

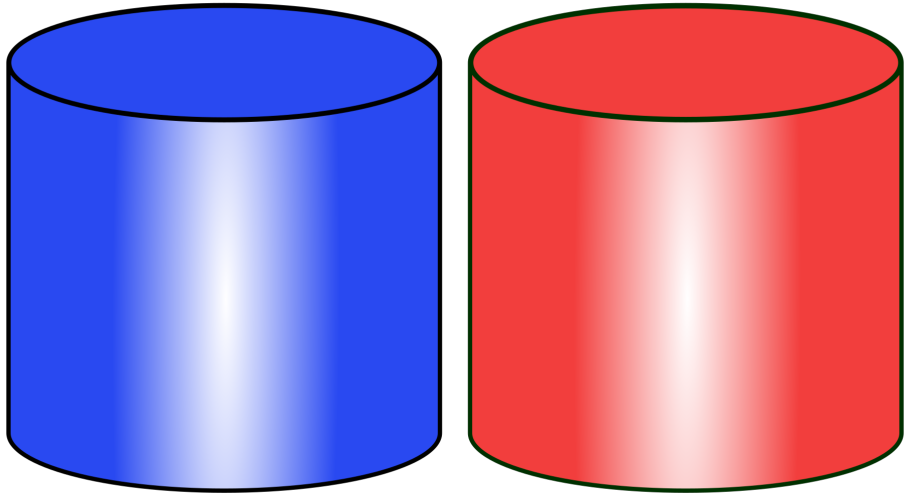
Obi-Wan for efficient and expressive data integration

Maxime Buron

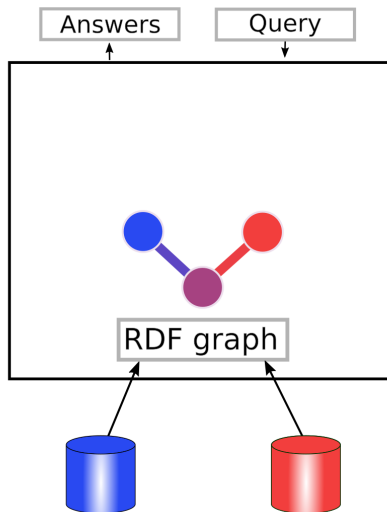
GR Big Data

Motivations

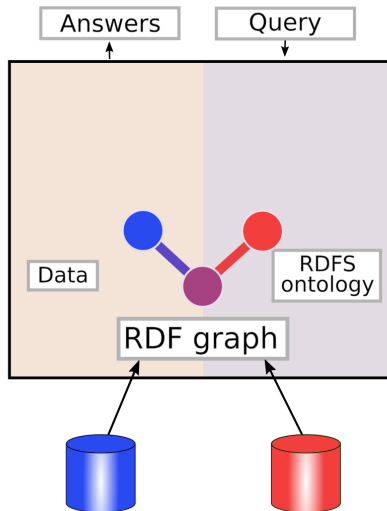
Heterogeneous data sources

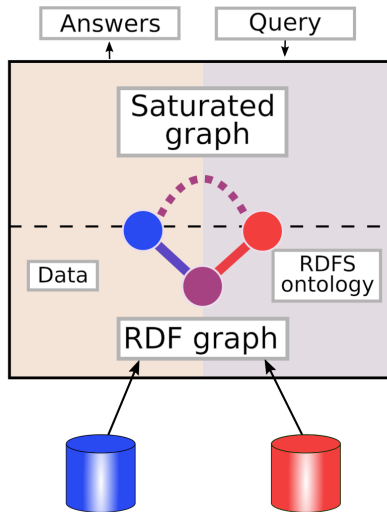


RDF data model for data integration



RDFS ontology





Preliminaries

RDF Triple

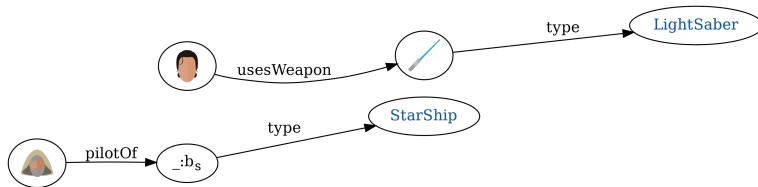
An RDF triple contains three values among:

- IRIs,
- blank nodes,
- literals.



