

# Reasoning on Data with Existential Rules

**Marie-Laure Mugnier**

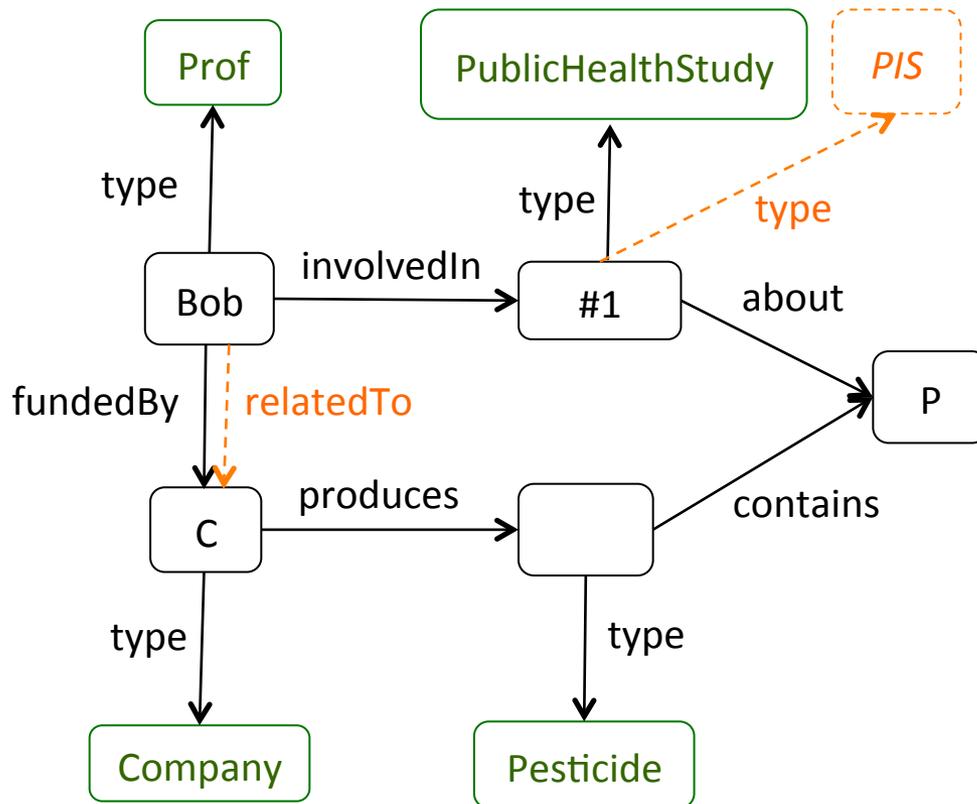
University of Montpellier



*Inria*

Stanford – April 14, 2020

# PART OF A KNOWLEDGE GRAPH



## Facts

```

    ∃ b (
      Prof(Bob) ∧
      PHS(#1) ∧
      Comp(C) ∧
      Pest(b) ∧
      involvedIn(Bob,#1) ∧
      fundedBy(Bob,C) ∧
      about(#1,P) ∧
      produces(C,b) ∧
      contains(b,P) )
  
