Optique

BootOX:
Practical Mapping of RDBs to OWL 2

Ernesto Jiménez-Ruiz[1], Evgeny Kharlamov[1], Dmitriy Zheleznyakov[1], Ian Horrocks[1], Christoph Pinkel, Martin G. Skjæveland, Evgenij Thorstensen, and Jose Mora

[1] University of Oxford

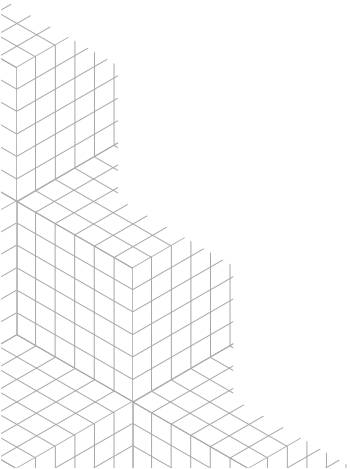
International Semantic Web Conference (IU&S)

October 14, 2015

Outline

EU project Optique

- Aims at facilitating **scalable end-user access to big data** in the oil and gas industry.
- Advocates for an **OBDA approach**
  - ontology provides a virtual access to the data
  - mappings connect the ontology with the data source.
- Focuses around two demanding use cases provided by the industry partners **Siemens** and **Statoil**



EU project Optique: current limitations

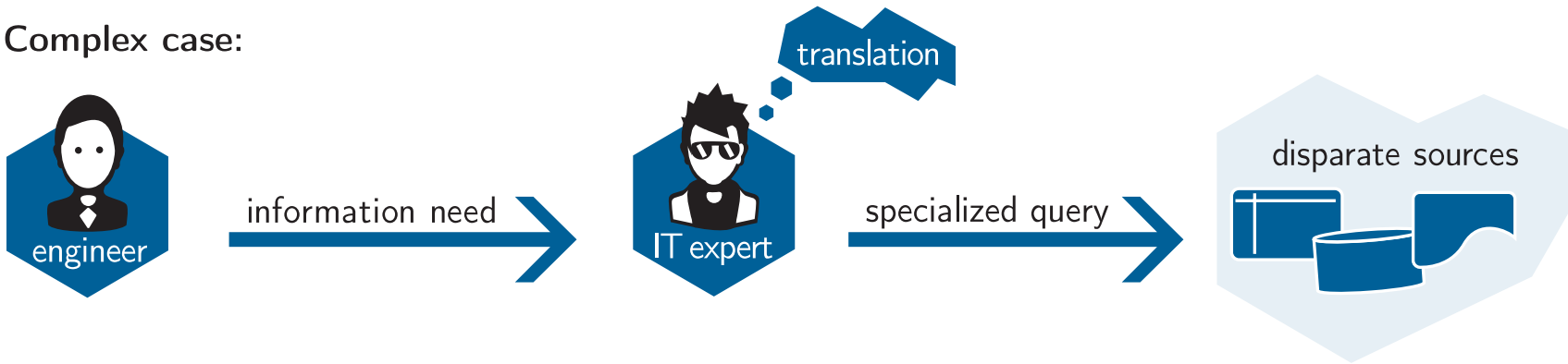
Optique

Simple case:

EU project Optique: current limitations

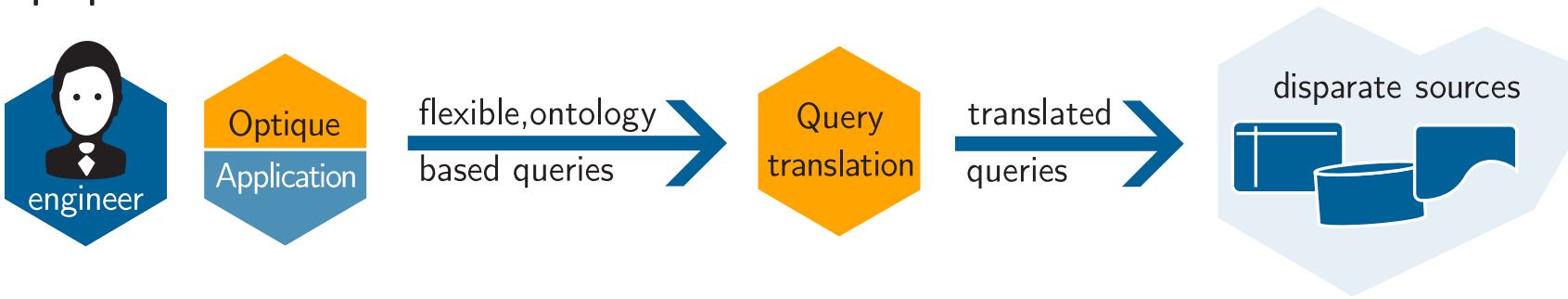
Optique

Complex case:

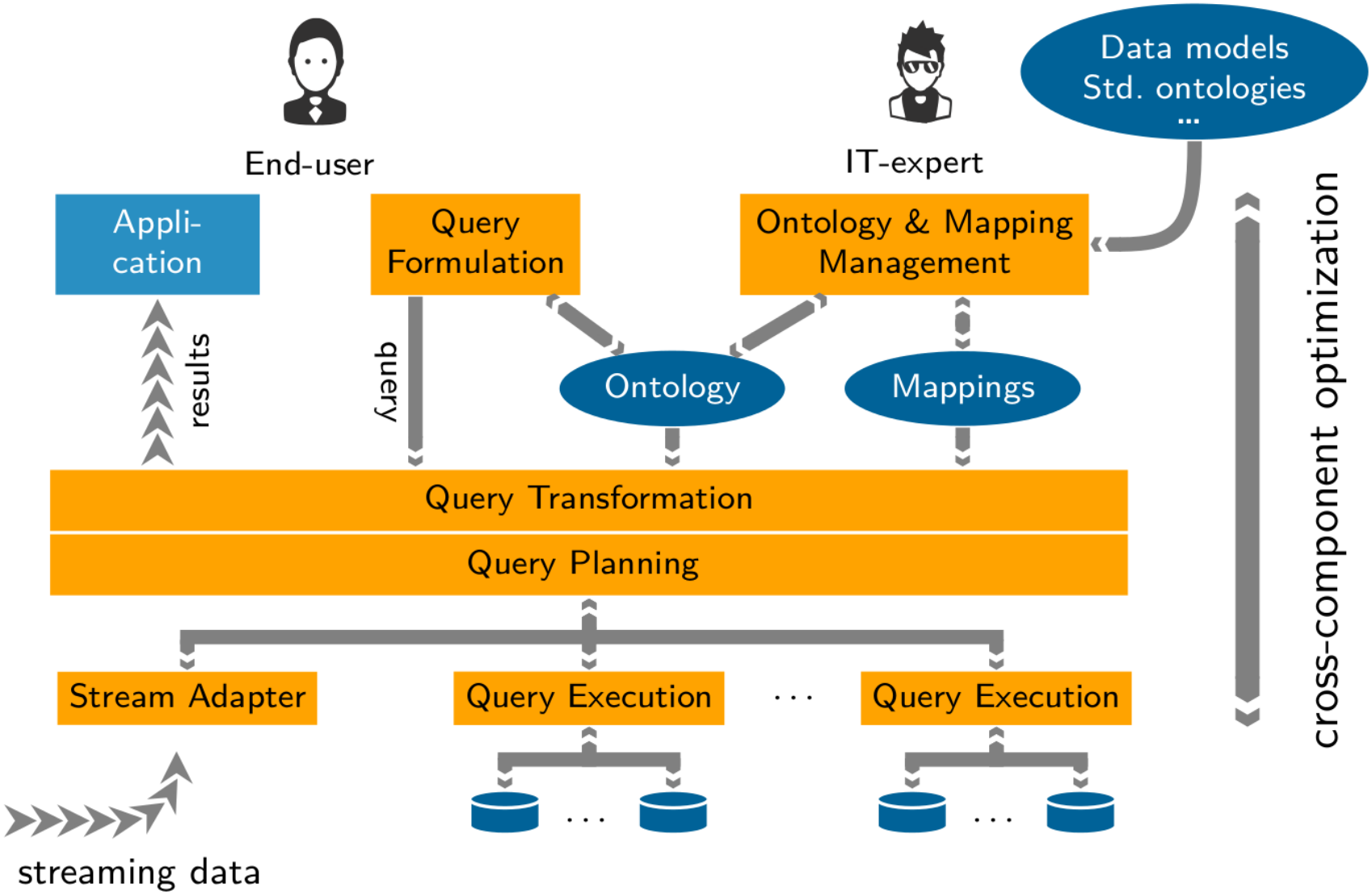


EU project Optique: current limitations

Optique solution



EU project Optique: architecture



Outline

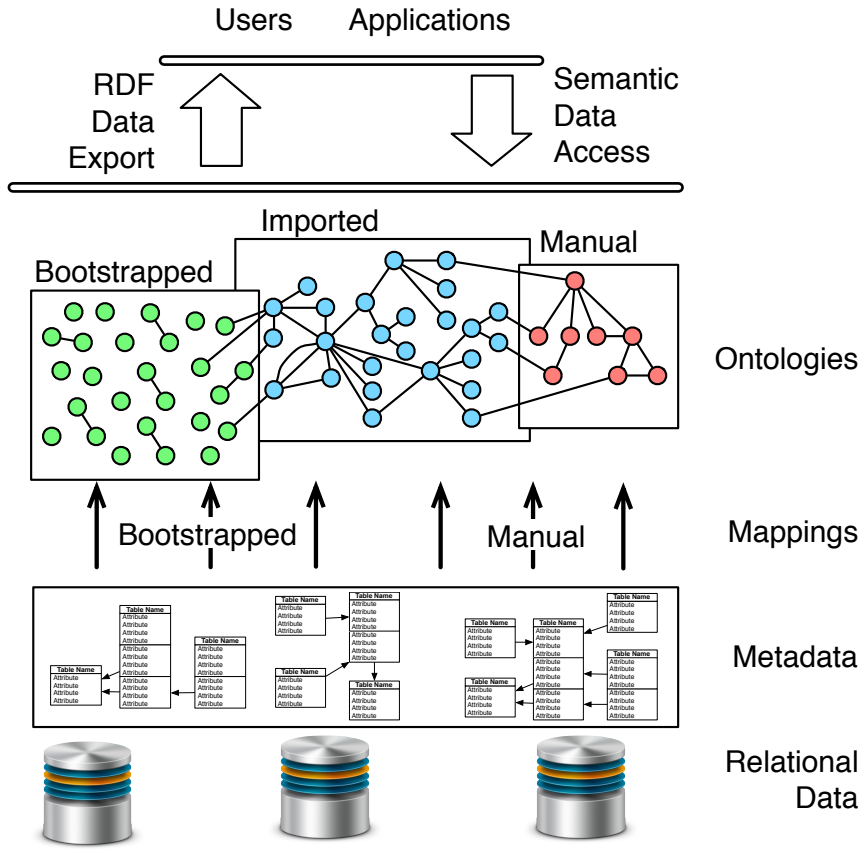
Exposing Relational Data via Ontologies

Optique

- Three ingredients
  - **Onto vocabulary:** defines the terms from the RDBs domain
  - **Axioms:** describe the structure of the RDBs domain
  - **Mappings:** relate ontology terms to queries over the RDB

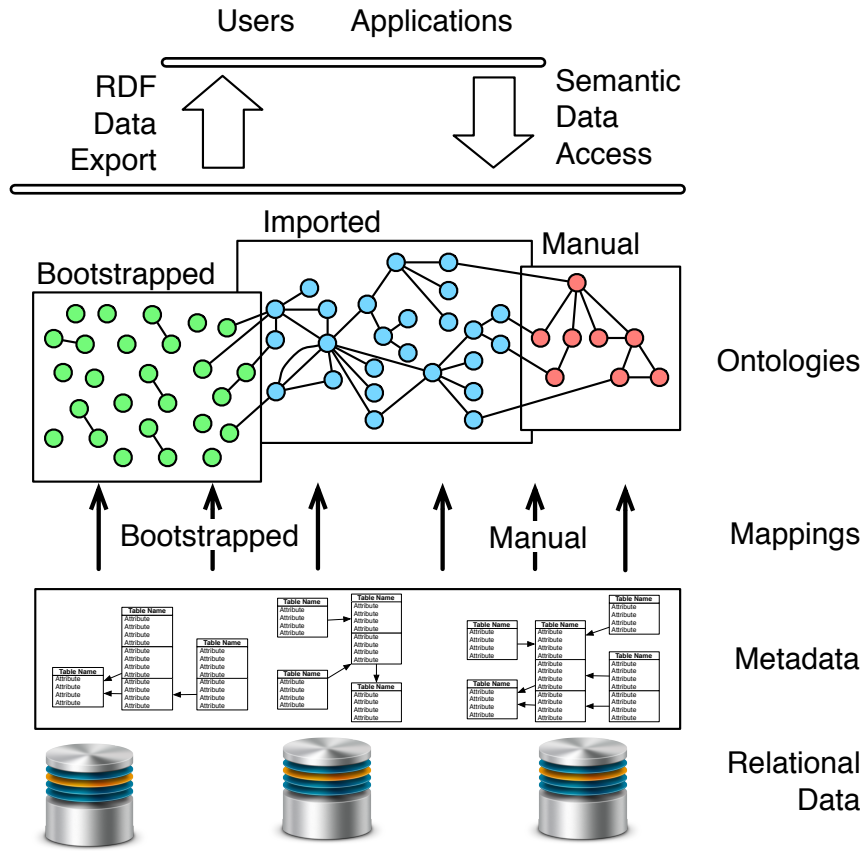
Mappings

$$Class(f_o(x)) \rightsquigarrow \texttt{SQL}(x)$$
$$objectProperty(f_o(x), f_o(y)) \rightsquigarrow \texttt{SQL}(x, y)$$
$$dataProperty(f_o(x), f_v(y)) \rightsquigarrow \texttt{SQL}(x, y)$$



