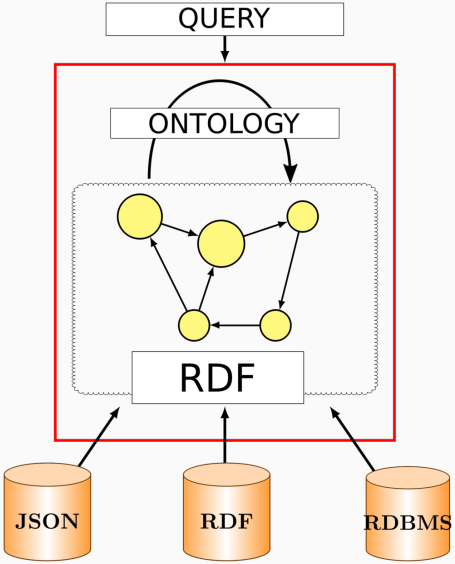
Ontology-Based RDF Integration of Heterogeneous Data

Maxime Buron, François Goasdoué, Ioana Manolescu, Marie-Laure Mugnier

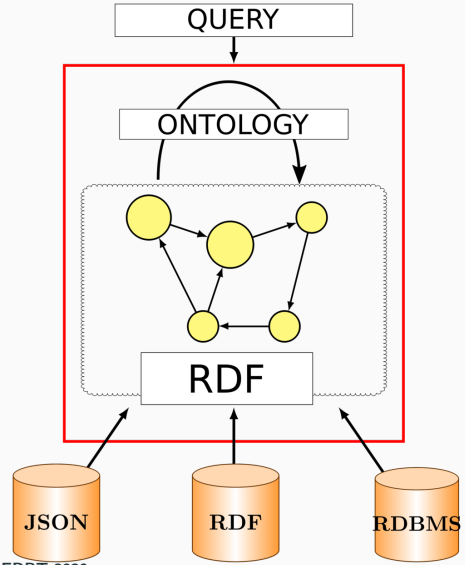
EDBT 2020

Context: Ontology-Based Data Access

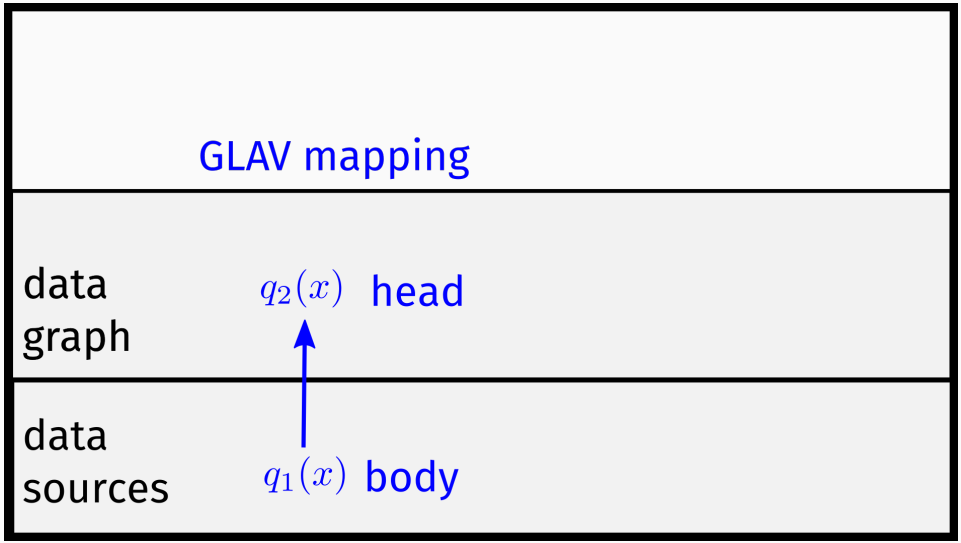


Contributions

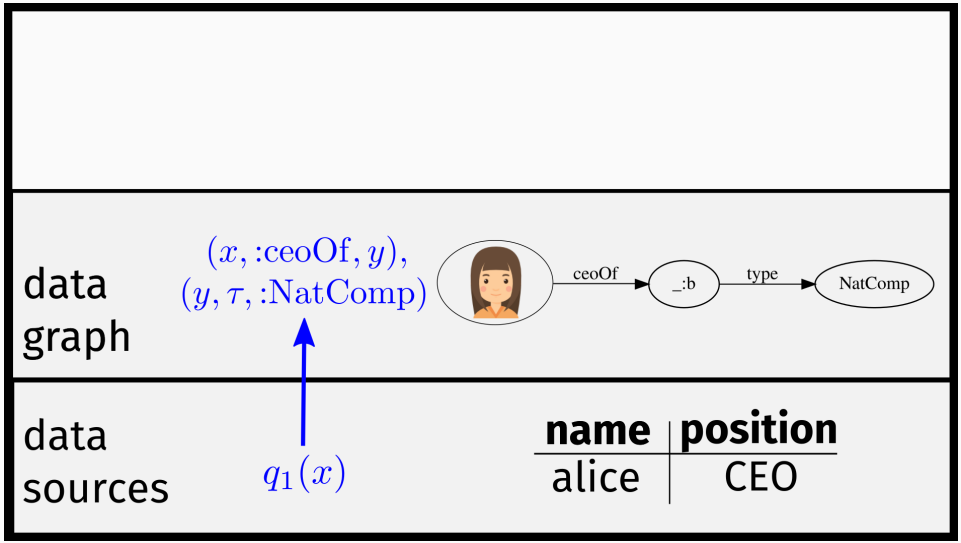
1. More powerful integration setting:
   - Global-Local-As-View mappings in an OBDA context
   - Queries on the data *and* the ontology
2. Two novel query answering methods: based on the amount of reasoning performed *offline* and at *query time*



