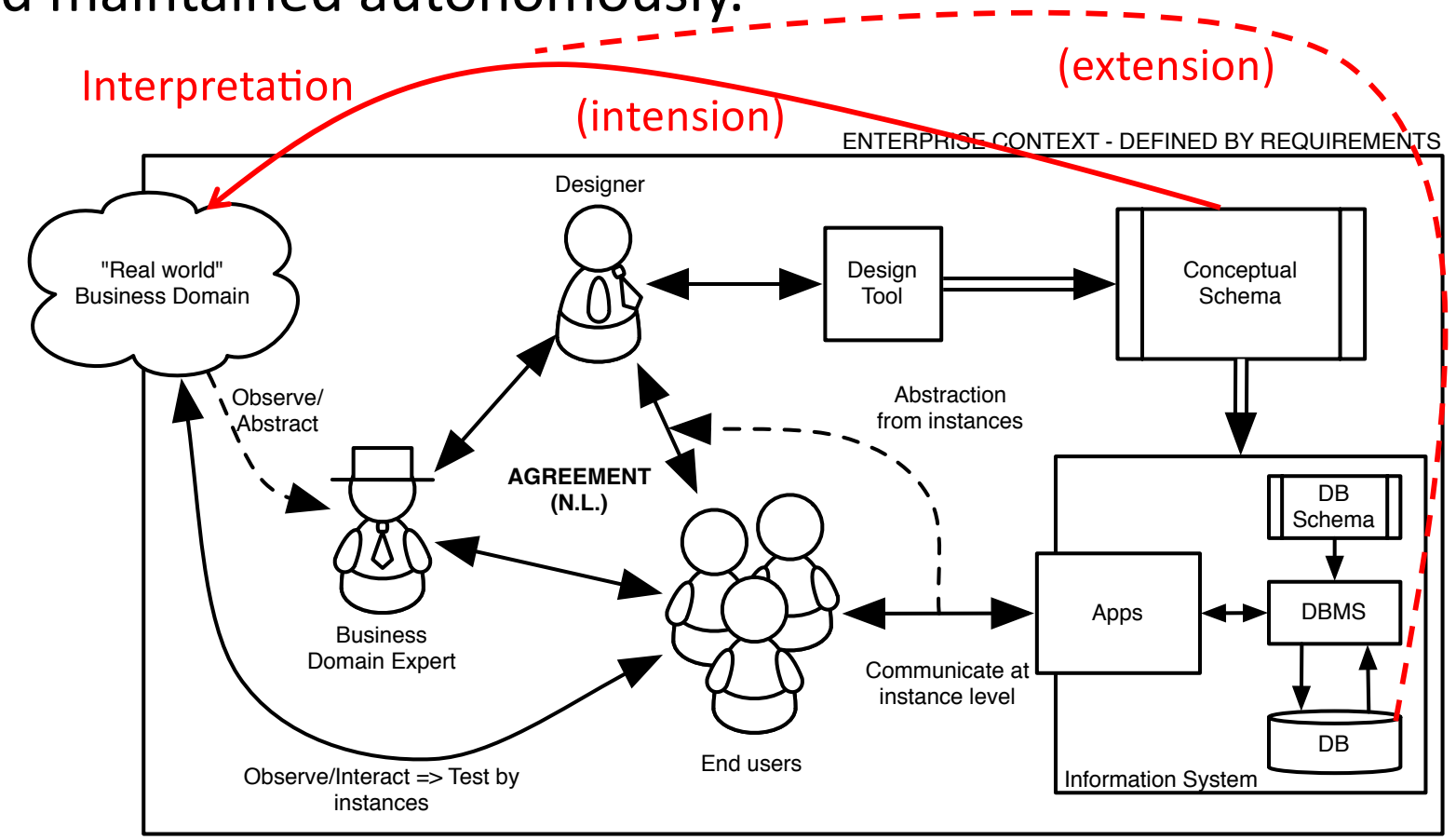
Open Information Systems 2019-2020

Lecture 1: Introduction to Ontology and Open Information Systems