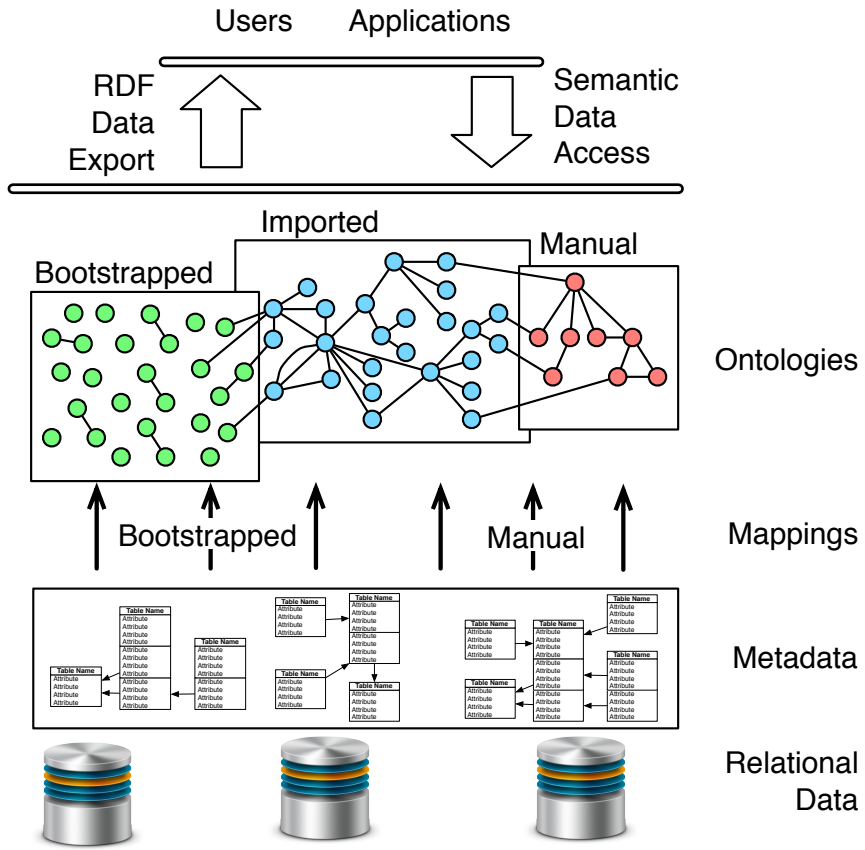
RDBs via Ontologies: Motivation

- **Virtual exposure of data** (OBDA) as in Optique
  - End-users friendly access to unfriendly relational data
  - Pay as you go data integration
- **Data Export**
  - Transformation into a standard/clear schema (e.g., unification)
  - Easy to exchange data (over the Web)



Existing Approaches: 5 Dimensions

- Level of automation
  - manual, semi-automatic, automatic
- Type of mappings
  - complex or direct mappings
- Language of mappings & ontology
  - R2RML, RDF, OWL 2
- Reuse of external vocabularies
  - domain ontologies or thesauri
- Purpose
  - OBDA, database integration, data export

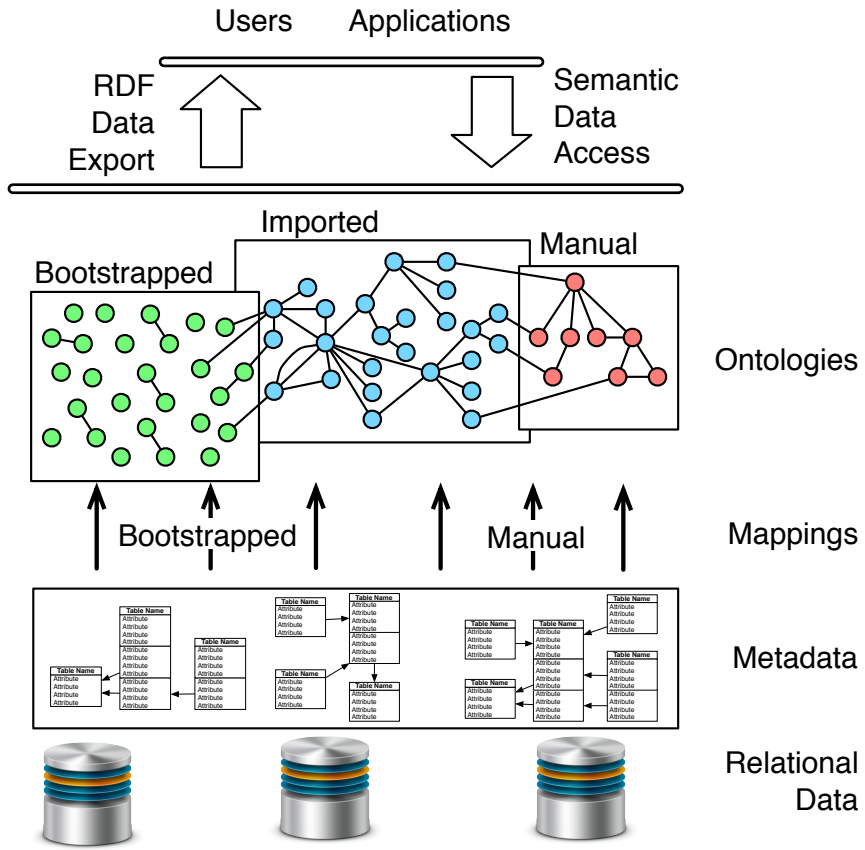


Our Goal: Bootstrapper with flexibility in 5 dimensions

Existing Approaches: 5 Dimensions

Optique

- Level of automation
  - manual, semi-automatic, automatic
- Type of mappings
  - complex or direct mappings
- Language of mappings & ontology
  - R2RML, RDF, OWL 2
- Reuse of external vocabularies
  - domain ontologies or thesauri
- Purpose
  - OBDA, database integration, data export



Our Goal: Bootstrapper with flexibility in 5 dimensions

Limitations of Current Approaches

- Other ontology languages (e.g. DLR-Lite, RDFs, F-Logic)
- Own native language before R2RML (e.g. Mastro)
- Concrete OWL 2 expressiveness (sometimes even outside OWL 2) and purpose
- Neglect the logical impact when reusing external ontologies
- Not available tool or integrated in commercial system

Outline

- **Main features**:
  - R2RML mappings based on the W3C directives
  - Schema and data driven bootstrapping
  - Propagation of database constraints customizable into...
    - OWL 2 expressiveness (OWL 2 and profiles)
    - Local vs. global constraints
    - Two attribute naming schemata
  - Extraction of new classes and properties
  - Importing of an external ontology
  - Creation of provenance mappings

