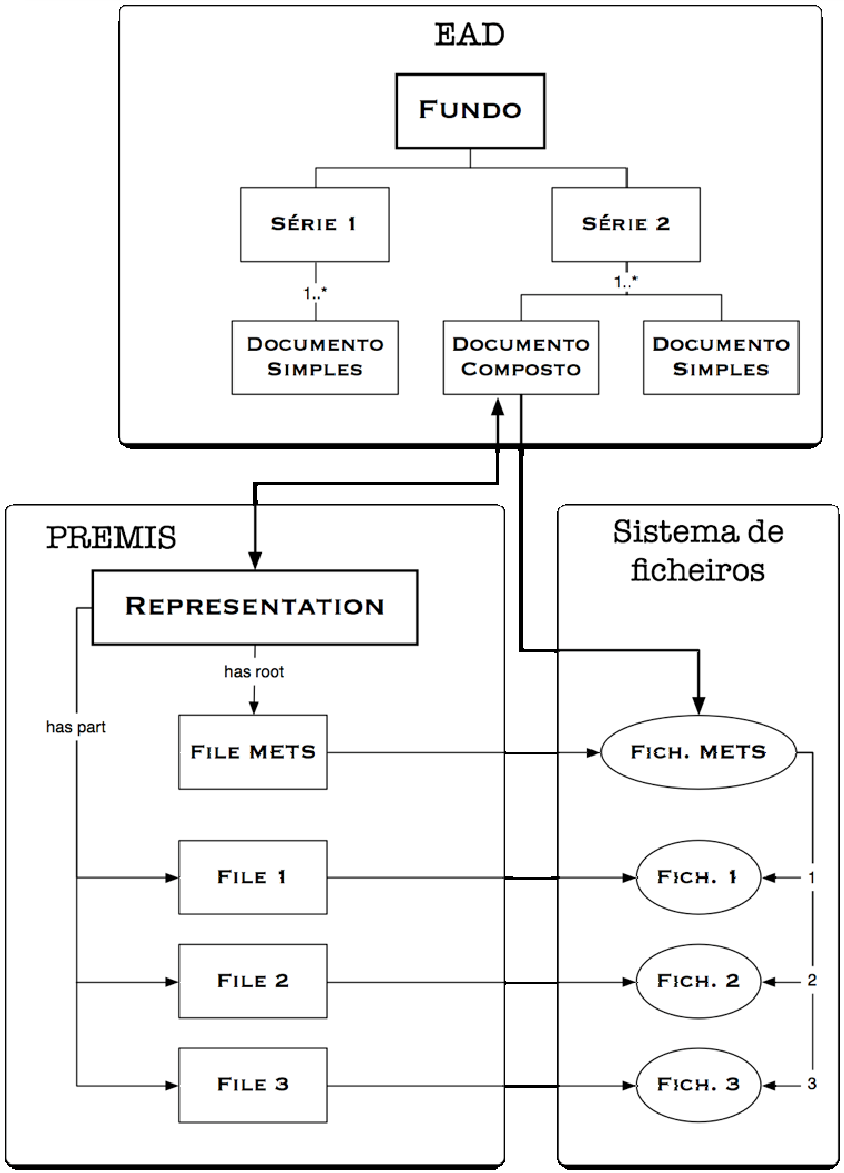
Topic Maps 2007

Universidade do Minho

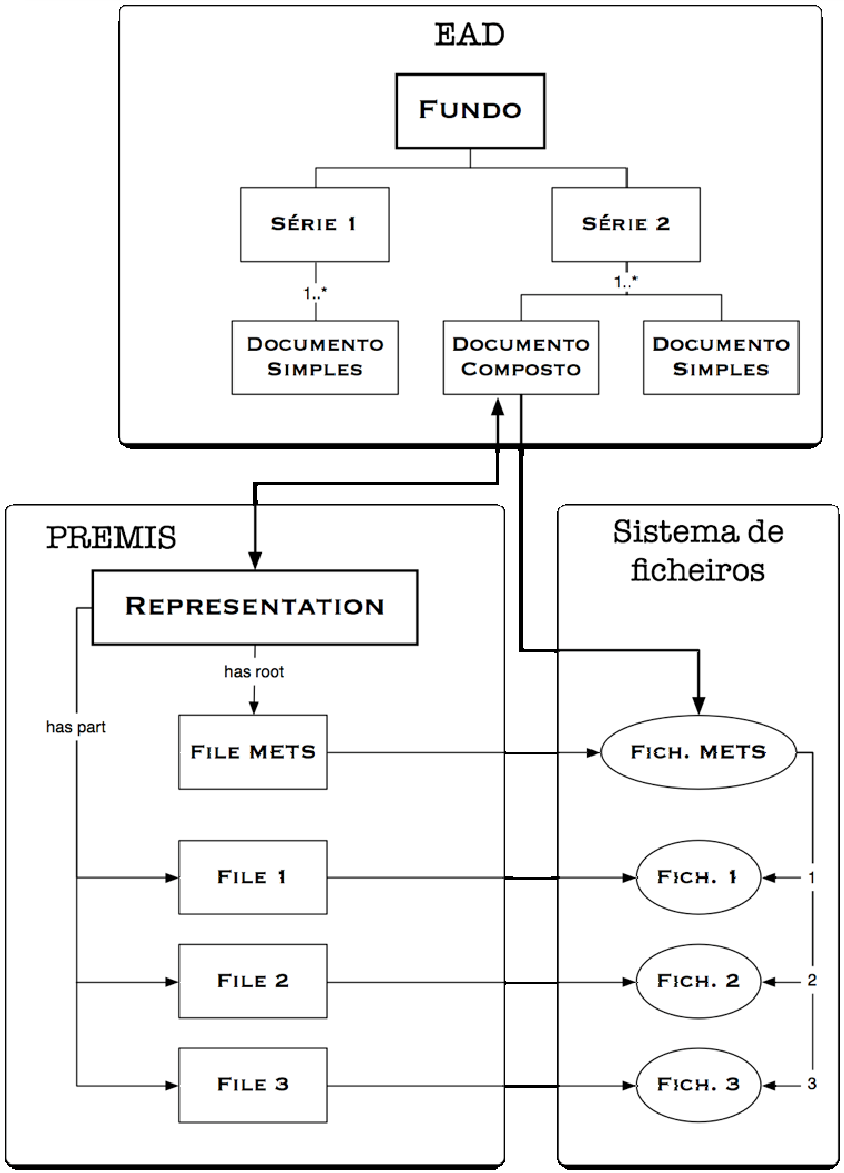
RODA: the model

Topic Maps 2007

RODA: the model



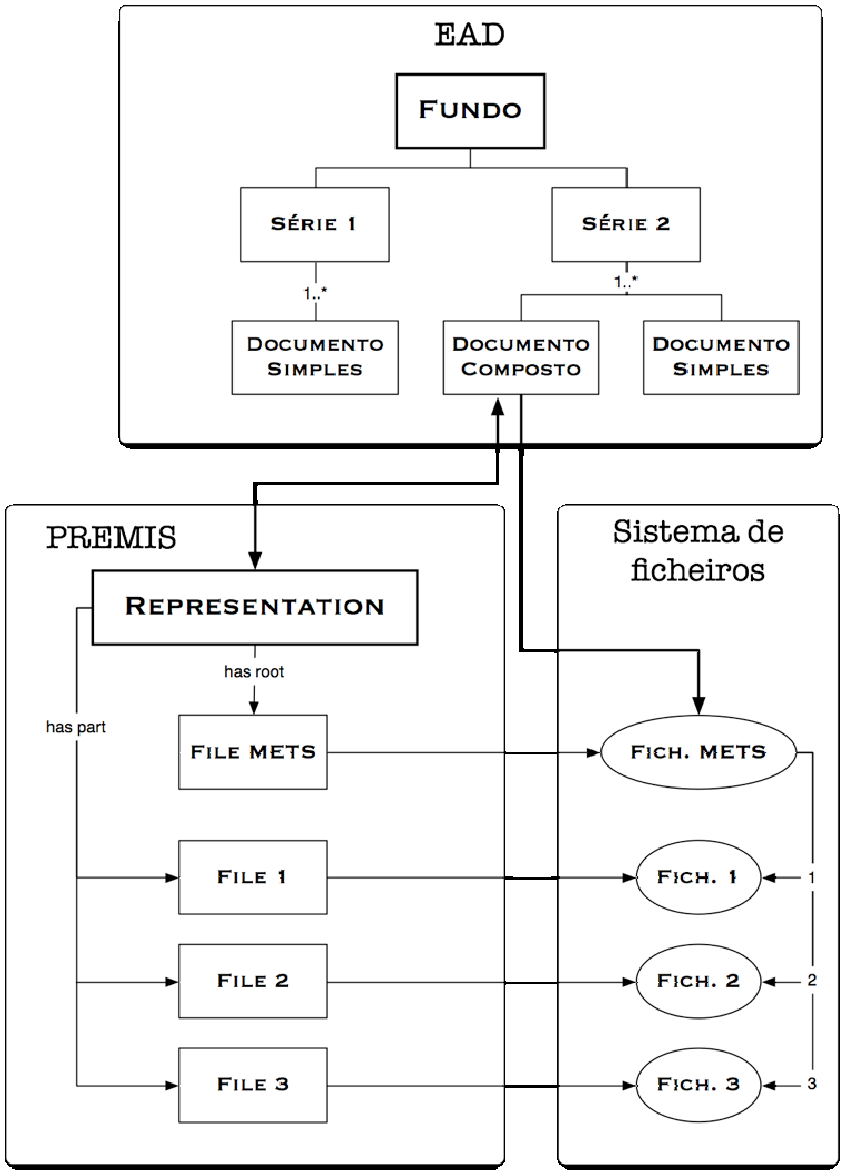