RDF Graph: Data and RDFS Ontology

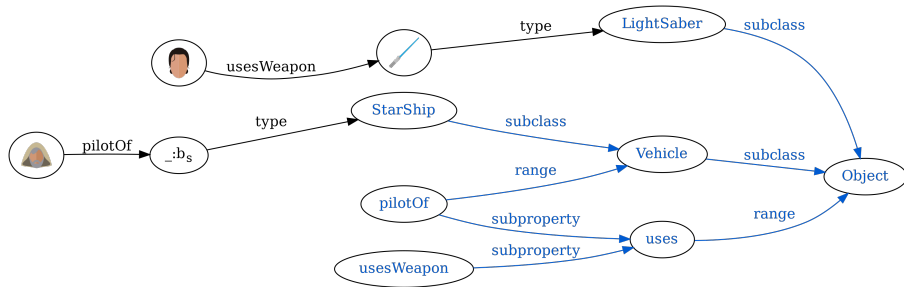
Data triples of an RDF graph G .



RDF Graph: Data and RDFS Ontology

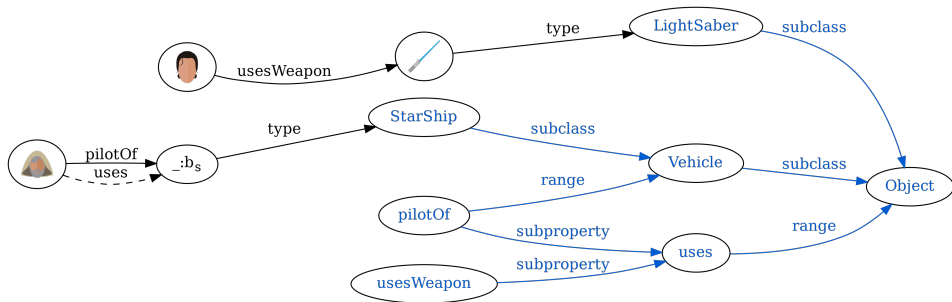
The RDFS triples use the properties:

- :subclass,
- :subproperty,
- :domain,
- :range.



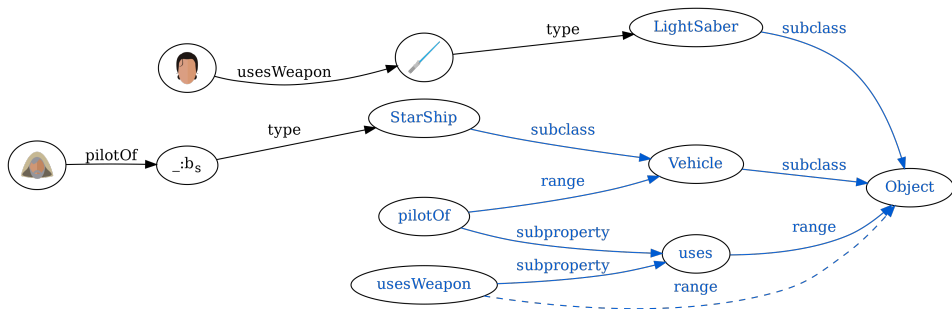
Data Entailment Rules $\mathcal{R}_{\text{data}}$

$$\mathcal{R}_{\text{data}} = \{ (p_a, :subproperty, p_b), (s, p_a, o) \rightarrow (s, p_b, o) \quad \dots \}$$