```

+ Basic ontological knowledge

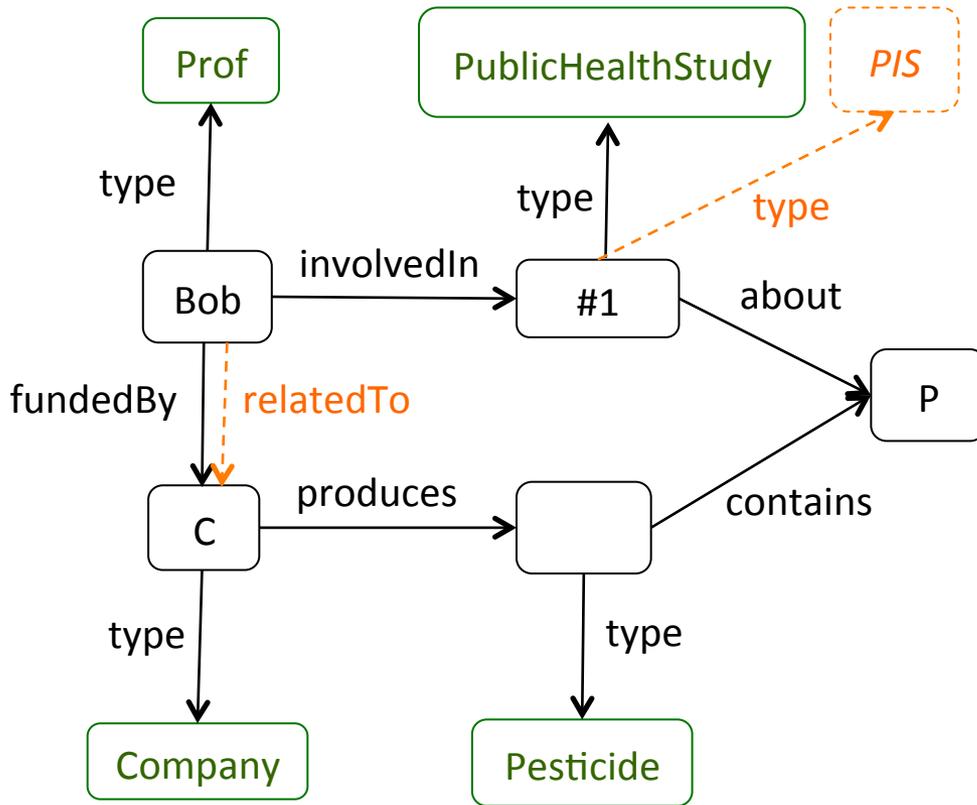
PublicHealthStudy **subclass of** PublicInterestStudy  
 fundedBy **subproperty of** relatedTo

## Rules

```

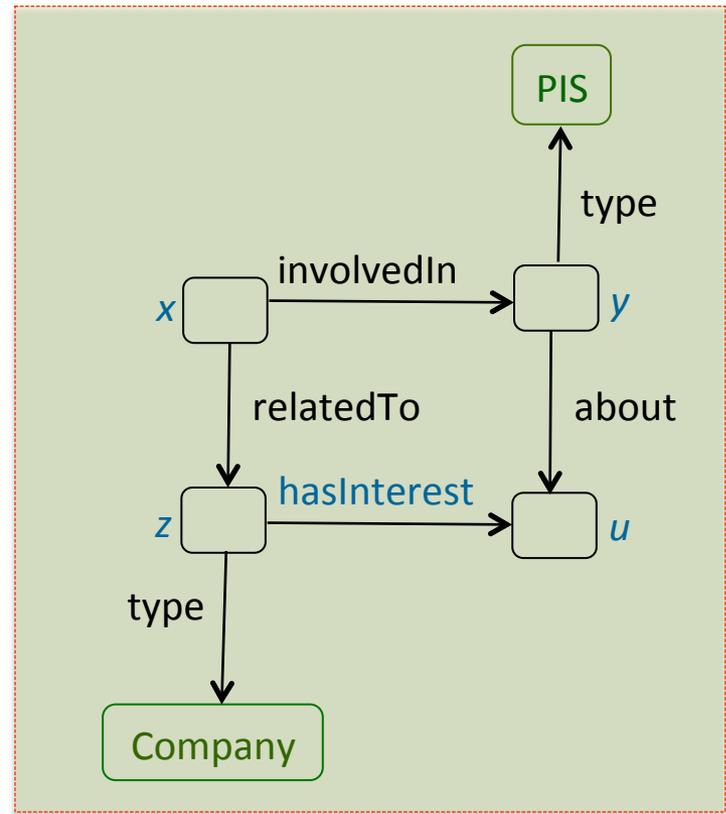
    ∀ x (PHS(x) → PIS(x))
    ∀ x ∀ y (fundedBy(x,y) → relatedTo(x,y))
  
```

# HOW TO INFER CONFLICTS OF INTEREST (COI) ?



What kind of **ontological knowledge** would allow to infer conflicts of interest?