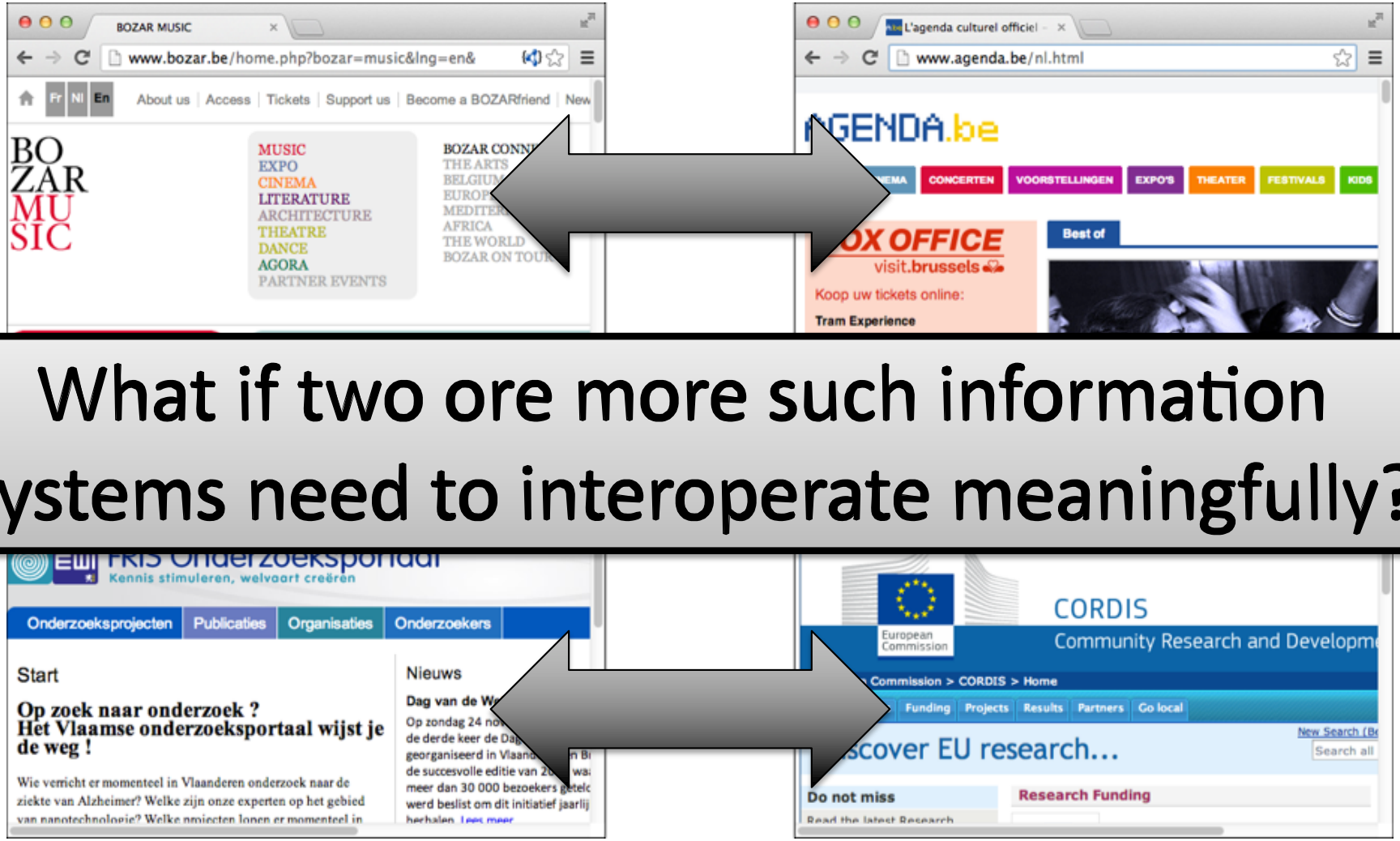
Christophe Debruyne

Context

- Information Systems (IS) on the Web are in general developed and maintained autonomously.