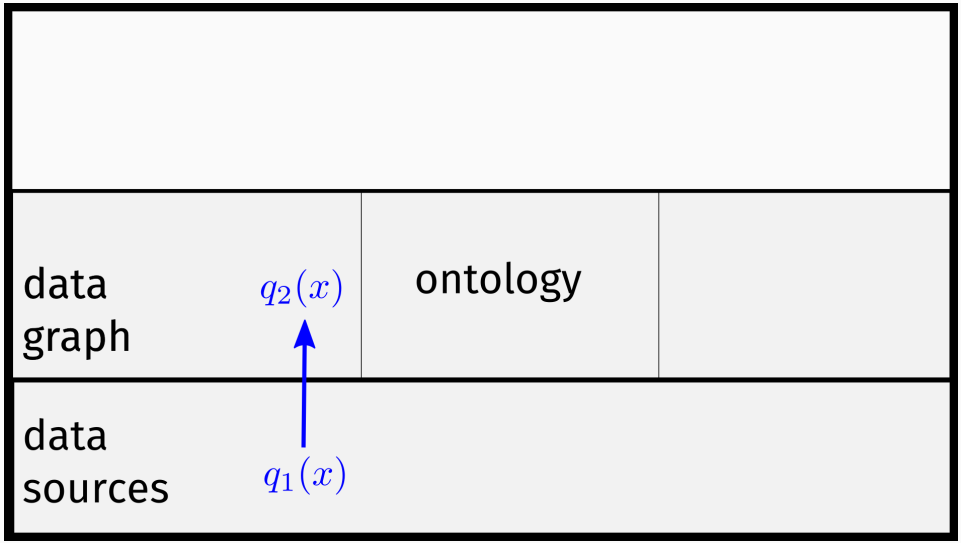
Global-Local-As-View Mapping



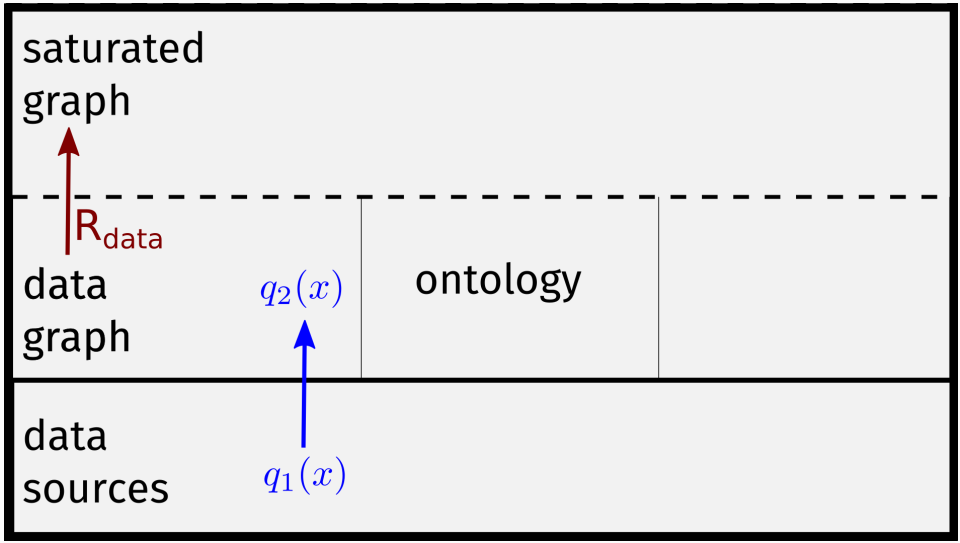
Global-Local-As-View Mapping Example