RODA: the model



First real digital object:

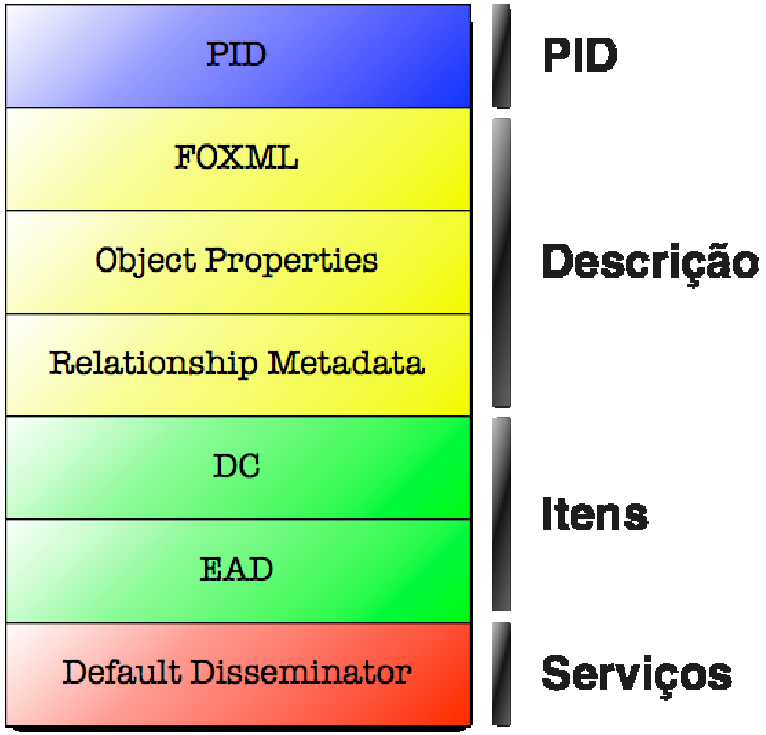
- 2000000 EAD nodes
- 2000000 PREMIS nodes
- 2000000 file nodes
- each preserv. event can add 2000000 nodes