Context



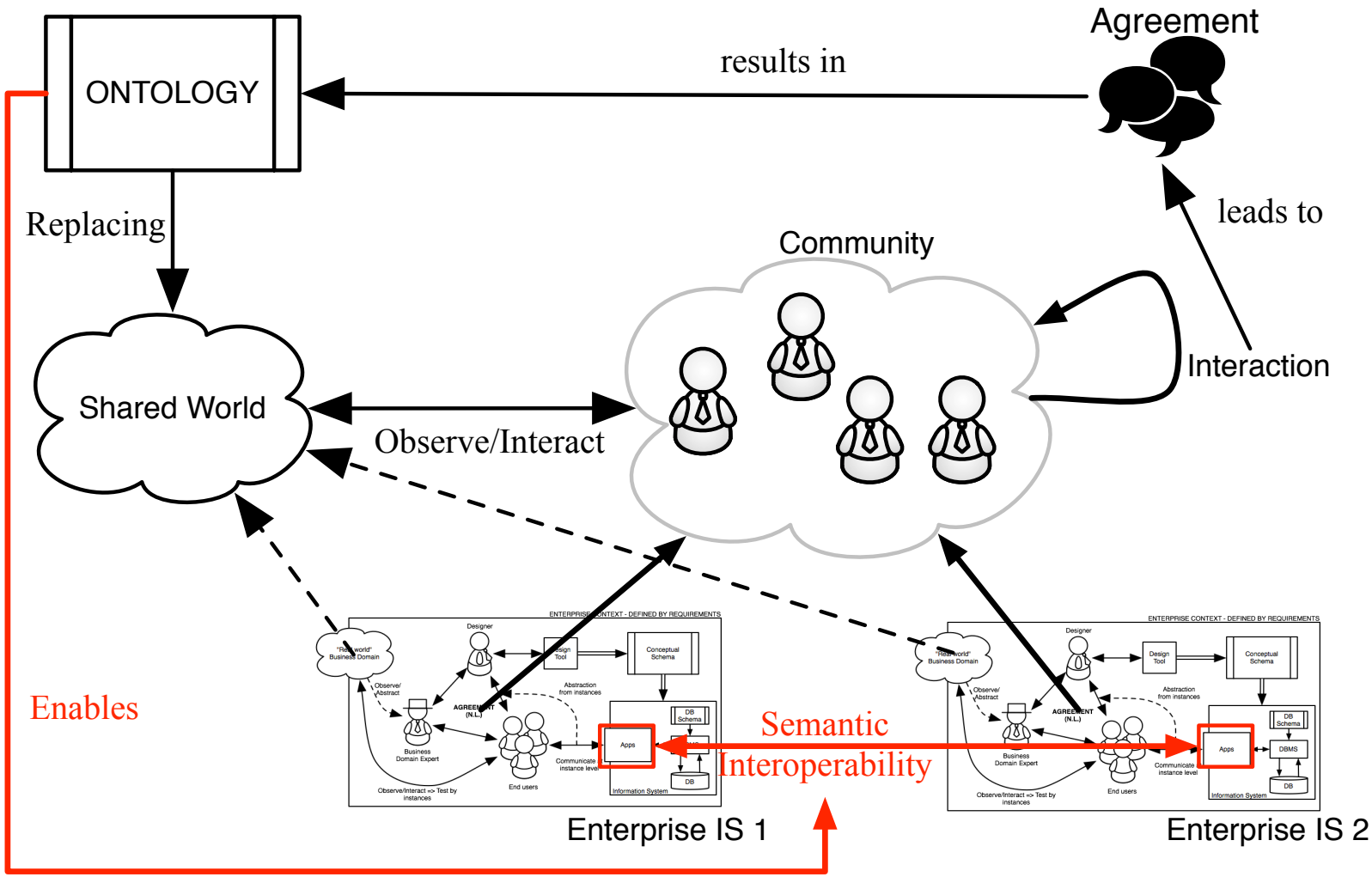
What if two ore more such information systems need to interoperate meaningfully?

Semantic Interoperability & Ontology

- **Semantic Interoperability** is defined as the ability of two or more information systems or their (computerized) components to exchange data, knowledge or resources and to interpret the information in those systems [DLM08].

- For IS to interoperate semantically, an **ontology** is needed
  - **"A [formal,] explicit specification of a [shared] conceptualization [Gru95] and extended by [Stu98]"**
  - Agreements among all the stakeholders;
  - Ontologies evolve while agreements are developed;
  - Ontologies are an externalization of the semantics outside an IS.

Semantic Interoperability & Ontology



Ontology Engineering is Critical

- We will see that the construction of ontology is not a trivial task, and that appropriate **methods** and **tools** for supporting this activity will be required.
- **Ontology engineering.** The set of activities that concern the ontology development process, the ontology life cycle, the principles, methods and methodologies for building ontologies and the tool suites and languages that support them, can be referred to as ontology engineering [GPFLC03].