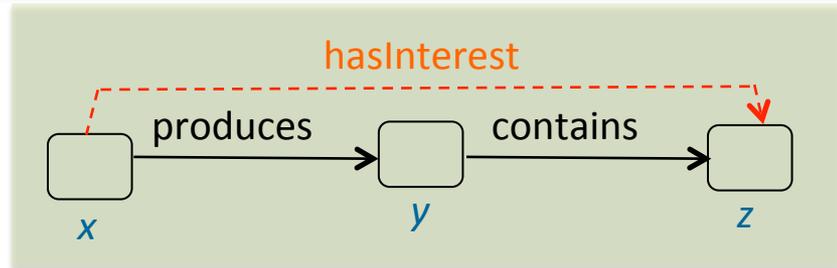
**Query:** "Find all  $x, y, z$  such that  $x$  has a conflict for study  $y$  because of its relationships with company  $z$ "



COI pattern

# DEFINING CONFLICTS OF INTEREST

$$R_1: \forall x \forall y \forall z ( \text{produces}(x,y) \wedge \text{contains}(y,z) \rightarrow \text{hasInterest}(x,z) )$$

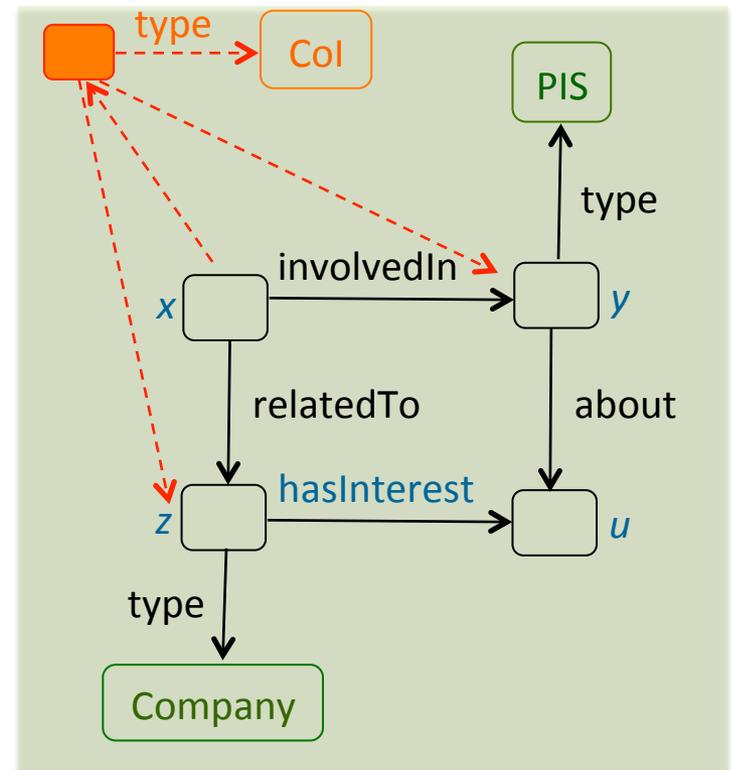


$$R_2: \forall x \forall y \forall z \forall u ( \text{involvedIn}(x,y) \wedge \text{PIS}(y) \wedge \text{about}(y,u) \wedge \text{relatedTo}(x,z) \wedge \text{Company}(z) \wedge \text{hasInterest}(z,u) \rightarrow \text{Col}(x,y,z) )$$

What if we only have unary and binary predicates ie graphs and not hypergraphs ?