RODA: the model



First real digital object:

- 2000000 EAD nodes
- 2000000 PREMIS nodes
- 2000000 file nodes
- each preserv. event can add 2000000 nodes



Universidade do Minho

Returning to constraints

- What is really a TM?
  - According to the XTM DTD
    - A list of ... Topics, associations, ocurrences
    - Everything else is constructed by reference
    - **It is easy to get lost in a TM!**
  - That DTD is completely abstract
    - There is room for “Specific Semantic Networks Schemas”

Universidade do Minho

TMCL: Implementation wishes

- Do it simple
- Do it with existing technology
- Do it user friendly
  - If possible do not force the user to learn a new language

Universidade do Minho

Topic Maps 2007

TMCL: Requirements from ISO JTC1 SC34

Schema constraints:
Topic of type T must have a specified number of explicit names/occurrences/subject-indicators (cardinality);
Topic of type T must have as name/occurrence/subject-indicators a value matching a particular pattern;
Topic of type T must (not) have a name/occurrence with scope S;
Topic of type T must have a name/occurrence, that is instance of topic type T, in scope S;
...

Contextual constraints:
Topic T can (only/not) be used for typing other topics;
Topic T can (only/not) be used for typing subject indicator;
Topic T can (only/not) be used for typing basenames;
Topic T can (only/not) be used for typing occurrences;
...

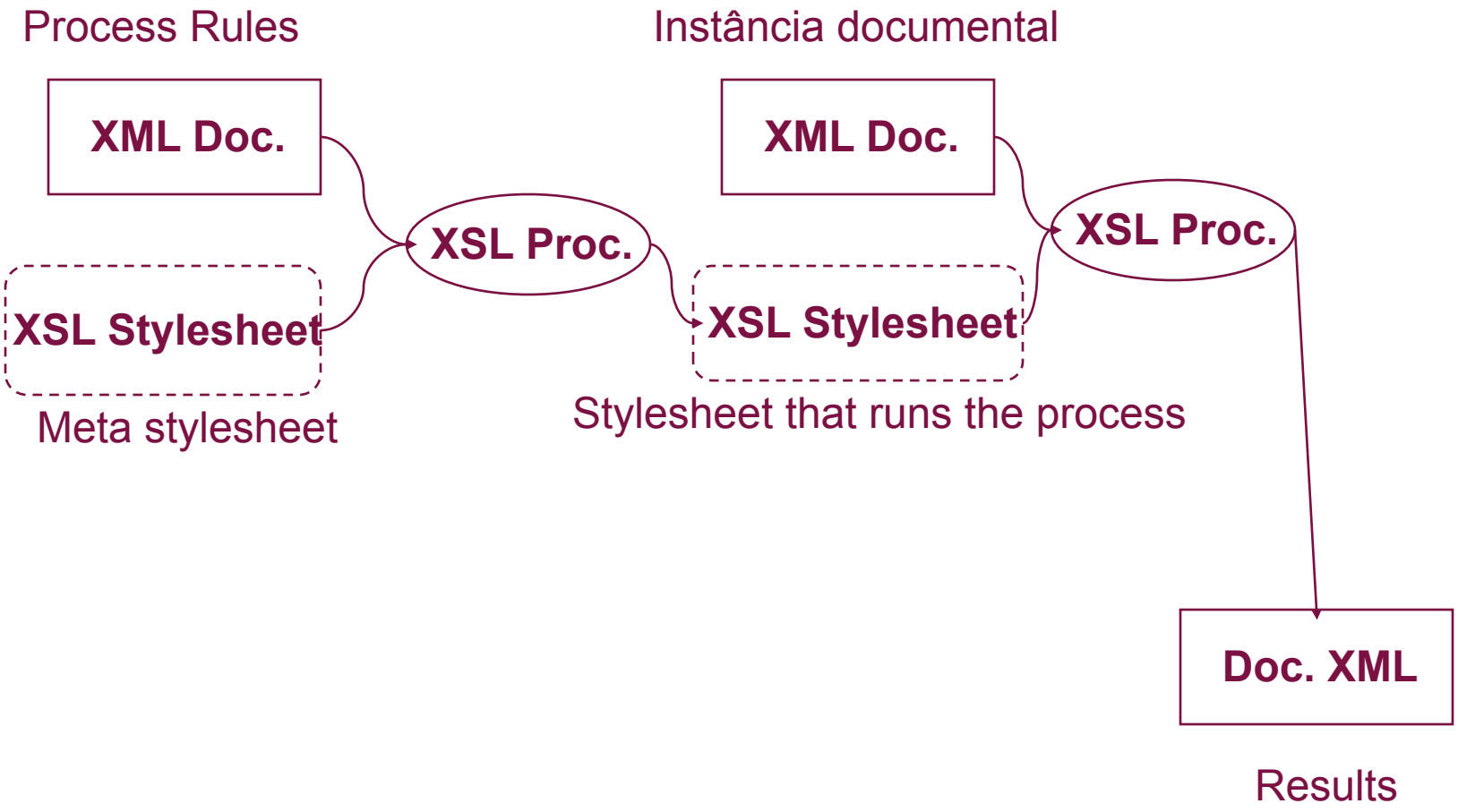
Universidade do Minho

How to ...?