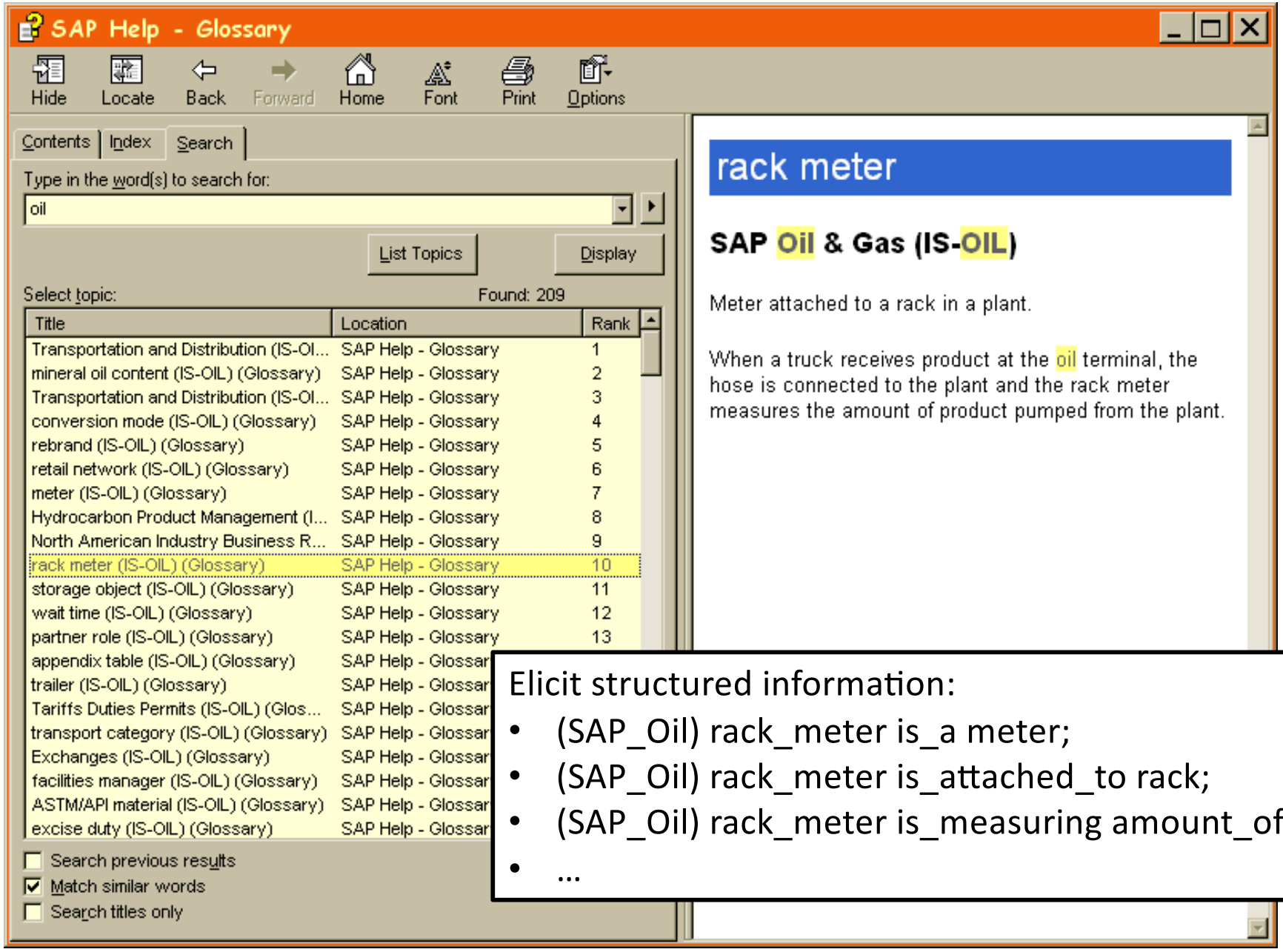
- **Important: Method != Methodology**

- Methods will be covered later in this course.

Where to find ontologies?