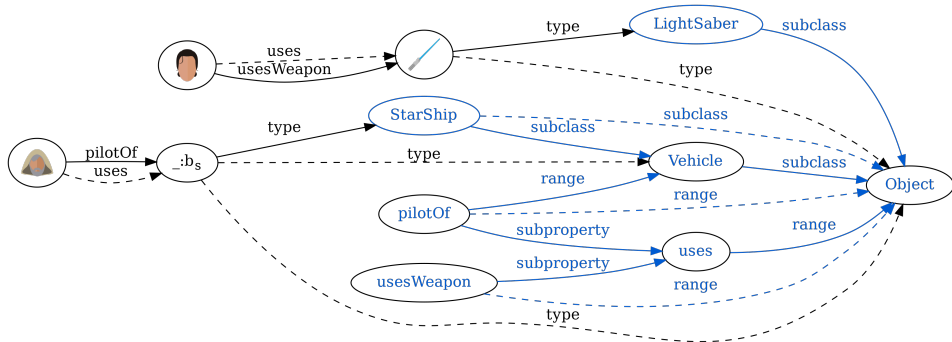
Ontological Entailment Rules $\mathcal{R}_{\text{onto}}$

$$\mathcal{R}_{\text{onto}} = \{ (p, :subproperty, p_a), (p_a, :range, o) \rightarrow (p, :range, o) \quad \dots \}$$



Full Saturation of the Graph w.r.t. $\mathcal{R}_{\text{data}}$ and $\mathcal{R}_{\text{onto}}$

The **full saturation** of G is $G^{\mathcal{R}_{\text{onto}} \cup \mathcal{R}_{\text{data}}}$:



Basic Graph Pattern Queries

We consider queries over the data **and the ontology**:

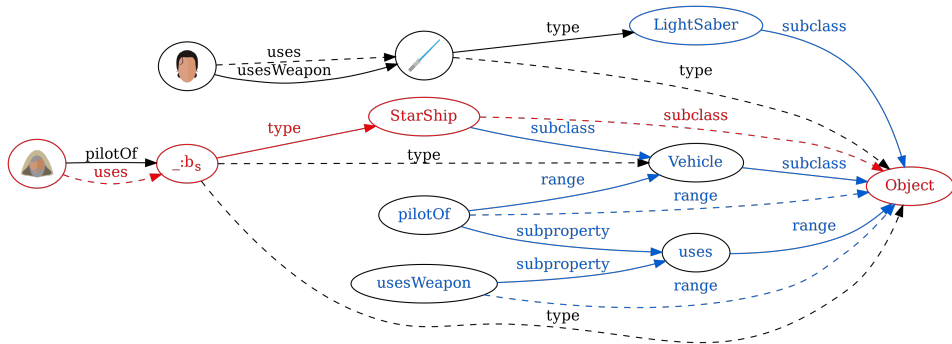
"Find anyone who uses something that is a kind of object"

$$q(x, y) \leftarrow (x, :uses, z), (z, :type, y), (y, :subclass, :Object)$$

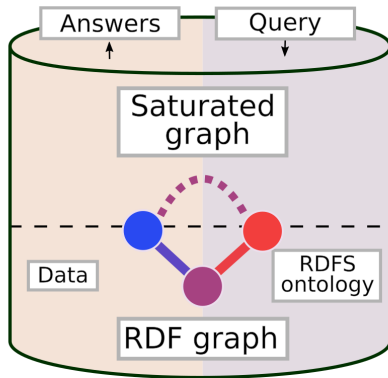
Saturation-Based Query Answering

$$q(x, y) \leftarrow (x, :uses, z), (z, :type, y), (y, :subclass, :Object)$$

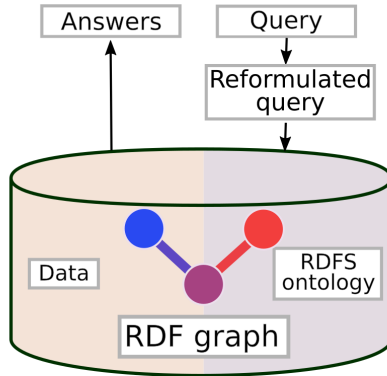
$$q(G^{\mathcal{R}_{\text{data}} \cup \mathcal{R}_{\text{onto}}}) = \begin{aligned} &\langle \text{👤}, :\text{LightSaber} \rangle \\ &\langle \text{🚗}, :\text{Vehicle} \rangle \\ &\langle \text{🚀}, :\text{StarShip} \rangle \end{aligned}$$