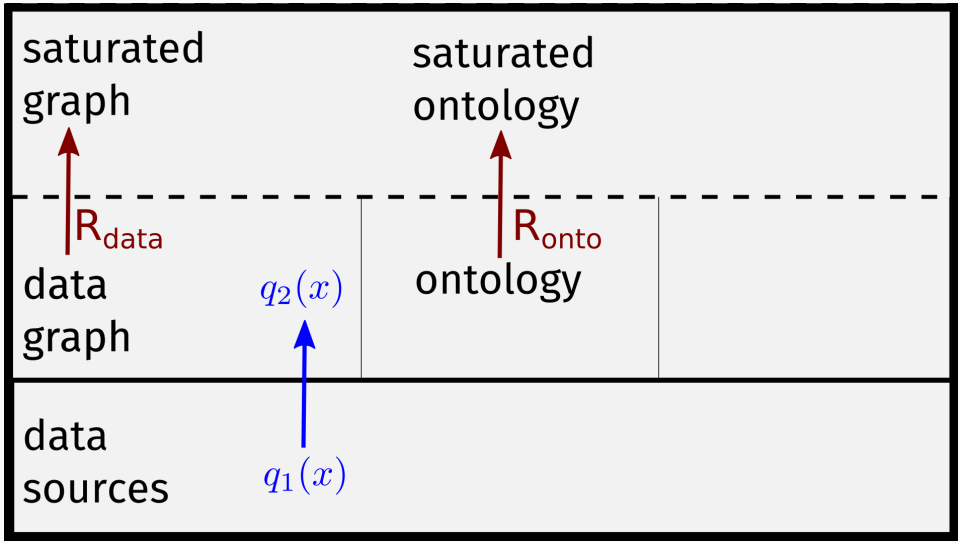
RDFS Ontology



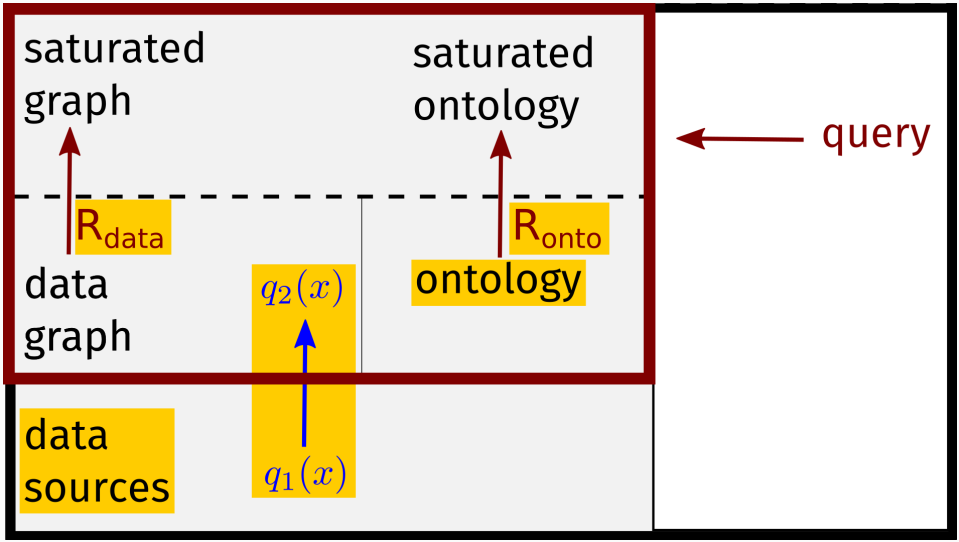
RDFS Entailment of Data Triples



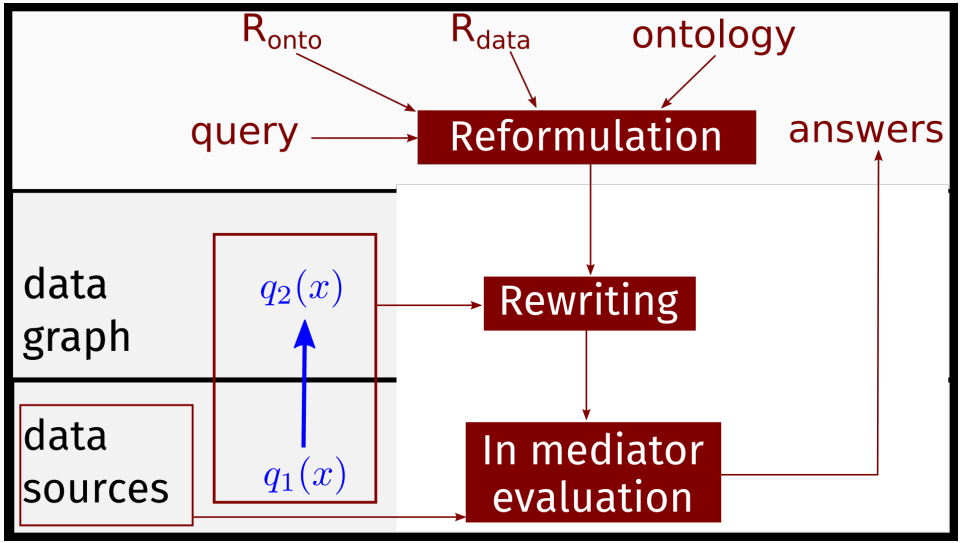