**Reification:** new object of type Col

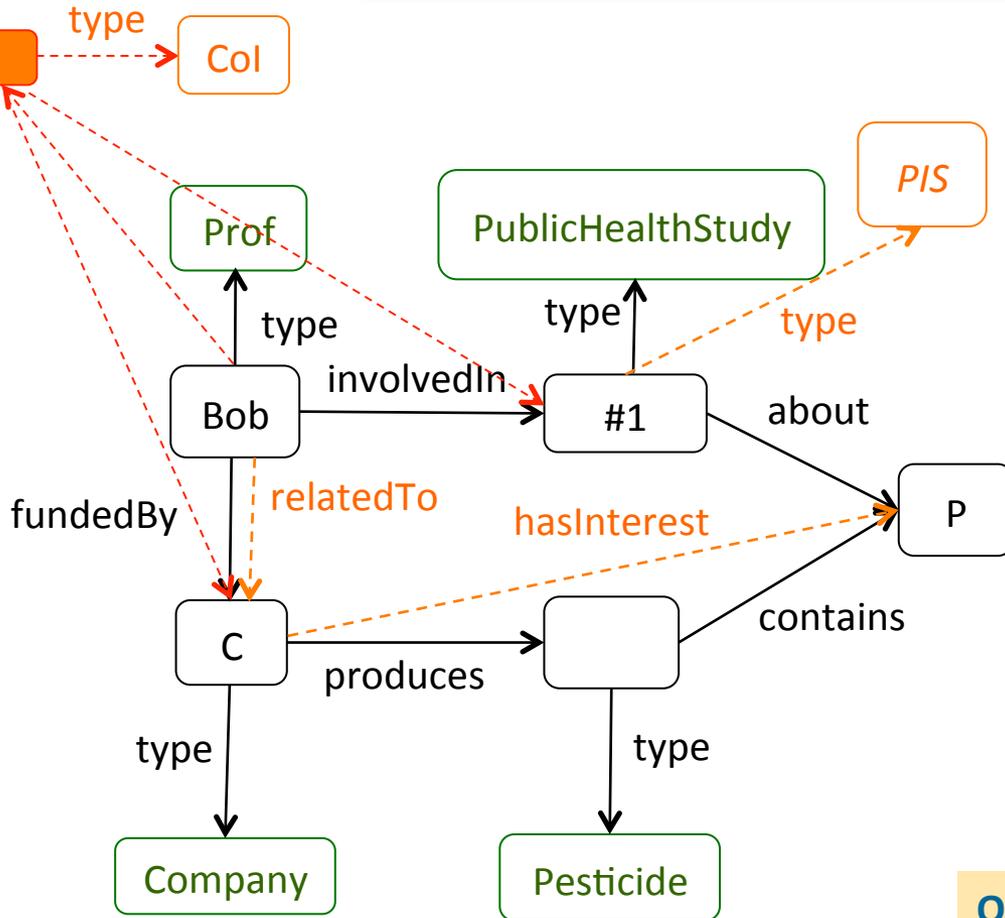
$$R_2: \forall x \forall y \forall z \forall u ( \text{body}[x,y,z,u] \rightarrow \exists o ( \text{Col}(o) \wedge \text{in}(x,o) \wedge \text{on}(o,y) \wedge \text{with}(o,z) ) )$$



Interest of creating a new object:

- **Flexible** description of Col (different patterns) instead of a fixed arity predicate
- Ability to **talk about** Col

# INFERRING CONFLICTS OF INTEREST



## Facts

$\exists x$  Prof(Bob), PHS(#1), Comp(C), Pest(x)  
 involvedIn(Bob,#1), fundedBy(Bob,C)  
 about(#1,P), produces(C,x), contains(x,P)

## Rules (universal quantifiers omitted)

PHS(x)  $\rightarrow$  PIS(x)  
 fundedBy(x,y)  $\rightarrow$  relatedTo(x,y)

$R_1$ : produces(x,y)  $\wedge$  contains(y,z)  
 $\rightarrow$  hasInterest(x,z)

$R_2$ : involvedIn(x,y)  $\wedge$  PIS(y)  $\wedge$  about(y,u)  $\wedge$   
 relatedTo(x,z)  $\wedge$  Company(z)  $\wedge$   
 hasInterest(z,u)  
 $\rightarrow \exists o$  Col(o)  $\wedge$  in(x,o)  $\wedge$  on(o,y)  $\wedge$  with(o,z)

## Inferred facts

PIS(#1), relatedTo(Bob,C), hasInterest(C,P)  
 Col(o<sub>1</sub>), in(Bob,o<sub>1</sub>), on(o<sub>1</sub>,#1), with(o<sub>1</sub>,C)