- Schematron/XCSL processing model has been used in similar contexts
- For prototyping this processing model we would only need open "standards":
  - XML
  - XSLT
  - XML Schemas

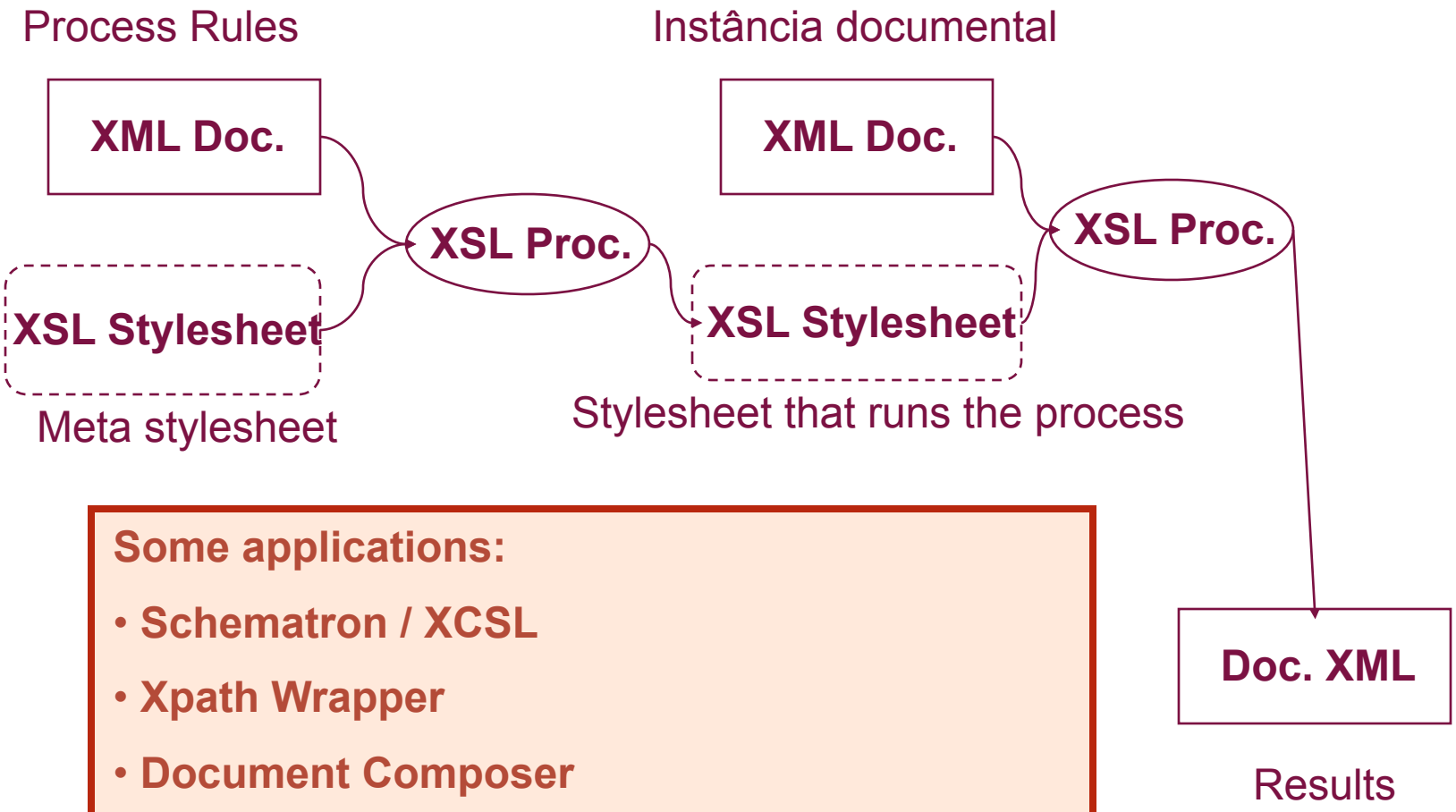
Pipeline

Multi-layer XSL Processing Architecture



Pipeline

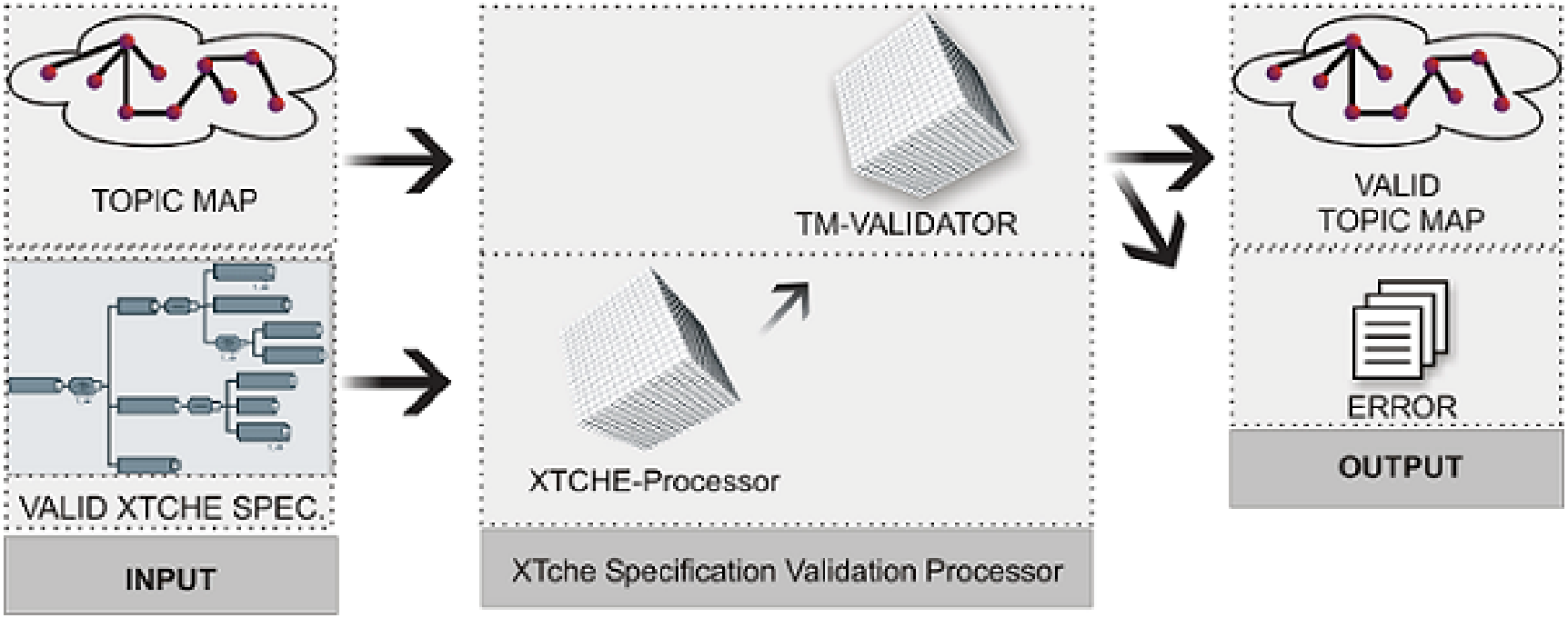
Multi-layer XSL Processing Architecture



Some applications:

- **Schematron / XCSL**
- **Xpath Wrapper**
- **Document Composer**
- **XDBT: XML Database Transformations**

XTche architecture

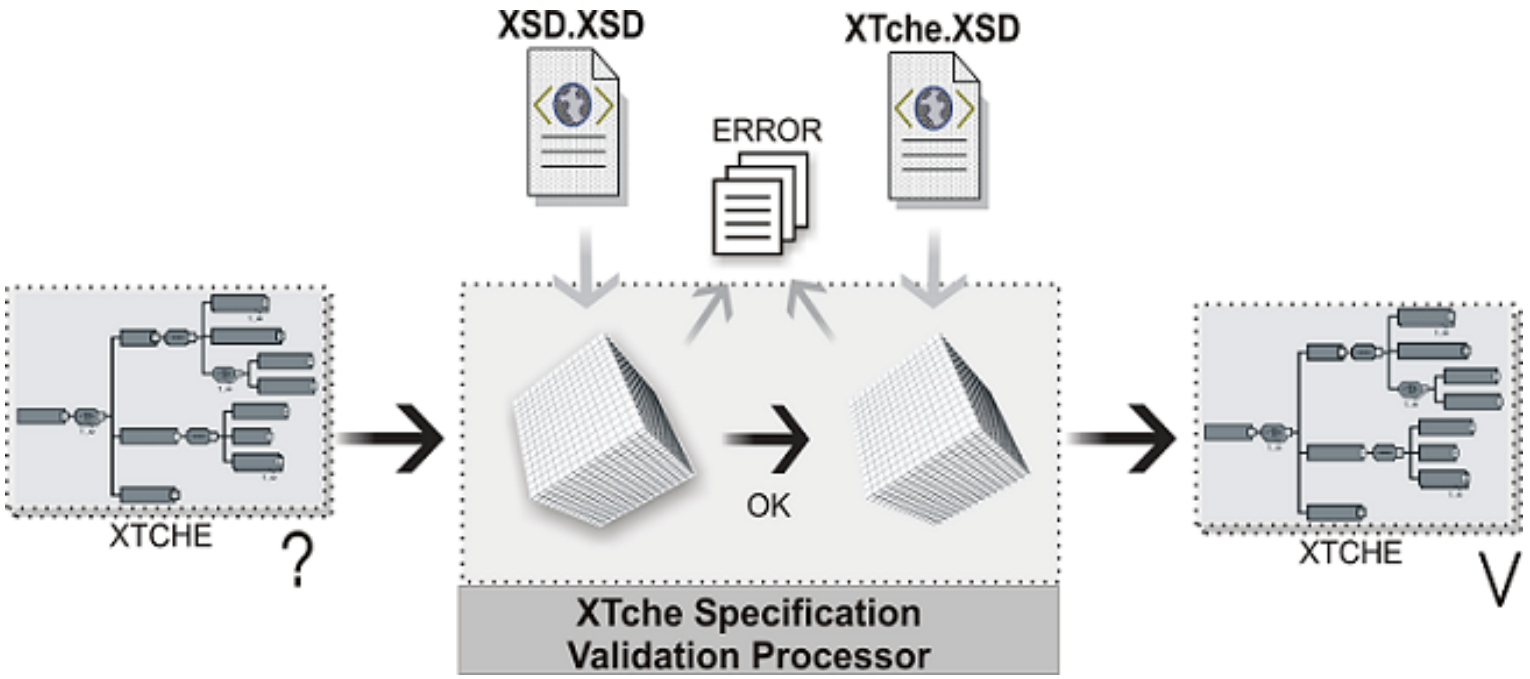


Universidade do Minho

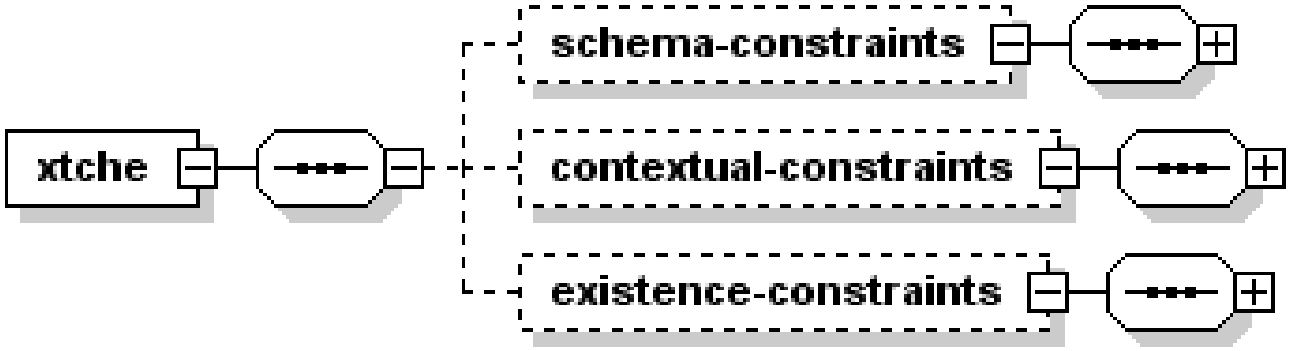
XTche Specification Language

- XML Schema based
  - Constraint lang. = Schema lang.
  - We have some experience with XML Schemas
  - The tools we use have an interesting graphical support that we wanted to use
  - If we choose an existing language to host ours part of the work is already done
- **XTche Spec. = XML Schema + semantic stamps**