- Convert ("align") existing near-ontologies
  - E.g. thesauri, lexicons, glossaries, …
- Engineer them from scratch
  - Top down
  - With/by domain experts
- "Mine" them from suitable sources
  - Bottom up
  - From databases ("lifting")
  - From text (e.g. Web, corpora)

Example from Prof. Dr. Em. Robert Meersman

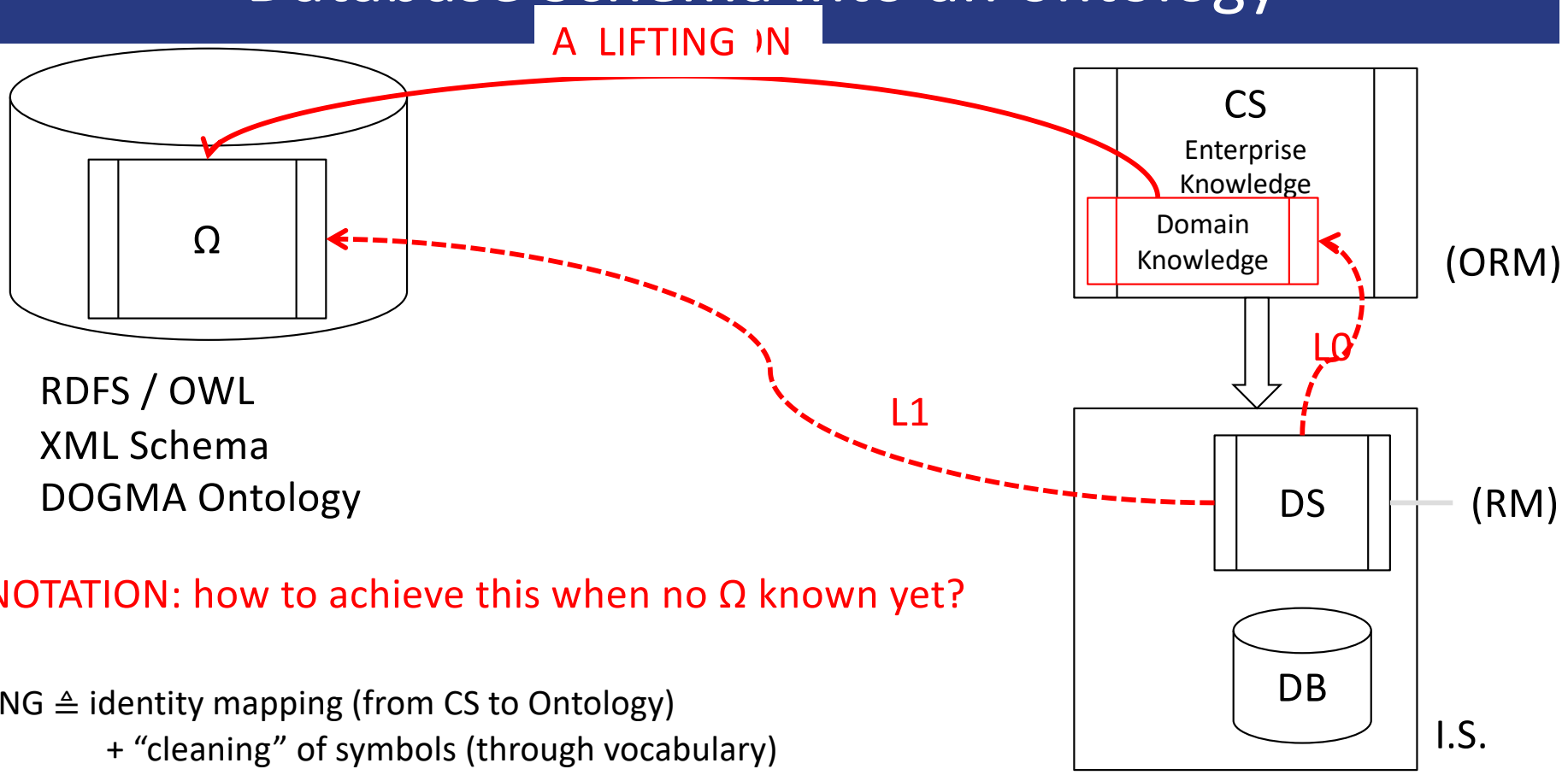


Elicit structured information:

- (SAP_Oil) rack_meter is_a meter;
- (SAP_Oil) rack_meter is_attached_to rack;
- (SAP_Oil) rack_meter is_measuring amount_of_produc
- …

CS = conceptual schema
Ω = ontology

Lifting a Conceptual Schema or Database Schema into an ontology



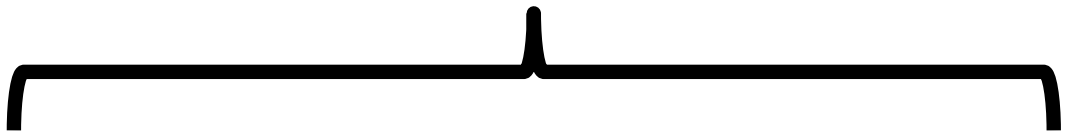
ANNOTATION: how to achieve this when no Ω known yet?