**Query:** find (x,y,z) such that  
 $\exists o$  Col(o)  $\wedge$  in(x,o)  $\wedge$  on(o,y)  $\wedge$  with(o,z)

**Answer:** (Bob,#1,C)

# EXISTENTIAL RULES AS AN ONTOLOGICAL LANGUAGE

---

$\forall X. \forall Y. \text{Body } [X,Y] \rightarrow \exists Z. \text{Head } [X,Z]$

$X, Y, Z :$   
*sets of variables*

any **positive conjunction** (without functional symbols except constants)

$\forall x. \forall y. \text{siblingOf}(x,y) \rightarrow \exists z \text{parentOf}(z,x) \wedge \text{parentOf}(z,y)$

*(Universal quantifiers will be omitted in examples)*

Key point: ability to assert **the existence of unknown entities**

Crucial for representing ontological knowledge in **open domains**

# RULE APPLICATION

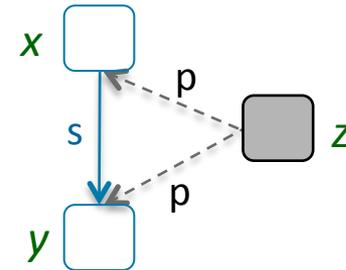
$R$  is **applicable** to  $F$  if there is a **homomorphism**  $h$  from  $body(R)$  to  $F$

ie a substitution  $h$  of the variables in  $body(R)$  by terms in  $F$  such that  $h(body(R)) \subseteq F$

$R = \text{siblingOf}(x,y) \rightarrow \exists z \text{parentOf}(z,x) \wedge \text{parentOf}(z,y)$

$F = \{ \text{siblingOf}(a,b) \}$

$h: x \mapsto a$   
 $y \mapsto b$



The **application** produces  $h(head(R))$

where a fresh variable (a **null**) is created for each existential variable in  $R$

$F' = \{ \text{siblingOf}(a,b), \text{parentOf}(z_0,a), \text{parentOf}(z_0,b) \}$

# THEORETICAL FOUNDATIONS

Graph-based KR

[Chein Mugnier  
1992, 2009]

logical translation of graph rules

$\forall \exists$ -rules, existential Rules [Baget+ IJCAI 2009]

Datalog+/- family [Cali+ PODS 2009]

Datalog (70-80s)

+ *existential variables  
in rule heads*

+ *complex relationships  
between objects*  
+ *unbounded predicate arity*

RDFS

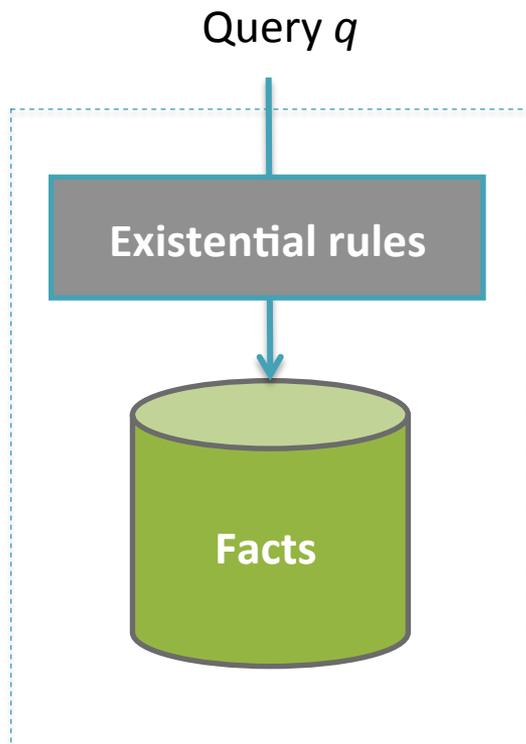
Lightweight Description Logics,

e.g. OWL 2 tractable profiles

More generally, Horn Description Logics

- Same logical form as « Tuple-Generating Dependencies » (TGDs)  
long studied in database theory

# QUERY ANSWERING PROBLEM



## Knowledge Base $K$

Find all the answers to  $q$   
*entailed by  $K$*

## Conjunctive query (CQ)

SELECT ... FROM ... WHERE *<join conditions>*

SELECT ... WHERE *<basic graph pattern>*

Existentially quantified conjunction of atoms,  
where free variables are answer variables