XTche Validation Process



XTche skeleton



- Namespace: **xmlns:xtche=**http://www.di.uminho.pt/~gepl/xtche
- Imports: http://www.di.uminho.pt/~gepl/xtche/xtche-schema.xd

Universidade do Minho

xtche-schema.xsd

- A set of flags / semantic stamps
- Each flag represents a semantic validation
- The user will associate flags to concrete topics and these will map to semantic constraints that will be verified
- Remaining: **Did we cover the wole TMCL?**

Universidade do Minho

Topic Maps 2007

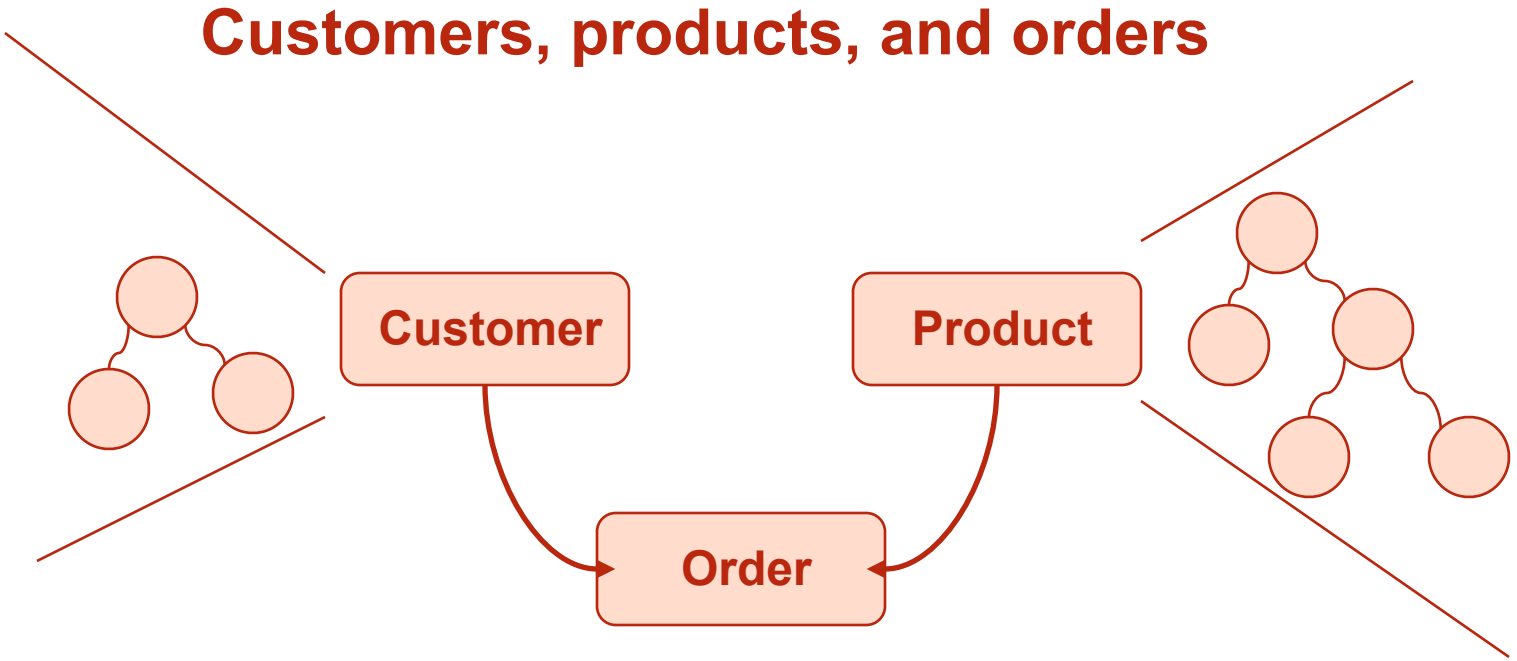
Flags / Semantic Stamps

- Schema constraints
  - **<xs:attribute name="topicType"/>**
  - "this element represents a Topic type"
- Contextual constraints
  - **<xs:attribute name="associationTypeExclusive"/>**
  - "this topic can only be used to qualify associations"
- Existence contraints
  - To guarantee the existence of at least a specific topic or association

