How to quickly detect the overlap and the consistency between LADM with LandInfra and LandXML
-
Application of schema matching techniques

UNIVERSITÉ LAVAL

Jacynthe Pouliot, Suzie Larrivée

Department of Geomatics Sciences

Hes·so

Haute Ecole Spécialisée de Suisse occidentale

Fachhochschule Westschweiz

University of Applied Sciences and Arts Western Switzerland

Claire Monney, Jens Ingensand

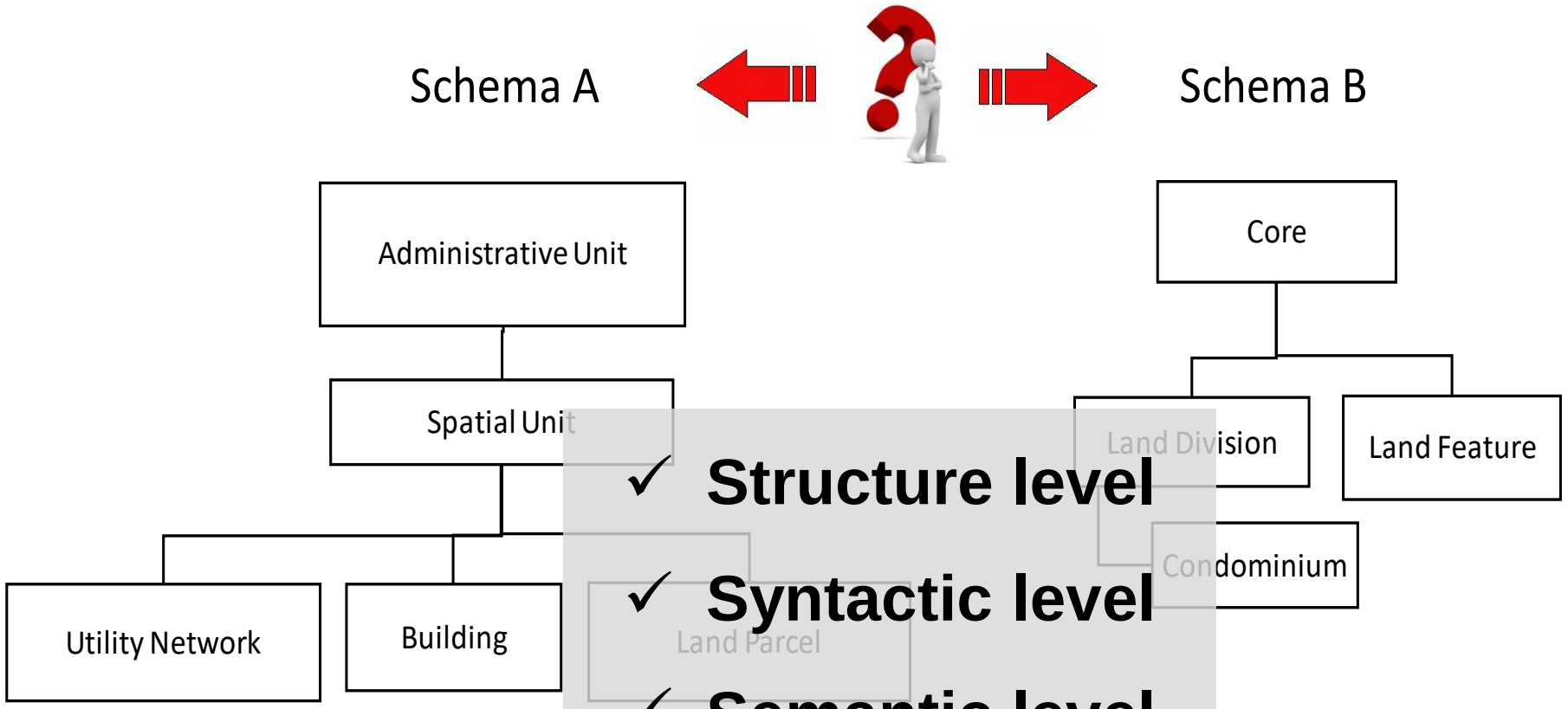
School of Business and Engineering Vaud (HEIG-VD)

Best standard to suit my needs ?