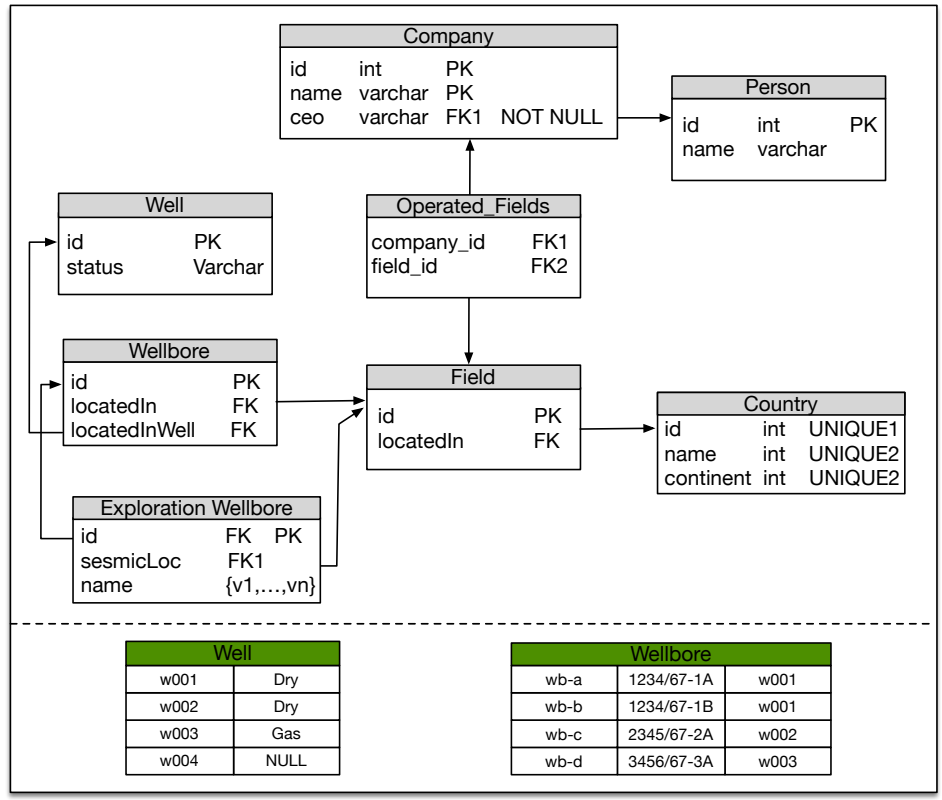
Mapping Generation in BootOX

- **R2RML mappings:**
  - W3C direct mapping guidelines (*schema driven*)
    - To a bootstrapped ontology vocabulary
    - To a given ontology vocabulary
  - Mappings beyond direct ones (*data driven*)
    - Clusters of tuples
    - Joinable tuples
  - Provenance mappings (*schema driven*)
    - URI level
    - Triple level
    - Graph level

Vocabulary Generation in BootOX

Optique

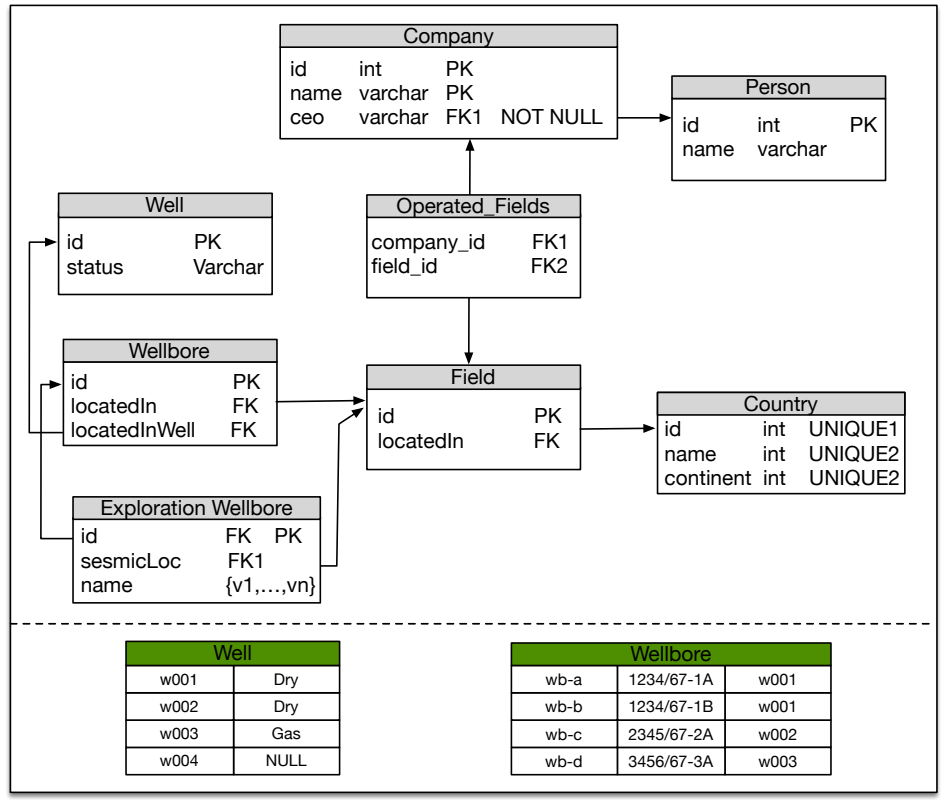
- **W3C directives**
  - Tables → classes
  - Foreign Keys → object properties
  - Data columns → data properties
  - Binary tables → fresh object properties
- **Attribute namming schema:**
  - Unique names (e.g. Person.name)
  - Reusable names (e.g. name)



Vocabulary Generation in BootOX

Optique

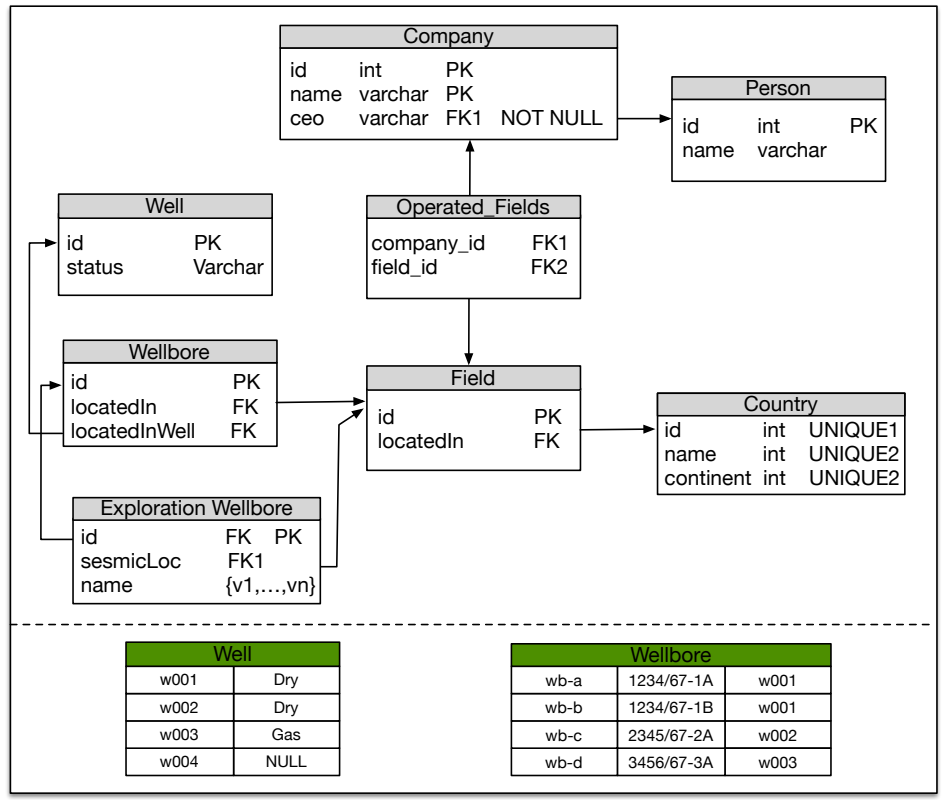
- **W3C directives**
  - Tables → classes
  - Foreign Keys → object properties
  - Data columns → data properties
  - Binary tables → fresh object properties
- **Attribute namming schema:**
  - Unique names (e.g. Person.name)
  - Reusable names (e.g. name)