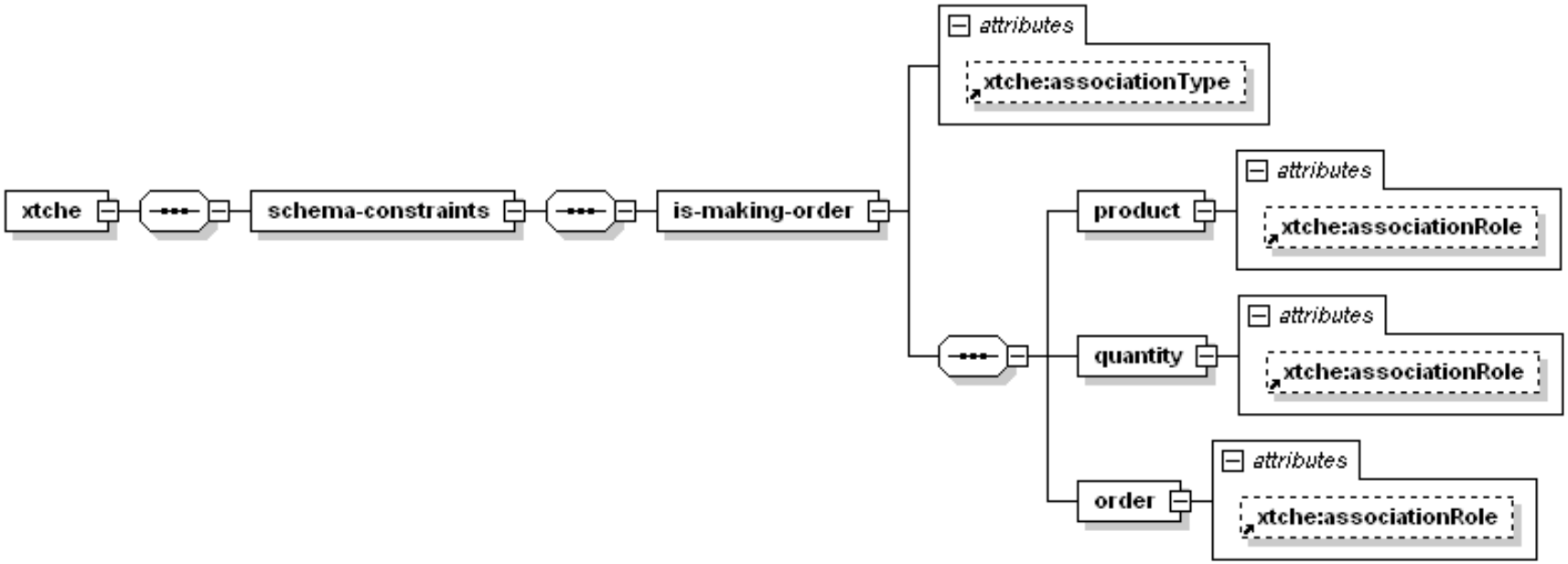
Universidade do Minho

A small case study

Topic Maps 2007



Schema constraint

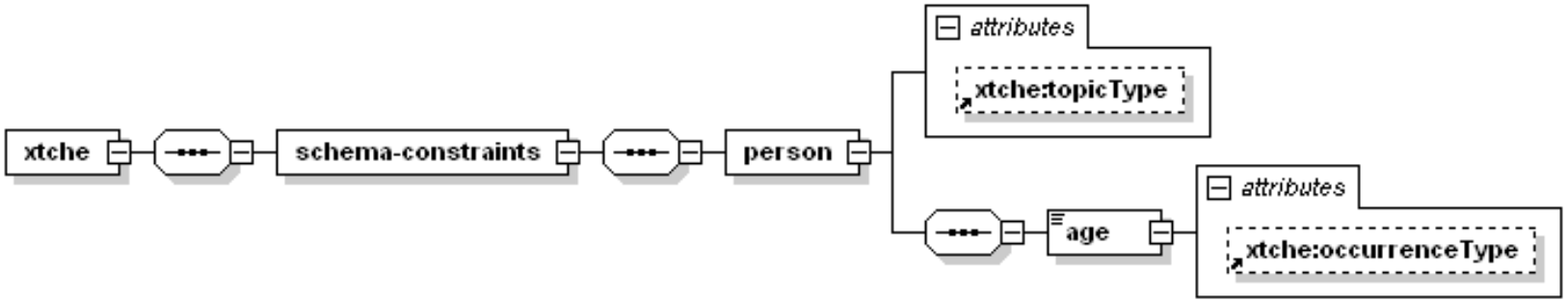


Restricting an Association structure:

$\forall a \in Associations$: if(Type(a) = "is-making-order"

then MemberRoles((order, product, quantity) a)

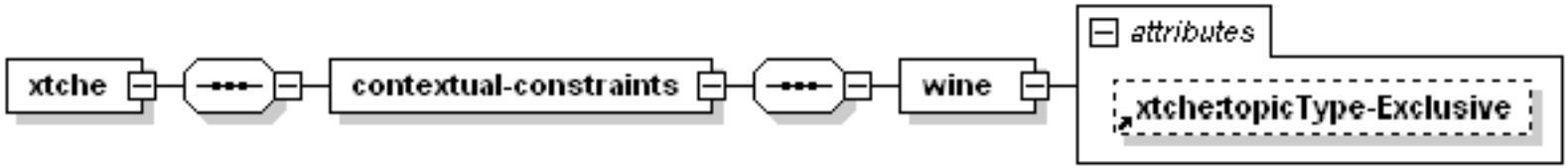
Schema constraint



Each person must have an age field:

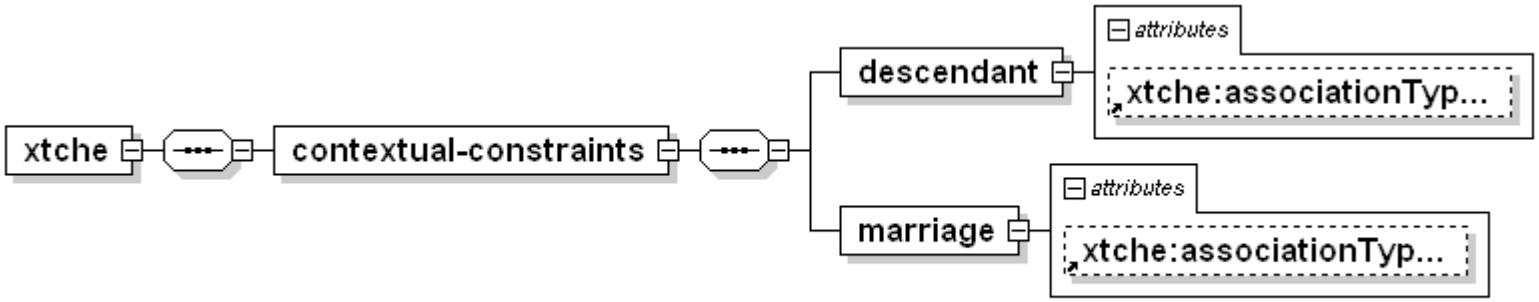
$\forall t \in \text{Topics}, \text{TopicType}(t)=\text{"person"} \Rightarrow \exists r \in \text{Occurrences} : r \subset t \wedge \text{OccType}(r) = \text{"age"}$

Contextual constraint



Restricting the use of a specific topic: topic “wine” can only be used as a Topic Type

Reporting Services



- Semantic Stamp = xtche:associationType-Forbidden
- topic “descendant” cannot be used to type associations
- topic “marriage” cannot be used to type associations