RDFS Entailment of Ontological Triples



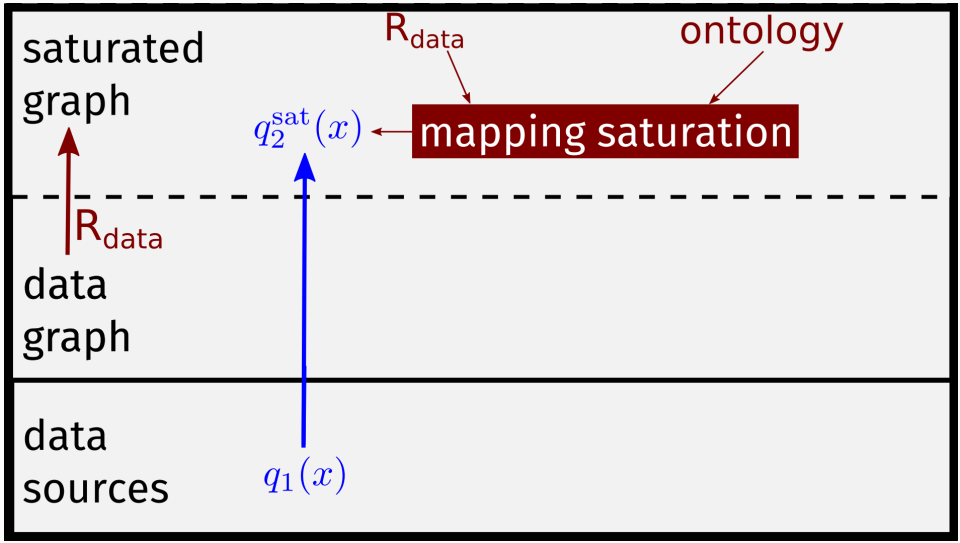
Query Answering Problem



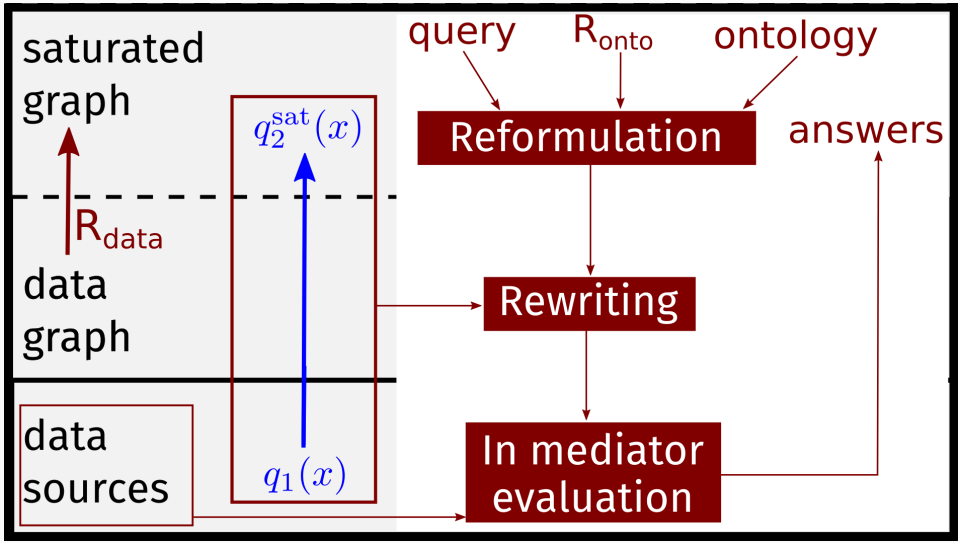