- Pouliot et al., 2018 (3D Geoinfo)
  - First experience about underground utility network

One possible solution…

Schema matching techniques



WordNet Search - 3.1

- WordNet home page - Glossary - Help

Lexical database of English

Word to search for: network Search WordNet

Display Options: (Select option to change) Change

Key: "S:" = Show Synset (semantic) relations, "W:" = Show Word (lexical) relations

Display options for sense: (frequency) {offset} <lexical filename > [lexical file number] (gloss) "an example sentence"

Display options for word: word#sense number (sense key)

Noun

- (23){08451269} <noun.group>[14] S: (n) **network#1 (network%1:14:00::)**, web#4 (web%1:14:00::) (an interconnected system of things or people) *"he owned a network of shops"; "retirement meant dropping out of a whole network of people who had been part of my life"; "tangled in a web of cloth"*
- (4){03826014} <noun.artifact>[06] S: (n) **network#2 (network%1:06:01::)** ((broadcasting) a communication system consisting of a group of broadcasting stations that all transmit the same programs) *"the networks compete to broadcast important sports events"*
  - *domain category*
  - *direct hypernym* / *inherited hypernym* / *sister term*
    - {03081962} <noun.artifact>[06] S: (n) communication system#2 (communication_system%1:06:00::), communication equipment#1 (communication_equipment%1:06:00::) (facility consisting of the physical plants and equipment for disseminating information)
- (1){03825135} <noun.artifact>[06] S: (n) net#6 (net%1:06:00::), **network#3 (network%1:06:00::)**, mesh#4 (mesh%1:06:00::), meshing#2 (meshing%1:06:01::), meshwork#1 (meshwork%1:06:00::) (an open fabric of string or rope or wire woven together at regular intervals)

Schema matching techniques for comparing standards…?

Good idea

Bad idea

Best standard to suit cadastre and land administration systems?

How quickly and efficiently compare LADM, LandInfra, LandXML ?

ISO 19152:2012

GEOGRAPHIC INFORMATION -- LAND ADMINISTRATION DOMAIN MODEL (LADM)

Land and Infrastructure (LandInfra)

OGC LandInfra / InfraGML

LANDXML 2.0

Second experiment

- Compare the schemas of Landinfra, LandXML and LADM
- Where (what format) can we get the schema?
  - Any kind of Database management system
  - UML (Unified Model Language)
    - Class diagram
  - **XSD (XML Schema Definition)**
    - **Textual presentation of the classes, the attributes, the domain of values, definition, and hierarchical relations**

Sample of LADM XSD

LADM_complet.XSD - Bloc-notes

Fichier Edition Format Affichage Aide

```
<?xml version="1.0" encoding="utf-8"?>
<xs:schema xmlns:xs="http://www.w3.org/2001/XMLSchema">
    <xs:element name="LA_SpatialUnit" type="LA_SpatialUnit"/>
        <xs:complexType name="LA_SpatialUnit">
            <xs:annotation>
                <xs:documentation>An instance of class LA_SpatialUnit is a spatial unit. A spatial unit may be
                associated to zero or more [0..*] basic administrative units (i.e. the spatial unit may be used to
                describe the extent – part of – a basic administrative unit).
                </xs:documentation>
            </xs:annotation>
            <xs:sequence>
                <xs:element name="area" type="xs:string" minOccurs="0" maxOccurs="unbounded"/>
                <xs:element name="dimension" type="xs:string" minOccurs="0" maxOccurs="1"/>
                <xs:element name="extAddressID" type="xs:string" minOccurs="0" maxOccurs="unbounded"/>
                <xs:element name="label" type="xs:string" minOccurs="1" maxOccurs="1"/>
                <xs:element name="referencePoint" type="xs:string" minOccurs="0" maxOccurs="1"/>
                <xs:element name="suID" type="xs:string" minOccurs="1" maxOccurs="1"/>
                <xs:element name="surfaceRelation" type="xs:string" minOccurs="0" maxOccurs="1"/>
                <xs:element name="volume" type="xs:string" minOccurs="0" maxOccurs="unbounded"/>
            </xs:sequence>
        </xs:complexType>
```