LIFTING ≙ identity mapping (from CS to Ontology)
+ "cleaning" of symbols (through vocabulary)
+ "leaving behind" constraints and enterprise specific knowledge
+ adding of extra domain-specific relations or relationships not present in CS

LO + LIFTING: lifting a DS = stripping of groupings (which may be implementation dependent)
+ mapping of symbols (remove possible encodings)

Groupings

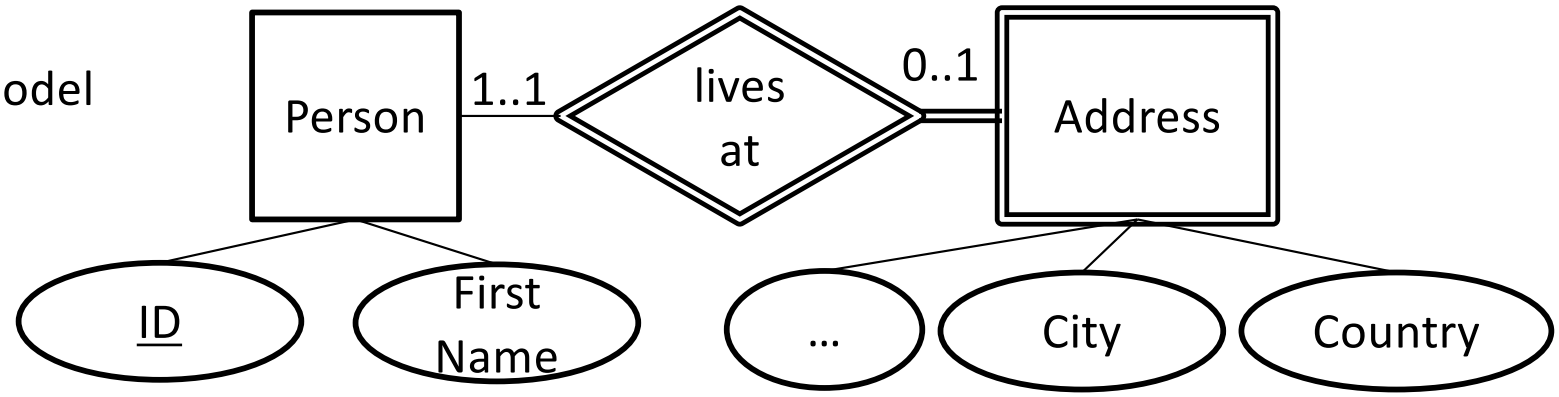
Database Schema



ID (PK)	First_Name	…	ADDRCity *NULLABLE*	ADDRCountry *NULLABLE*
1	Christophe		Ghent	Belgium
2	Julia		Bethune	France

Groupings

Conceptual Model

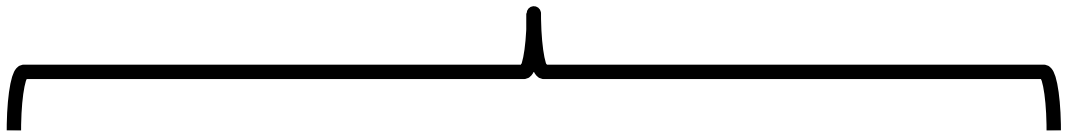


Database Schema

ID (PK)	First_Name	…	ADDRCity *NULLABLE*	ADDRCountry *NULLABLE*
1	Christophe		Ghent	Belgium
2	Julia		Bethune	France