Axiom Generation in BootOX

Optique

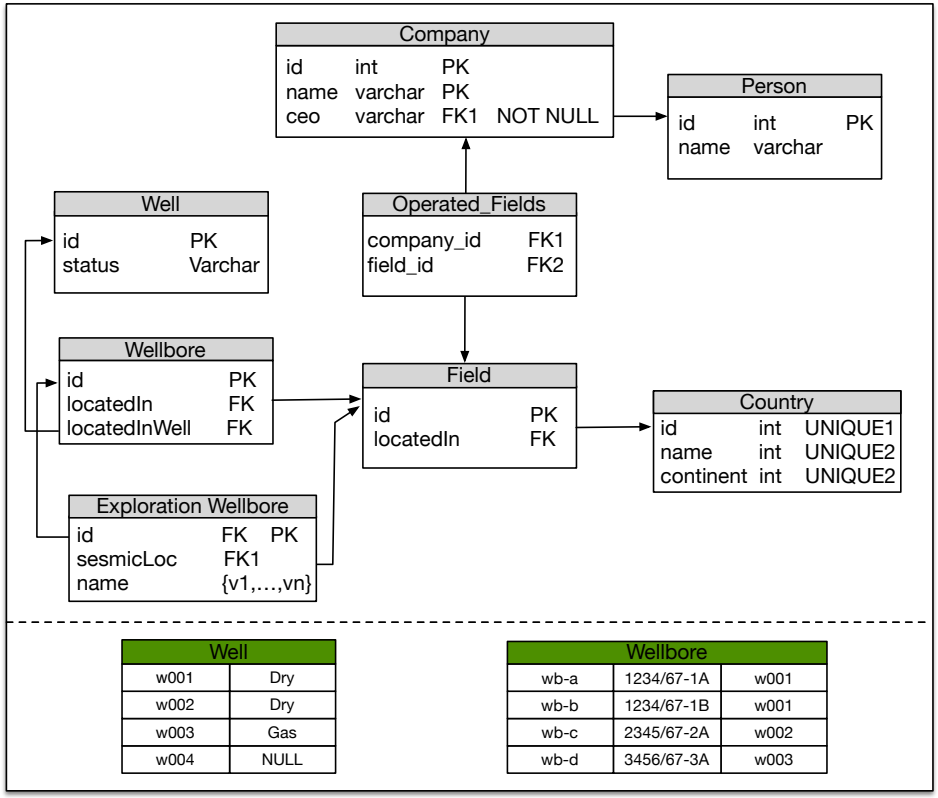
- **OWL 2 expressiveness**
  - OWL 2 QL (e.g. OBDA/Optique)
  - OWL 2 EL (e.g. EOLO)
  - OWL 2 RL (e.g. RDFox)
  - OWL 2 (e.g. PAGOdA)



Axiom Generation in BootOX

Optique

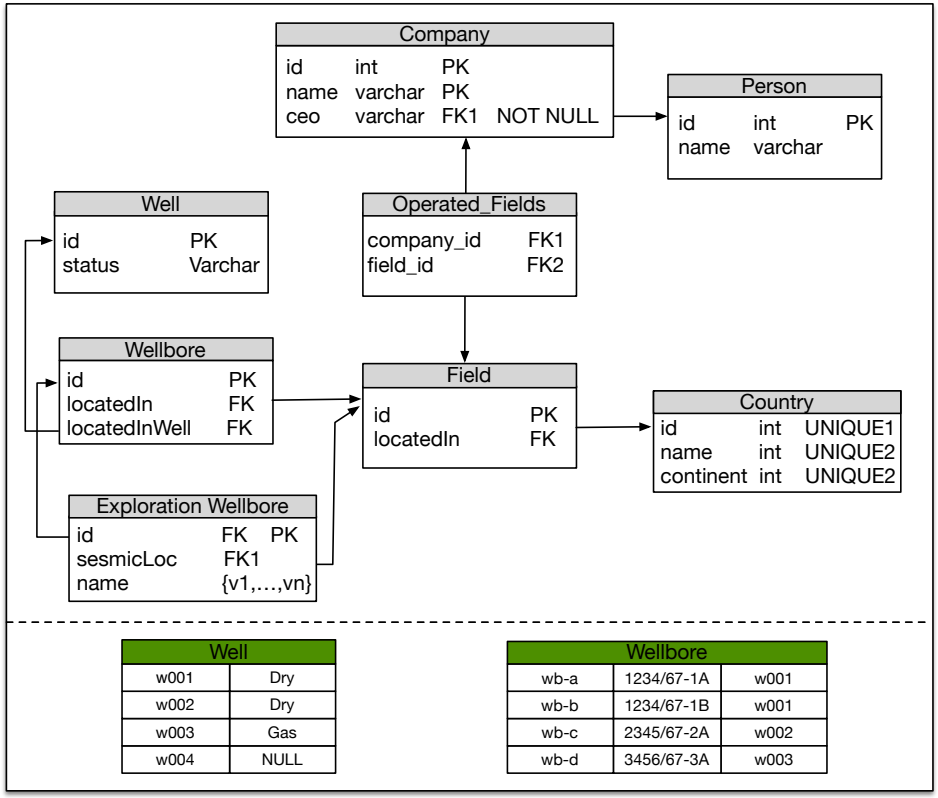
- Unique constraints
  - Person **HasKey:** id (OWL 2 RL/EL)
- Global onto. constraints
  - Person.name **Functional** (OWL 2 RL)
  - name **Domain:** Person (all profiles)
  - Person.name **Range:** xsd:string (all profiles)
- Local onto. constraints
  - Person **subclassOf:** name **some** xsd:string (OWL 2 QL/EL)
  - Person **subclassOf:** name **only** xsd:string (OWL 2 RL)
  - Person **subclassOf:** name **exactly 1** xsd:string (OWL 2)