$\exists o (\text{Col}(o) \wedge \text{in}(x,o) \wedge \text{on}(o,y) \wedge \text{with}(o,z))$

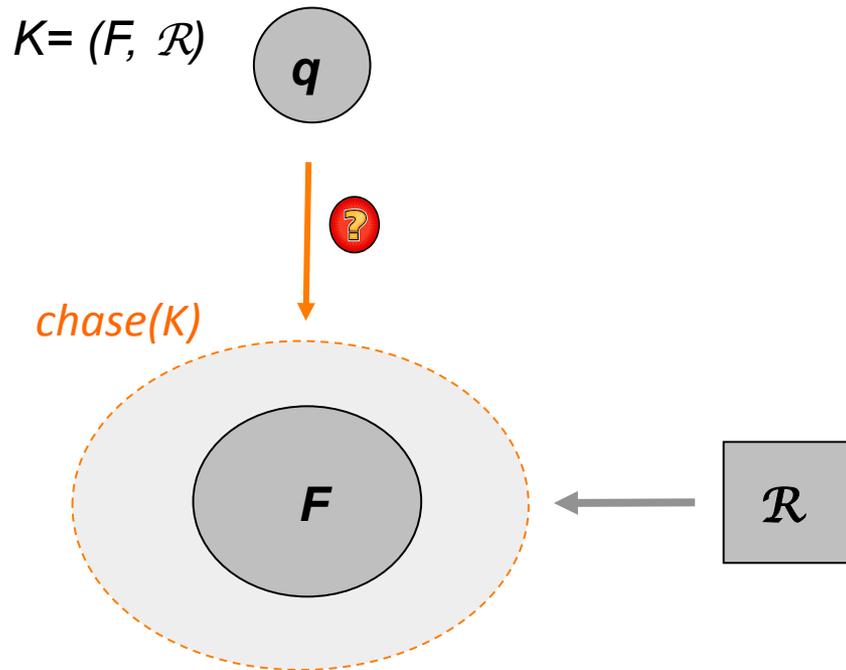
## Query answering **decision problem**:

Given a KB  $K$  and CQ  $q$ , does  $K$  provide an answer to  $q$  ?

**Undecidable**

but many decidable and tractable classes are known

# 1. FORWARD CHAINING (CHASE) - MATERIALISATION



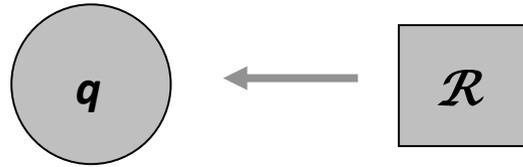
For any CQ  $q$ ,  
 $Answers(q, K) = Answers(q, chase(K))$

Of course  $chase(K)$  may be infinite

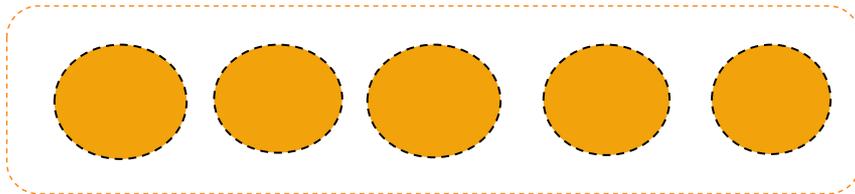
e.g.  $Human(x) \rightarrow \exists z \text{ parentOf}(z, x) \wedge Human(z)$  with  $F = Human(Bob)$

## 2. BACKWARD CHAINING BY QUERY REWRITING

$K = (F, \mathcal{R})$

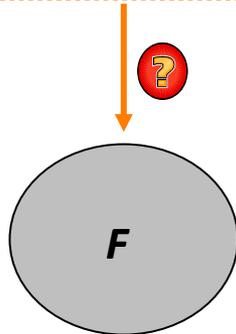


1.  $q$  is rewritten into a **union of CQs**, and more generally into a **first-order query** (core SQL query)