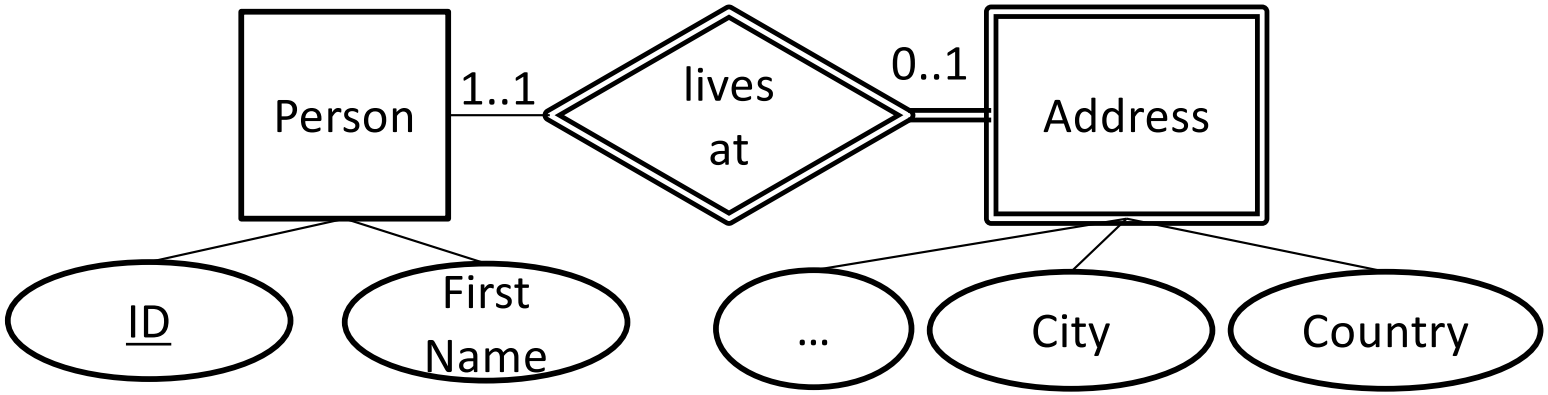
Groupings

Database Schema



ID (PK)	First_Name	…	ADDRCity *NULLABLE*	ADDRCountry *NULLABLE*
1	Christophe		Ghent	Belgium
2	Julia		Bethune	France

What would be removed or changed for an ontology?



- The artificial ID for Person
- Address being identified by people
- (Maybe) each address belongs to exactly one person
- (Maybe) each person has at most one address
- (Maybe) replace “lives at” by something more meaningful. E.g. ”resides at”

By the way …

- What is the difference between a relation and a relationship?
- The latter is an instance of the former.

Source: https://www.instagram.com/p/BtTnIOUhRtG/

Term/Concept complexity

Some terms and concepts are known by “everybody” while, for others, the definition of one term introduces the need to define other terms.

- Tweety
- **Open Market Value** is defined as "the estimated **amount** for which a property should exchange on the date of valuation between a willing buyer and a willing seller in an arm’s-length transaction after proper marketing wherein the parties had each acted knowledgeably, prudently, and without compulsion.” *(IVS 1 - Market Value Basis of Valuation, 7th Ed.)*

Relation(ship) complexity

The definition of a term often requires the specification of relations to other terms, and rules on these relations to constrain their meaning.

- Tweety *is a* Bird
- Christophe *is a* Resident.
  "Your residence status for tax purposes is determined by the number of days that you are present in Ireland in a tax year. You will be resident in Ireland for a tax year in either of the following circumstances: (i) If you spend 183 days or more in Ireland during a tax year or, (ii) If you spend 280 days or more in Ireland over a period of two consecutive tax years, you will be regarded as resident for the second tax year." (source: https://www.citizensinformation.ie/)

Ontology vs. Database and Conceptual Models

- Ontologies are…
  - a priori shared, *yet developed autonomously*
  - independent of application purpose, *yet often obtained from specific contexts in domain*
  - language independent, *yet engineered using (natural) language*
- Translates into difficult new requirements…
  - standardization, openness, extensibility
  - scalability of design and implementation

- **Peter Spyns, Robert Meersman, Mustafa Jarrar: Data Modelling versus Ontology Engineering. SIGMOD Record 31(4): 12-17 (2002)**

Open versus Closed Systems

Closed

- designed for the purpose of one organization
- requirements and functionality known a priori
- data model agreed locally and refers to organizational concepts
- all facts about the domain are already stored; facts not stored are presumed false
- usually cryptically stored in proprietary format (vendor lock-in) only understood by designer

Open

- developed for deployment on Internet
- domain-specific, not application-specific
- users, usage context, applications largely unknown a priori
- ontologies refer to language- and context independent concepts