Materialization and saturation-based query answering (MAT)

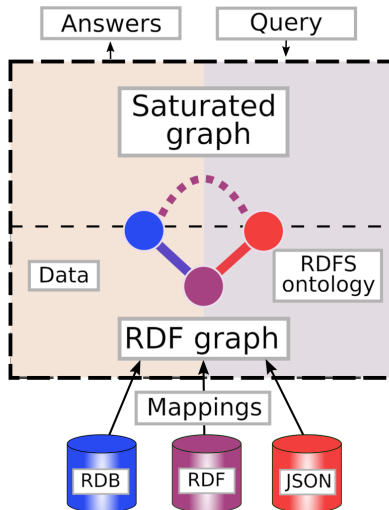


Materialization and reformulation-based query answering



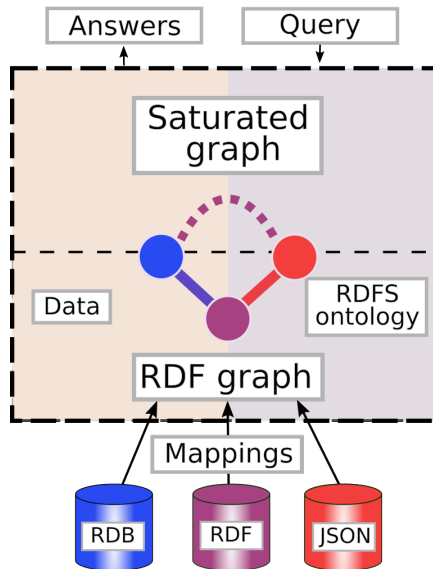
Obi-Wan : an RDF integration system

Ontology-Based Data Access