Classical Method: *All* Reasoning at Query Time



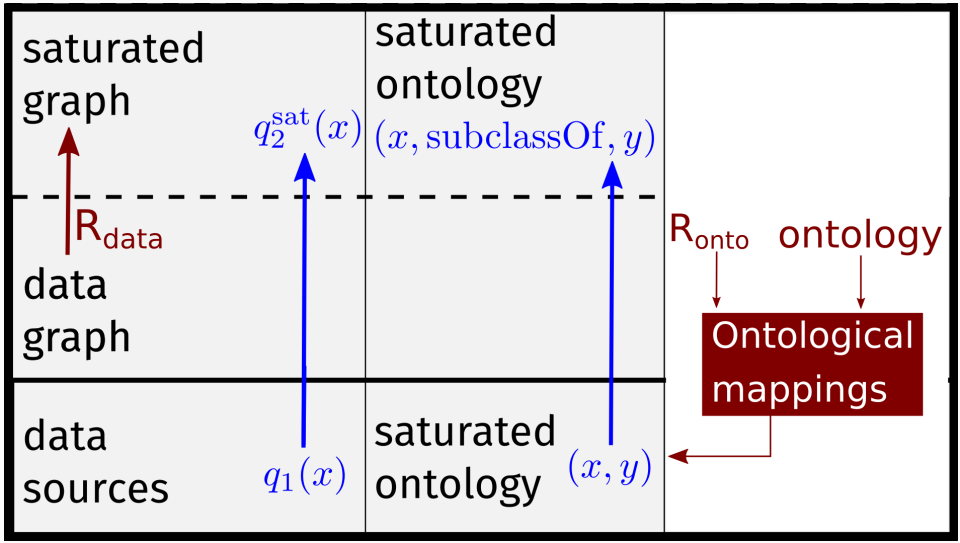
Some reasoning at Query Time Method: Preprocessing