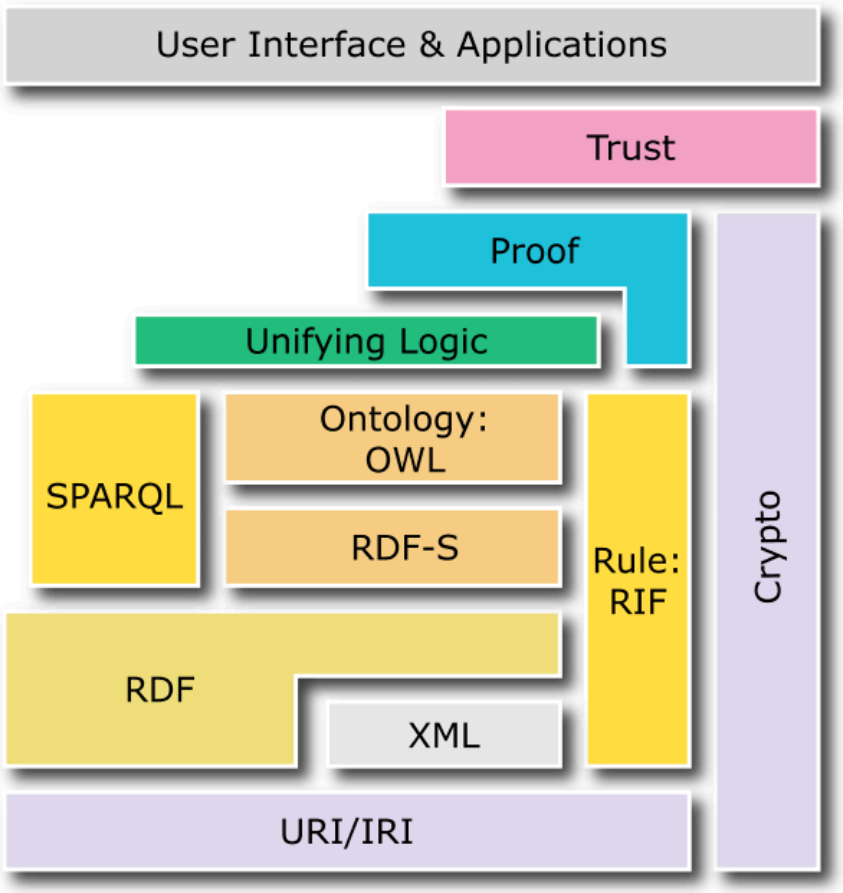
The Semantic Web

- We as humans can interpret the relationships denoted by a hyperlink. This is a more difficult task for machines.

- “The Semantic Web is an extension of the current Web in which information is given well-defined meaning, better enabling computers and people to work in co-operation.” [BLHL01]

- Represent content on the Web in a way that is more easily machine-accessible, and to use intelligent techniques to take advantage of these representations.

The Semantic Web Stack



https://www.w3.org/2007/Talks/0130-sb-W3CTechSemWeb/#(24)

There are many variations

- Some considered RDF on top of XML (which we will see is now outdated).
- Some consider rules (RIF/SWRL) on top of RDF or RDFS

References

[BLHL01] T. Berners-Lee, J. Hendler, and O. Lassila. The Semantic Web. Scientific American, 284(5):34-43, 2001

[DLM08] P. De Leenheer and T. Mens. Ontology evolution: State of the art and future directions. In Hepp et al. [HDdS08], pages 131–176.

[GPFLC03] A. Gómez-Pérez, M. Fernández-López, and Ó. Corcho. Ontological Engineering with examples from the areas of Knowledge Management, e-Commerce and the Semantic Web. Springer, 2003.

[HDdS08] M. Hepp, P. De Leenheer, A. de Moor, and Y. Sure, editors. Ontology Management, Semantic Web, Semantic Web Services, and Business Applications, volume 7 of Semantic Web And Beyond Computing for Human Experience. Springer, 2008.

[Gru95] T. Gruber. Toward principles for the design of ontologies used for knowledge sharing? Int. J. Hum.-Comput. Stud., 43(5-6):907–928, 1995.

[Stu98] R. Studer, R. Benjamins, and D. Fensel. Knowledge engineering: Principles and methods. Data & Knowledge Engineering, 25(1–2):161–198, 1998.