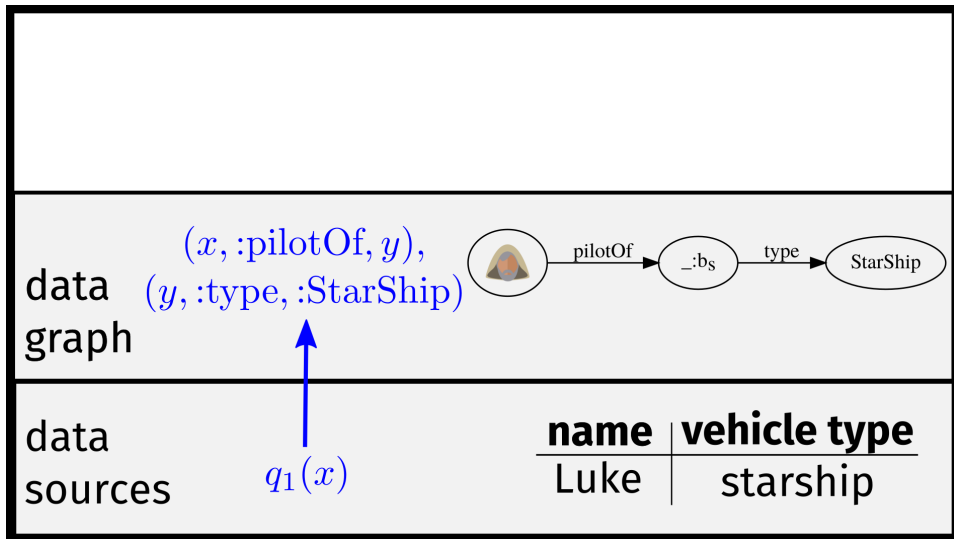


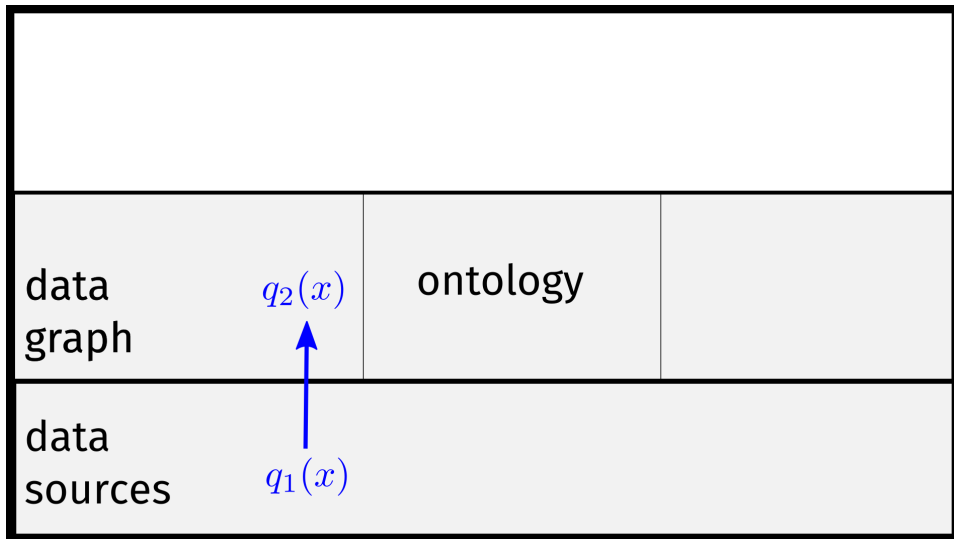
Expressivity contributions

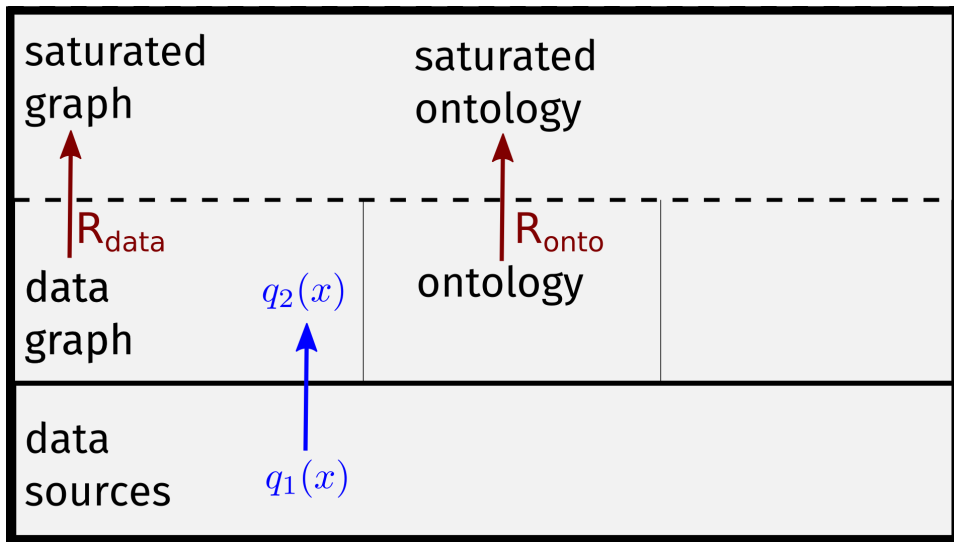
- Queries on the data *and* the ontology
- Global-Local-As-View mappings in an OBDA context