Axiom Generation in BootOX

Optique

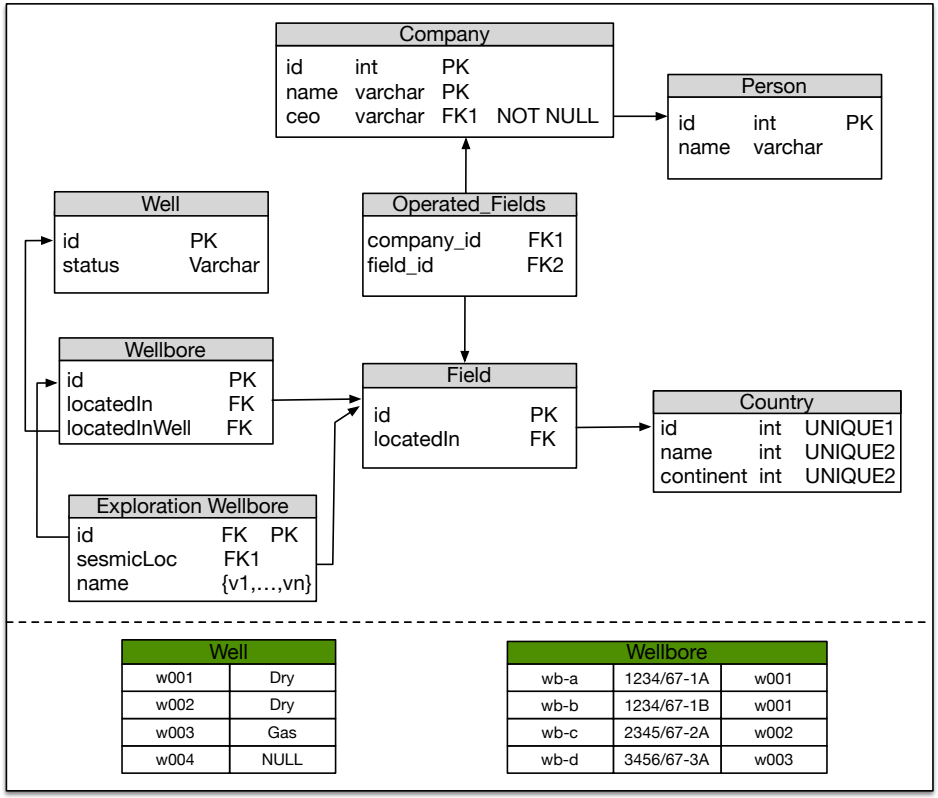
- Unique constraints
  - Person **HasKey:** id (OWL 2 RL/EL)
- Global onto. constraints
  - Person.name **Functional** (OWL 2 RL)
  - name **Domain:** Person (all profiles)
  - Person.name **Range:** xsd:string (all profiles)
- Local onto. constraints
  - Person **subclassOf:** name **some** xsd:string (OWL 2 QL/EL)
  - Person **subclassOf:** name **only** xsd:string (OWL 2 RL)
  - Person **subclassOf:** name **exactly 1** xsd:string (OWL 2)



Axiom Generation in BootOX

Optique

- Unique constraints
  - Person **HasKey:** id (OWL 2 RL/EL)
- Global onto. constraints
  - Person.name **Functional** (OWL 2 RL)
  - name **Domain:** Person (all profiles)
  - Person.name **Range:** xsd:string (all profiles)
- Local onto. constraints
  - Person **subclassOf:** name **some** xsd:string (OWL 2 QL/EL)
  - Person **subclassOf:** name **only** xsd:string (OWL 2 RL)
  - Person **subclassOf:** name **exactly 1** xsd:string (OWL 2)



Datatypes in BootOX

- Clear mapping between SQL and XML schema datatypes
- Not all XML Schema datatypes are included in OWL 2
  - *xsd:date* not in OWL 2
  - *xsd:boolean* and *xsd:double* not in OWL 2 QL/EL
- Value spaces of primitive datatypes are disjoint (e.g. *xsd:double* and *xsd:decimal*)

× Problems when materalizing the data or using the ontology for virtual access to data

✓ Solution: *rdfs:Literal*

Datatypes in BootOX

- Clear mapping between SQL and XML schema datatypes
- Not all XML Schema datatypes are included in OWL 2
  - *xsd:date* not in OWL 2
  - *xsd:boolean* and *xsd:double* not in OWL 2 QL/EL
- Value spaces of primitive datatypes are disjoint (e.g. *xsd:double* and *xsd:decimal*)

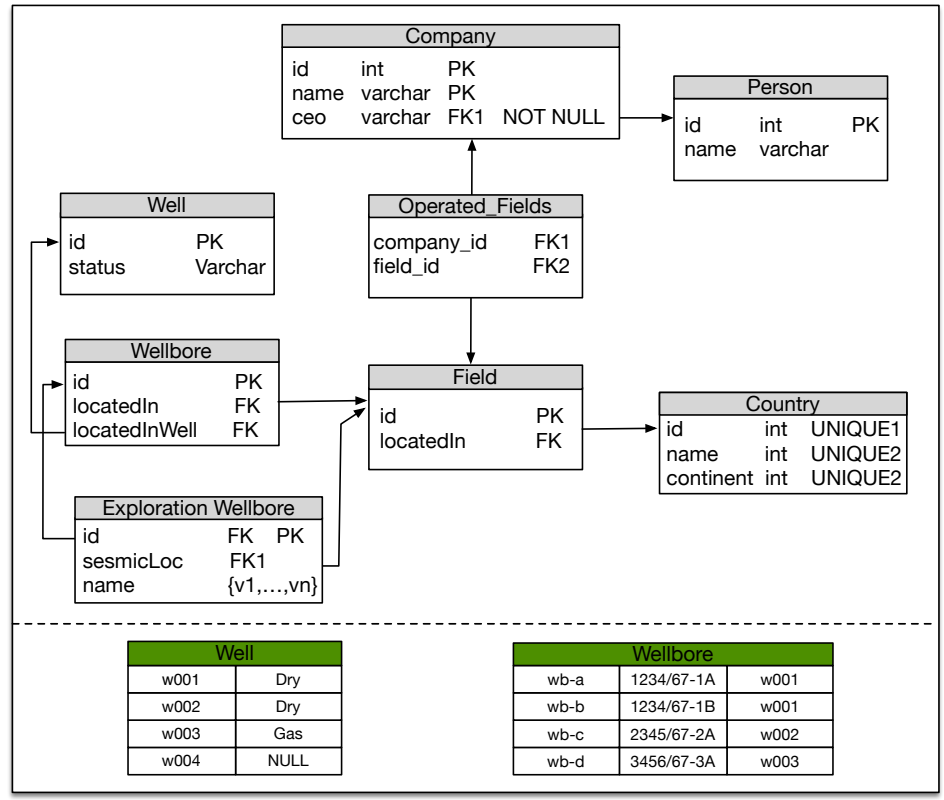
✗ Problems when materalizing the data or using the ontology for virtual access to data

✓ Solution: *rdfs:Literal*

Taxonomy Generation in BootOX

Optique

- **Data driven**
  - Clusters of tuples
  - Joinable tuples
  - e.g. Well_Dry
    SubclassOf: Well
- **Schema driven**
  - A single-column Foreign Key and Primary key
  - e.g. Exploration_Welllbore
    SubclassOf: Wellbore