Class name

Definition

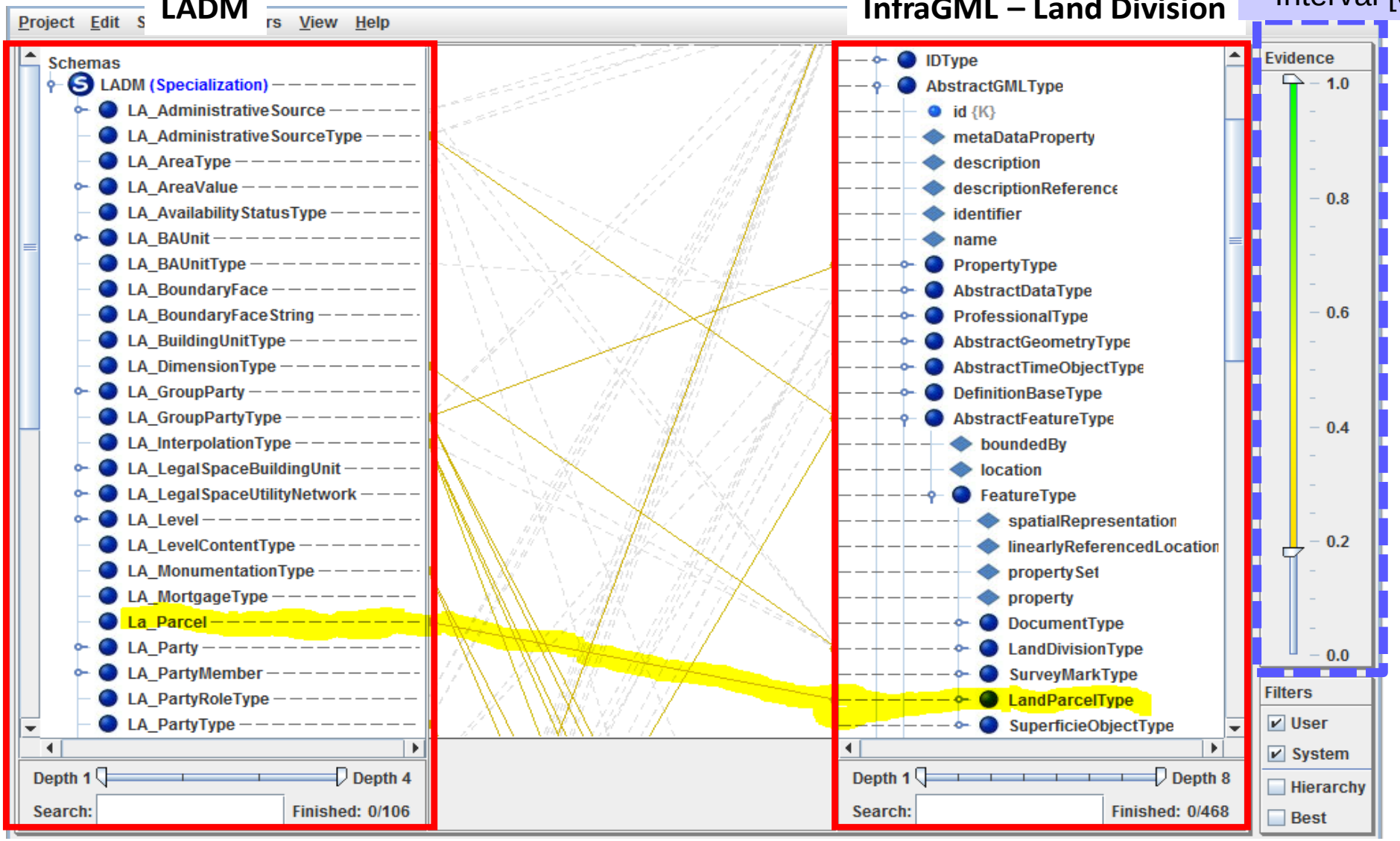
Attributes

XSD files under comparison

- **LADM**
  - 1 XSD file = 50 classes
- **LandInfra (InfraGML)**
  - 8 parts (Core / LandFeatures / Facilities and Projects / Alignments / Roads / Railways / Survey / 7 Land Division)
  - 15 separate XSD files, when combined = 446 classes
- **LandXML 2.0**
  - 14 packages (Alignments / Application / CgPoints / CoordinateSystem / GradeModel / Monuments / Parcels / PipeNetworks / PlanFeatures / Project / Roadways / Surfaces / Survey / Units)
  - 1 XSD file = 223 classes

OpenII – Harmony Diagram

Matching score
Interval [0,1]



Comparison performed

- **Only class name** (no attribute)
  - Too much attributes = same name as ID, NAME, TYPE
- Only 3 options tested:

  Option 1. **Only Syntax**: Matching score = edit distance

  Option 2. **Only Semantic (Wordnet)**: Sense of the class name and synonym, hyponym, hypernym of a lexical database.

  Option 3. **Syntax and Semantic**: Combination of options 1+2
- Result's interpretation
  - A major challenge…

Strategy to interpret the results

- LADM-LandInfra
  - 50 * 446 = 20 300 possible matches
- LADM-LandXML
  - 50 * 223 = 11 150 possible matches
- With option 1 (only syntax)
  - LADM-LandInfra = 18 798 matches (scores higher than 0.001)

Strategy to interpret the results

- 3 groups of level of similarity:
  - Tightly match
    - Matching scores higher than 0.4
  - Loosely match
    - Matching scores between 0.2 and 0.4
  - No match
    - Matching scores between 0 and 0.19

Results

Read the papers … (all detail and tables)

Only Syntax (no tightly match)

LADM	LandInfra (InfraGML)
LA_AdministrativeSourceType	AdministrativeDivisionPropertyType
LA_ResponsibilityType	CI_ResponsibleParty_Type
LA_BuildingUnitType	BuildingType
LA_Parcel	LandParcelType
LA_SpatialUnit	SpatialUnitType