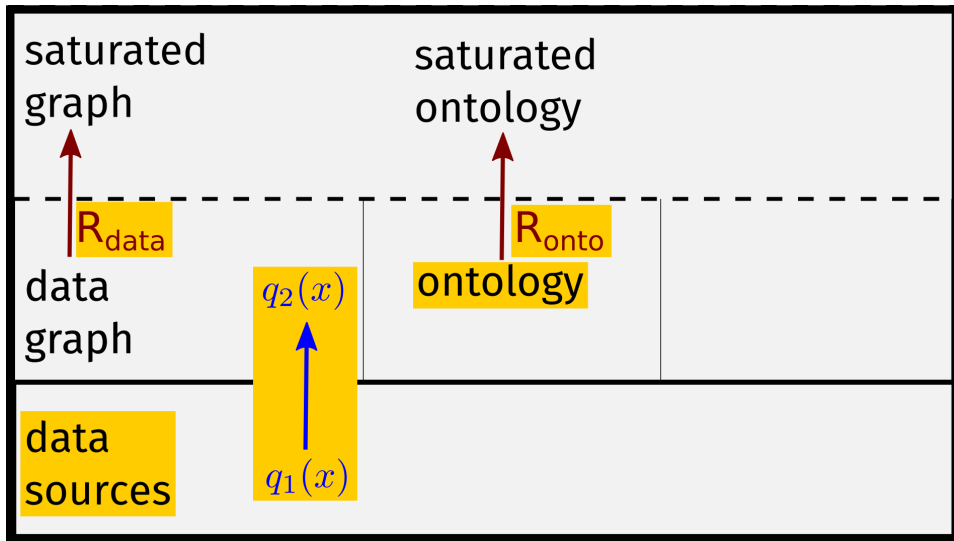
Global-Local-As-View Mapping Example



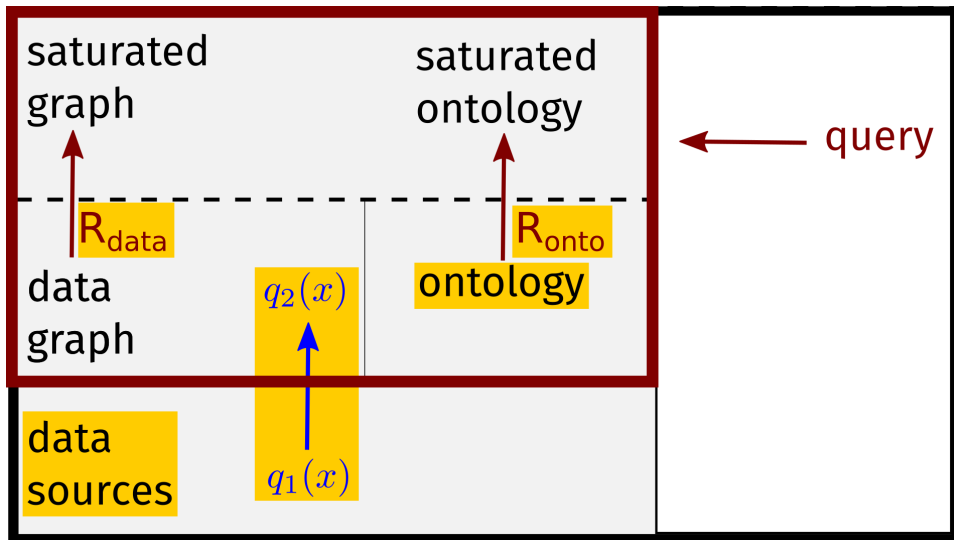




RDF Integration System

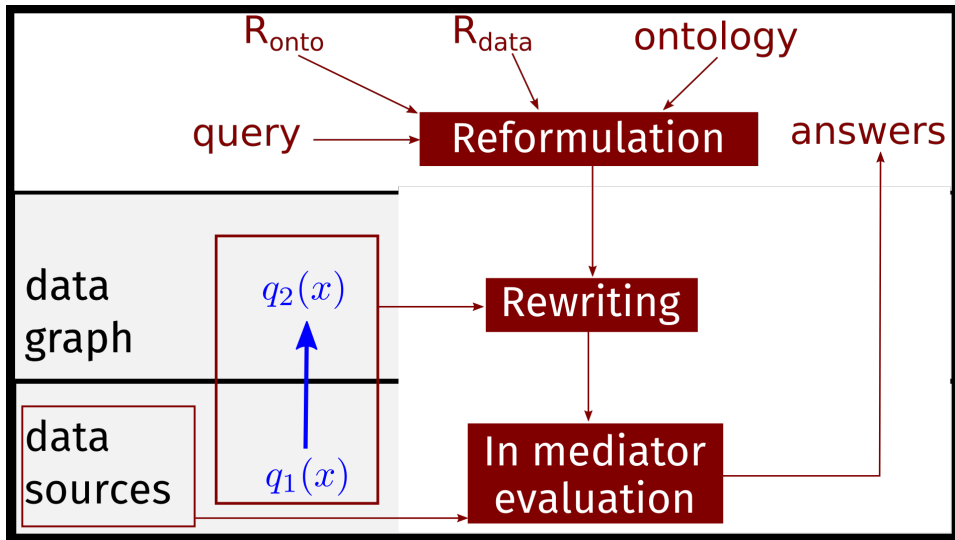


Query answering problem

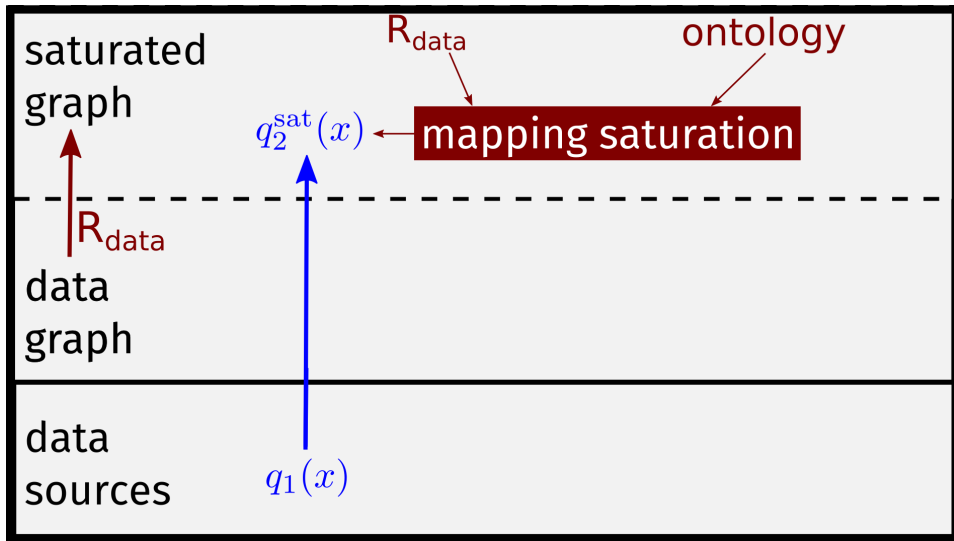


Query answering strategies

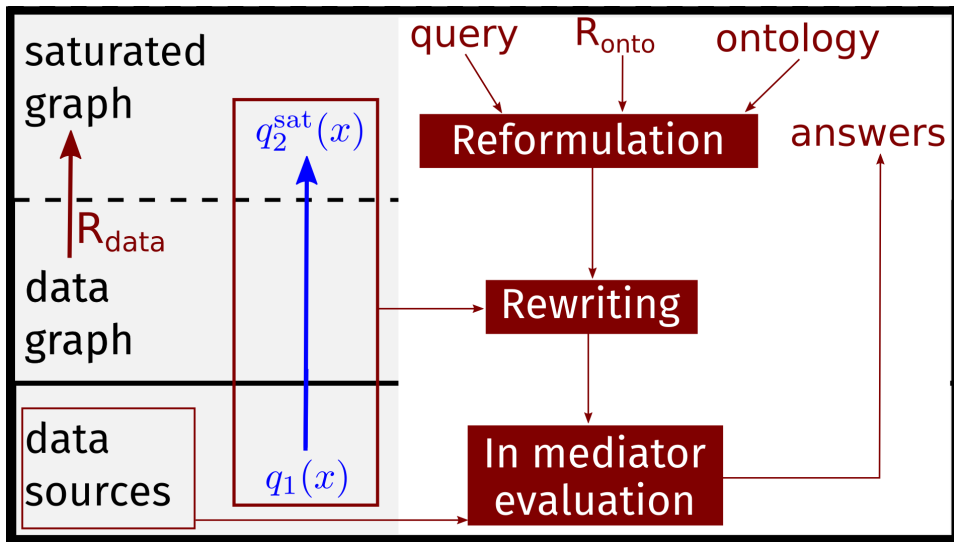
Classical strategy: *all* reasoning at query time



Some reasoning at query time method: mapping saturation