Taxonomy Generation in BootOX

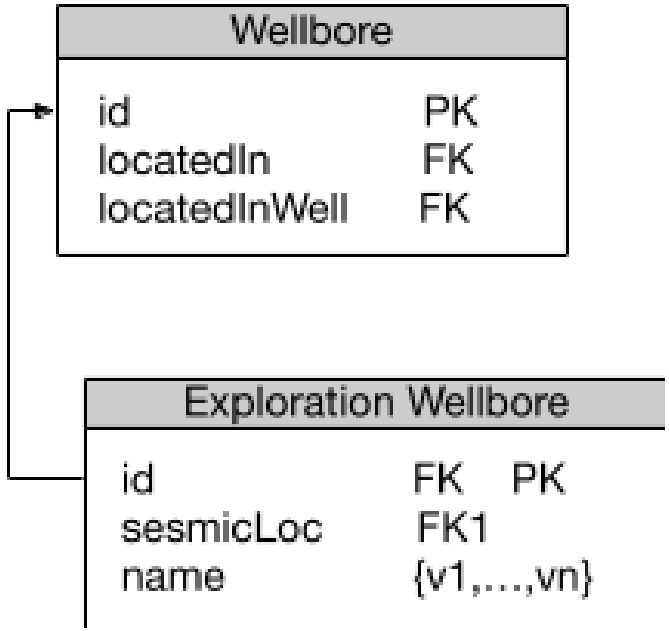
- **Data driven**
  - Clusters of tuples
  - Joinable tuples
  - e.g. Well_Dry
    SubclassOf: Well
- **Schema driven**
  - A single-column Foreign Key and Primary key
  - e.g. Exploration_Welllbore
    SubclassOf: Wellbore

Well	
w001	Dry
w002	Dry
w003	Gas
w004	NULL

Wellbore		
wb-a	1234/67-1A	w001
wb-b	1234/67-1B	w001
wb-c	2345/67-2A	w002
wb-d	3456/67-3A	w003

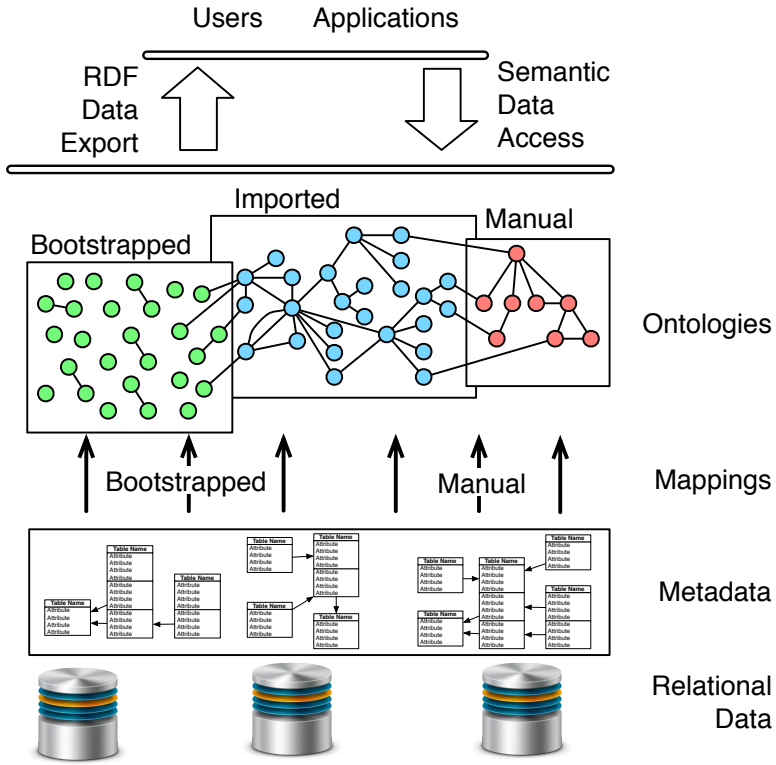
Taxonomy Generation in BootOX

- **Data driven**
  - Clusters of tuples
  - Joinable tuples
  - e.g. Well_Dry
    SubclassOf: Well
- **Schema driven**
  - A single-column Foreign Key and Primary key
  - e.g. Exploration_Welllbore
    SubclassOf: Wellbore



Importing of a Domain Ontology

- Based on the **LogMap** system
  - Ontology alignment system
  - Discovers OWL 2 equivalence and sub-class(property) axioms
  - Implements logical techniques to minimize undesired consequences in the integration

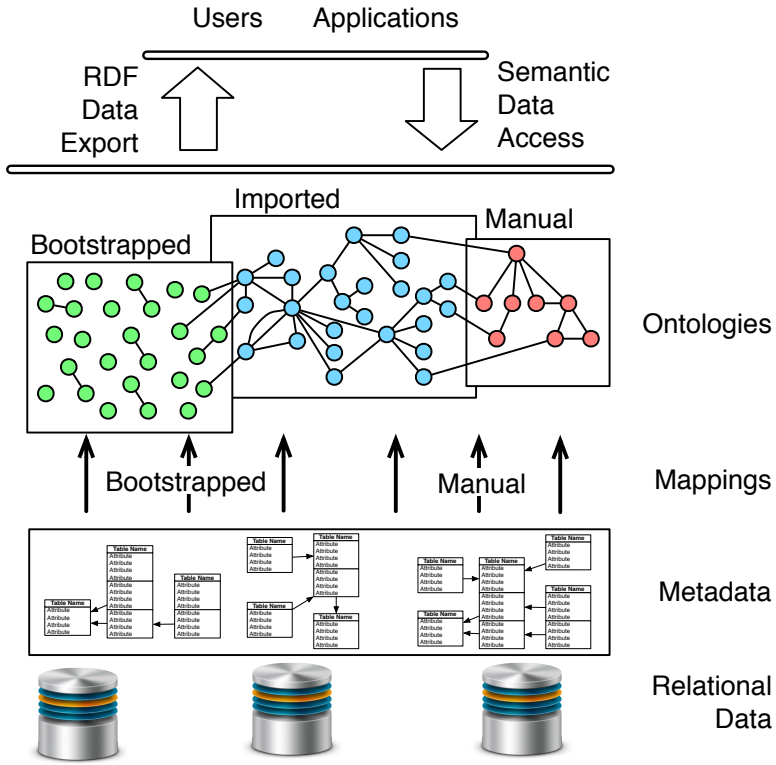


- Large-scale Interactive Ontology Matching: Algorithms and Implementation. ECAI 2012
- Detecting and Correcting Conservativity Principle Violations in Ontology-to-Ontology Mappings. ISWC 2014

Importing of a Domain Ontology

Optique

- Based on the **LogMap** system
  - Ontology alignment system
  - Discovers OWL 2 equivalence and sub-class(property) axioms
  - Implements logical techniques to minimize undesired consequences in the integration



- **Large-scale Interactive Ontology Matching: Algorithms and Implementation. ECAI 2012**
- **Detecting and Correcting Conservativity Principle Violations in Ontology-to-Ontology Mappings. ISWC 2014**