Only Semantic (mainly tightly match)

LADM	LandInfra (InfraGML)
LA_Right	InterestInLandType
LA_LegalSpaceBuildingUnit	TimePositionUnion
LA_LegalSpaceBuildingUnit	DirectPositionType
LA_SpatialUnitGroup	LandDivisionType
LA_RequiredRelationshipSpatialUnit	OwnershipAttributeGroup

LADM class name match with …

	Syntax-Name		Semantic (Wordnet)		Syntax+Semantic (Wordnet)	
	LandInfra	LandXML	LandInfra	LandXML	LandInfra	LandXML
Tightly matches	0	0	14	11	22	3
Loosely matches	27	5	28	23	24	22
No match	23	45	8	16	4	25
Single match rate	**54%**	**10%**	**84%**	**59%**	**92%**	**47%**

LADM class name match with …

	Syntax-Name		Semantic (Wordnet)		Syntax+Semantic (Wordnet)	
	LandInfra	LandXML	LandInfra	LandXML	LandInfra	LandXML
Tightly matches	0	0	14	11	22	3
Loosely matches	27	5	28	23	24	22
No match	23	45	8	16	4	25
Single match rate	**54%**	**10%**	**84%**	**59%**	**92%**	**47%**

LADM class name match with …

	Syntax-Name		Semantic (Wordnet)		Syntax+Semantic (Wordnet)	
	LandInfra	LandXML	LandInfra	LandXML	LandInfra	LandXML
Tightly matches	0	0	14	11	22	3
Loosely matches	27	5	28	23	24	22
No match	23	45	8	16	4	25
Single match rate	**54%**	**10%**	**84%**	**59%**	**92%**	**47%**

LADM class name match with …

	Syntax-Name		Semantic (Wordnet)		Syntax+Semantic (Wordnet)	
	LandInfra	LandXML	LandInfra	LandXML	LandInfra	LandXML
Tightly matches	0	0	14	11	22	3
Loosely matches	27	5	28	23	24	22
No match	23	45	8	16	4	25
Single match rate	**54%**	**10%**	**84%**	**59%**	**92%**	**47%**

- **A large part of LADM content is covered by LandInfra, much less with LandXML**
- **LADM much closer to LandInfra compared with LandXML**

BUT are the matches correct??

Accuracy assessment

Accuracy assessment

- Comparison with independent and trustworthy works
  - Annex D of LandInfra official document (OGC 15-111r1, 2016)
  - Tables of Stubkjaer (2015)

LADM	Landinfra
LA_Source	7.2.1.3 Document
LA_RRR (Right, Restric. Resp.)	7.10.2 InterestInLand
LA_BAUnit	7.10.2.2 LandPropertyUnit
LA_Parcel alias LA_SpatialUnit	7.10.2.3 LandParcel
LA_Restriction	7.10.2.5 Easement
LA_SpatialUnitGroup	7.10.3 AdministrativeDivision
LA_AdministrativeSource	7.10.4 Statement
LA_SpatialUnit	7.10.6 SpatialUnit

Accuracy assessment

- **Percentage of correct and omitted matches**

	Correct matches	Omitted matches
LADM-LandInfra	9/15 (60%)	6/15 (40%)
LADM-LandXML	2/11 (18%)	9/11 (82%)

Lessons learned (1/4)

How applicable are schema matching techniques to compare geospatial standards?

- **Yes, applicable**
  - **Simple and fast with XSD files**
  - **Standards are well organized, good fit with schema matching**
- **But**
  - **XSD files not fully normalized**
  - **XSD content vary in levels of detail**
  - **Having many XSD files require lots of processing**

Lessons learned (2/4)

- What is the usefulness (rapidity and automation) of schema matching techniques?
  - Rapid and automate for overall comparison **BUT** not for detail comparison (the one that was of interest…)

Lessons learned (3/4)

- What is the accuracy of schema matching techniques?
  - Varying (not able to have a clear statement)
    - The number of samples tested is too limited
  - Depends on :
    - Naming principles (conceptual levels)
    - Sense of concepts (semantics)
    - For semantic: the quality of the external sources

Lessons learned (4/4)

- How do we define similarity levels between geospatial standards?
  - Interpretation of the results still a big challenge
  - Semantic is an important aspect
  - 3 proposed levels well adapted

Further works

- Supplementary tests and improve strategy
  - Integrate the attributes, description, code lists
  - Explore incremental schema matching
- Improve the integration of semantic
  - Quality of external sources
  - Definition within the standards
  - Natural Language Processing (NLP)
- Consider others standards or sources of validation

Acknowledgements