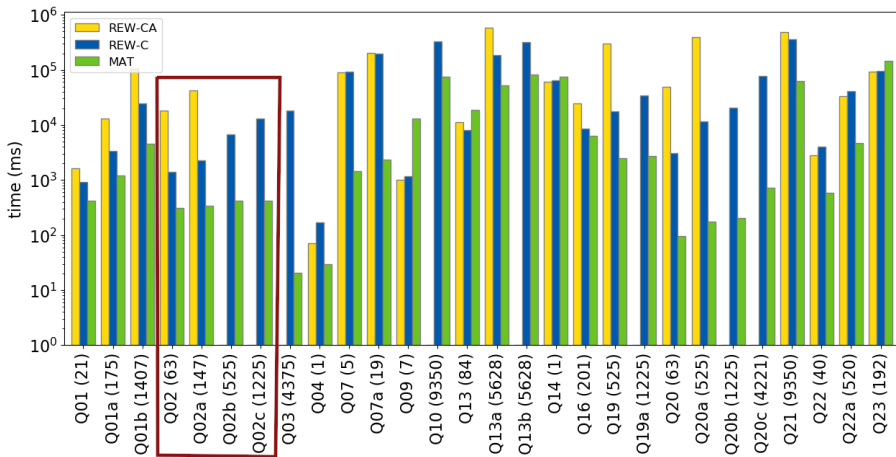
Some reasoning at query time method: query time



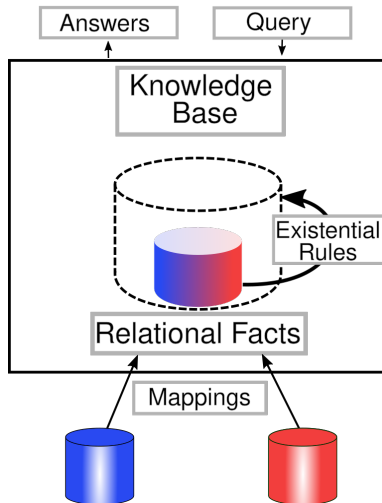
- **Software:**
 - *OntoSQL* (reformulation and materialization)
 - *Graal* (rewriting)
 - *Tatooine* (mediation)
- **RDF Integration System:**
 - Extension of BSBM
 - 3863 GLAV mappings
 - RDFS ontology of 2011 triples
 - Induced graph with 108M triples (185M triples when saturated)
 - Two data sources: One **relational** and one **JSON**

Query Answering Times on Heterogeneous Data Sources

- Materialization (MAT) - kind of reference time
- Full reformulation + rewriting (REW-CA)
- Mapping saturation + partial reformulation + rewriting (REW-C)



Perspective : mapping saturation with existential rules



Thank you

- Demo:
<https://perso.limos.fr/~maburon/projects/obi-wan/app/>
- <https://gitlab.inria.fr/cedar/obi-wan>