Outline

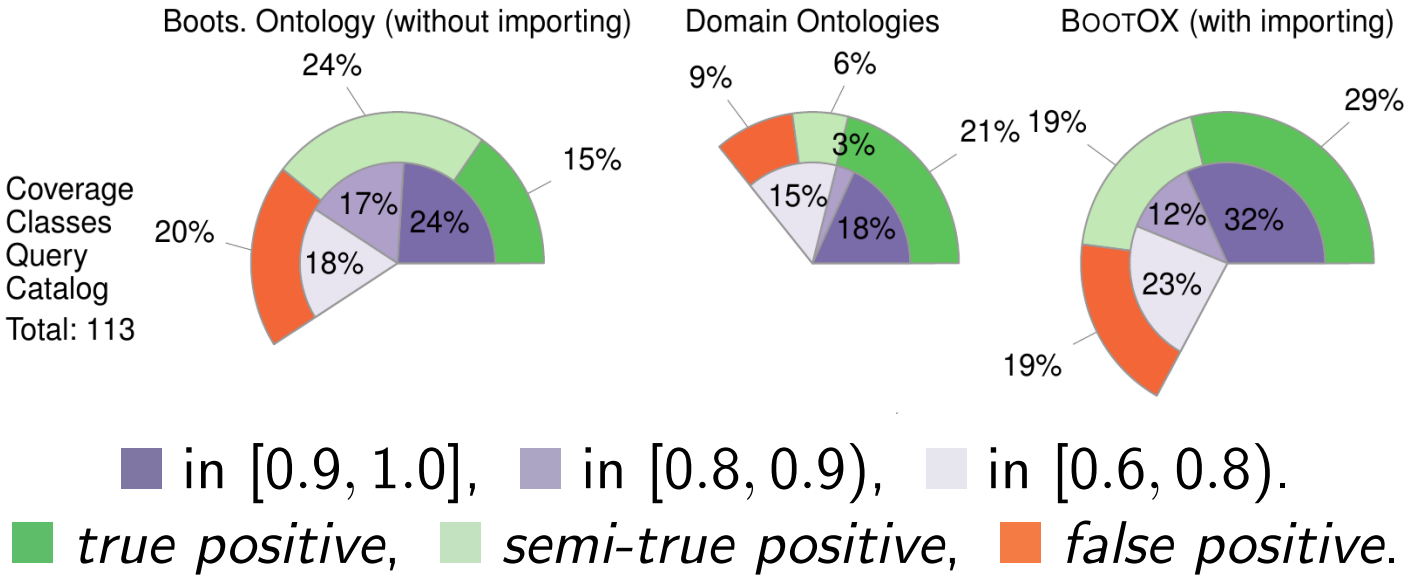
Conducted evaluation

- Ability of formulating queries with the bootstrapped bocabulary
  - **Query coverage in Optique**

- Ability of enabling the answering of queries in a controlled scenarios
  - **Evaluation with RODI benchmark**
  - IncMap, MIRROR, BootOX and -ontop-

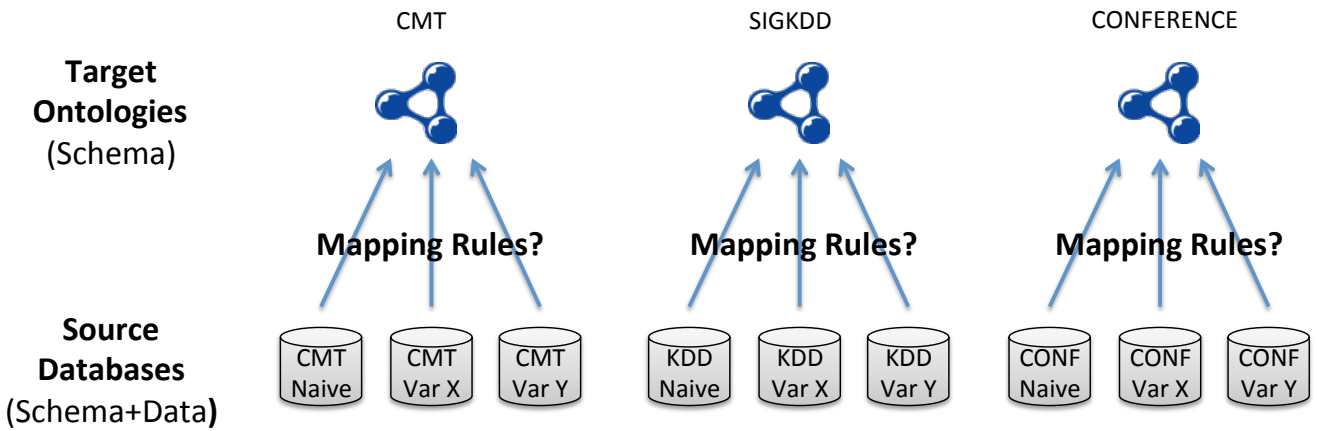
Query coverage in Statoil

- ≈ 67 % of query catalogue terms covered (>60 % confidence) by aligned ontology
  - 29 % true positives
  - 19 % semi-true positives
  - 19 % false positives



Evaluation with RODI benchmark

Optique



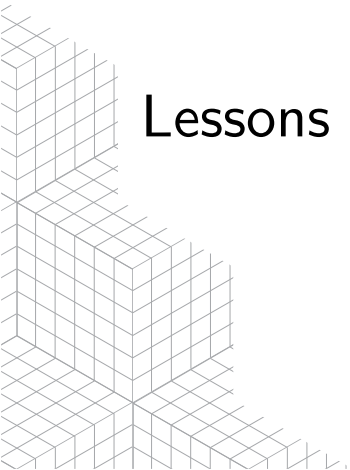
RODI: A Benchmark for Relational-to-Ontology Data Integration. ESWC 2015.

Evaluation with RODI benchmark

Scenario	IncMap	MIRROR	-ontop-	BOOTOX
	Adjusted naming			
CMT	**0.5**	0.28	0.39	0.39
CONFERENCE	0.26	0.27	**0.37**	**0.37**
SIGKDD	0.21	0.3	**0.45**	**0.45**
	Cleaned hierarchies			
CMT	**0.44**	0.17	0.28	0.28
CONFERENCE	0.16	0.23	**0.3**	**0.3**
SIGKDD	0.11	0.11	**0.16**	**0.16**
	Combined case			
SIGKDD	0.05	0.11	**0.16**	**0.16**
	Missing FKs			
CONFERENCE	0.03	**0.17**	-	**0.17**
	Denormalised			
CMT	0.22	0.22	**0.28**	**0.28**

RODI: A Benchmark for Relational-to-Ontology Data Integration. ESWC 2015.

Outline



Lessons learnt

- **Feedback from use cases...**
  - ✓ Good as a first approximation of the ontology and mappings
  - ✓ Ontology and mappings ready to use for one of the Statoil data sources
  - × For the largest Statoil datasources, the solution is far from perfect
  - × Ontology close to the original database
  - × Large amount of ontology entities and R2RML mappings
  - ✓ Implementation of an incremental/interactive bootstrapping
- **Future work...**
  - Stand-alone verion of BootOX
  - New ways to exploit the data and schema (e.g. views)

Lessons learnt

- **Feedback from use cases...**
  - ✓ Good as a first approximation of the ontology and mappings
  - ✓ Ontology and mappings ready to use for one of the Statoil data sources
  - × For the largest Statoil datasources, the solution is far from perfect
  - × Ontology close to the original database
  - × Large amount of ontology entities and R2RML mappings
  - ✓ Implementation of an incremental/interactive bootstrapping
- **Future work...**
  - Stand-alone verion of BootOX
  - New ways to exploit the data and schema (e.g. views)

Questions?

Thank you for your attention

- **Optique project:**
  `http://www.optique-project.eu/`
- **BootOX Project:**
  `http://www.cs.ox.ac.uk/isg/tools/BootOX/`
- `ernesto.jimenez.ruiz@gmail.com`
- **Acknowledgements**:
  - Seventh Framework Program (FP7) of the European Commission under Grant Agreement 318338, “Optique”