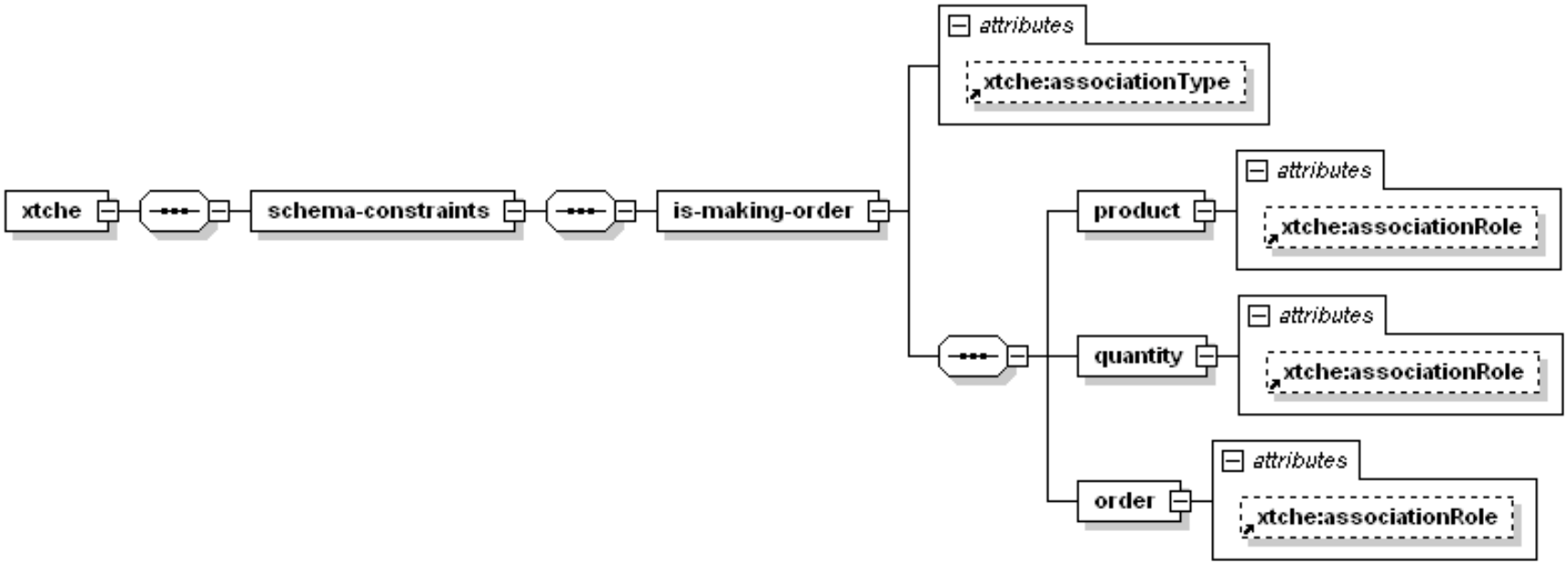
- Expected result: number of descendant and marriage associations

Universidade do Minho

Astma language set

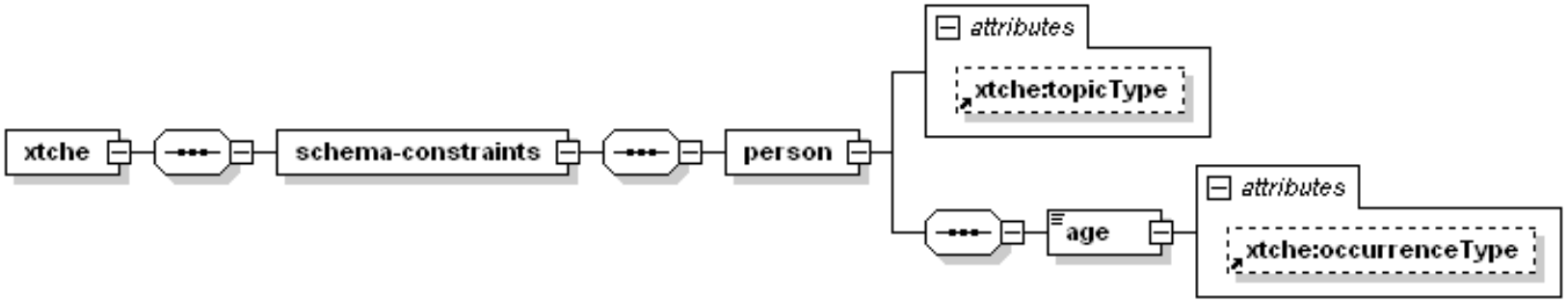
- Non-XML syntax
- Language for: structuring, adding, querying, constraining...
- An object algebra...
- It has its own model. The Topic Map must be Uploaded to this model.

Schema Constraint



forall [* (is-making-order)
product: *
quantity: *
order: *]

Schema Constraint



forall [* (person)
  age: *]

Astma does not distinguish between Topics, Associations, Occurrences, ...

Universidade do Minho

Ontopia Schema Language

- XML Schema based
- Allows the user to specify a schema for a specific Topic Map
- Since we are in the schema domain we are not allowed to use instance values

Universidade do Minho

Comparing the 3 Languages

	XTche	ASTMA	OSL
Background	XML Schema + Stamps	Specific Syntax	XML Schema
Model	Abstract Doc. Tree (XTM)	Specific (should be open)	Abstract Doc. Tree (XTM)
Range of App.	Til now... it looks promising	Some doubts...	Only structural constraints
Proc. Engine	Schema and XSL proc.	Specific	Schema proc.
Scalability	Does not scale	???	Does not scale

Universidade do Minho

Conclusions

- There is work to be done
  - to optimize (in our case, in the XSL domain, it can be done a lot...)
  - to find a better model
    - Object: (id, [Property], [RelationShip])
  - or better implementations for it

Some ongoing projects (Digital Preservation):

RODA: http://roda.iantt.pt

CRiB: http://crib.dsi.uminho.pt

Questions...?

Giovani Rubert Librelotto
giovani@unifra.br

José Carlos Ramalho
jcr@di.uminho.pt