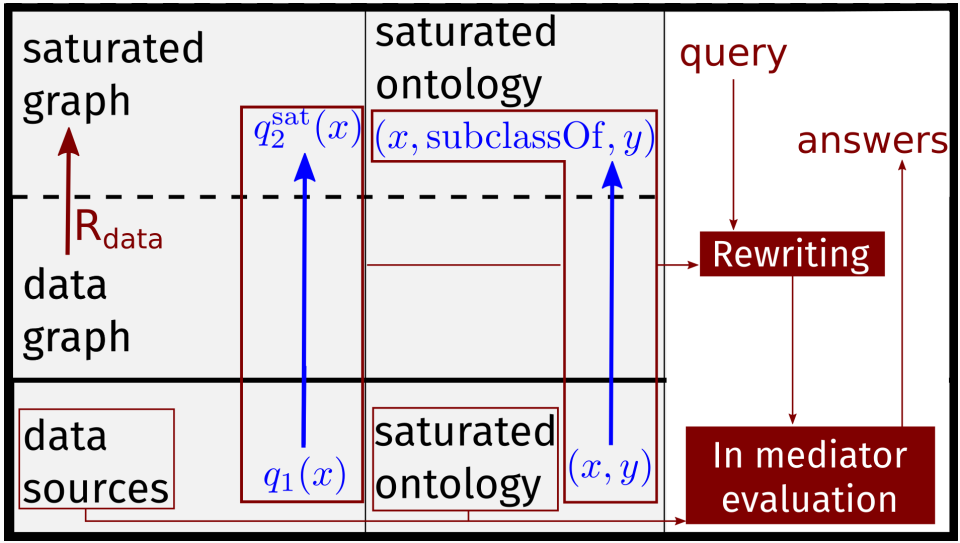
Some reasoning at Query Time Method: Query Time



No Reasoning at Query Time: Pre-Processing



No Reasoning at Query Time: Query Time

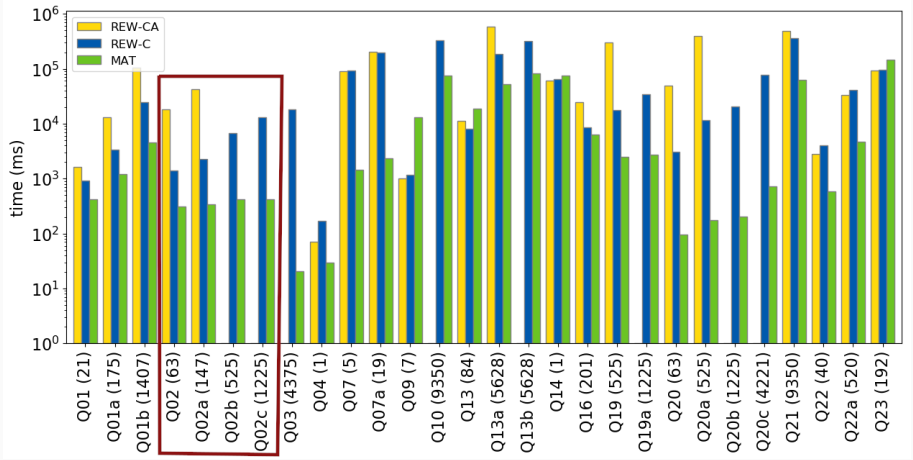


Experiments

- **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)



Conclusion

- **Global-Local-As-View mappings** in OBDA Context
- **Queries on data and ontology**
- A new **scalable query answering strategy** using partial reformulation and saturated mappings

Obi-Wan demo at: `http://pages.saclay.inria.fr/maxime.buron/projects/obi-wan/`