$rew(q, \mathcal{R})$

2. The rewriting is evaluated on the facts

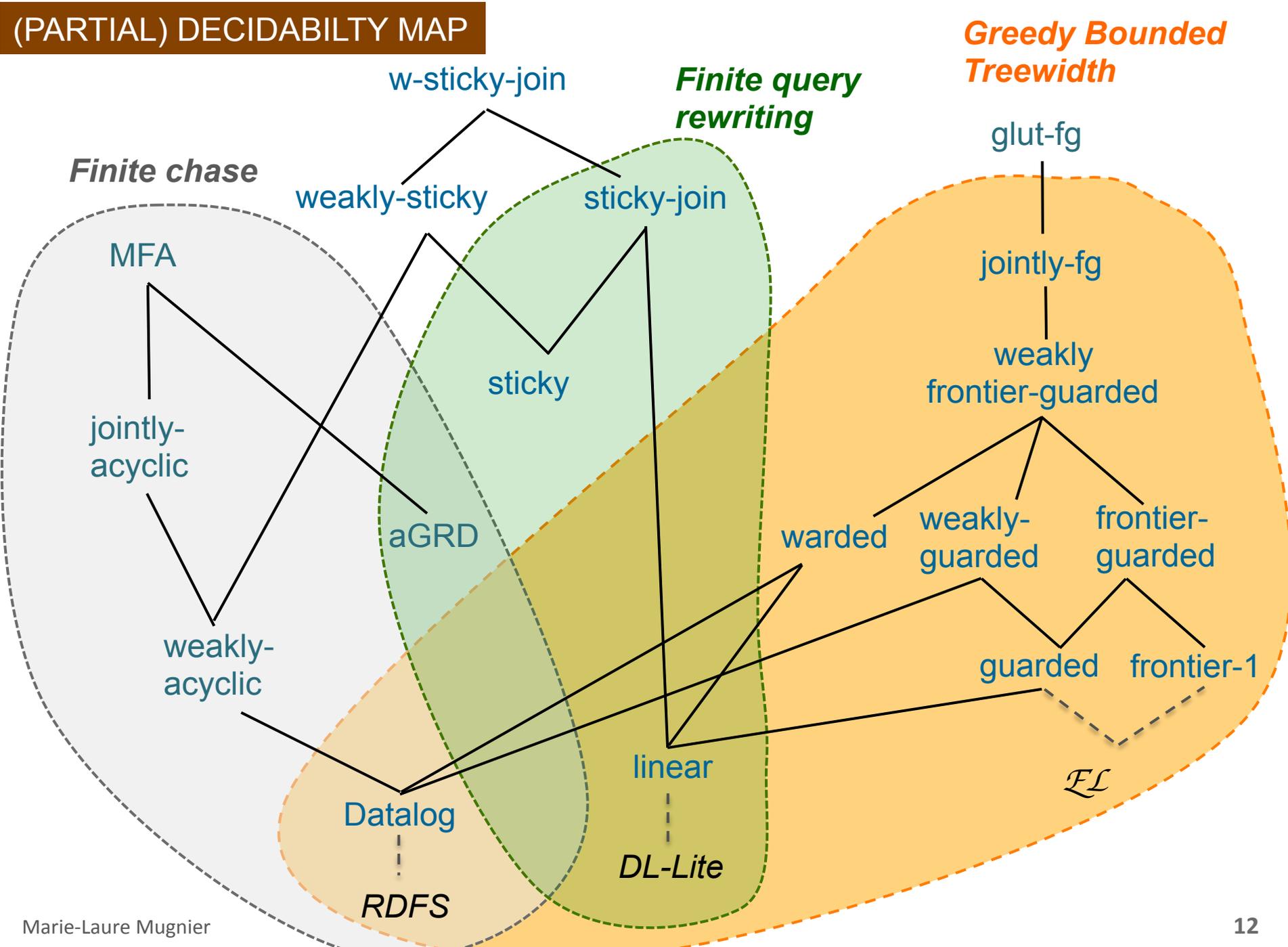


Query rewriting is independent from any set of facts

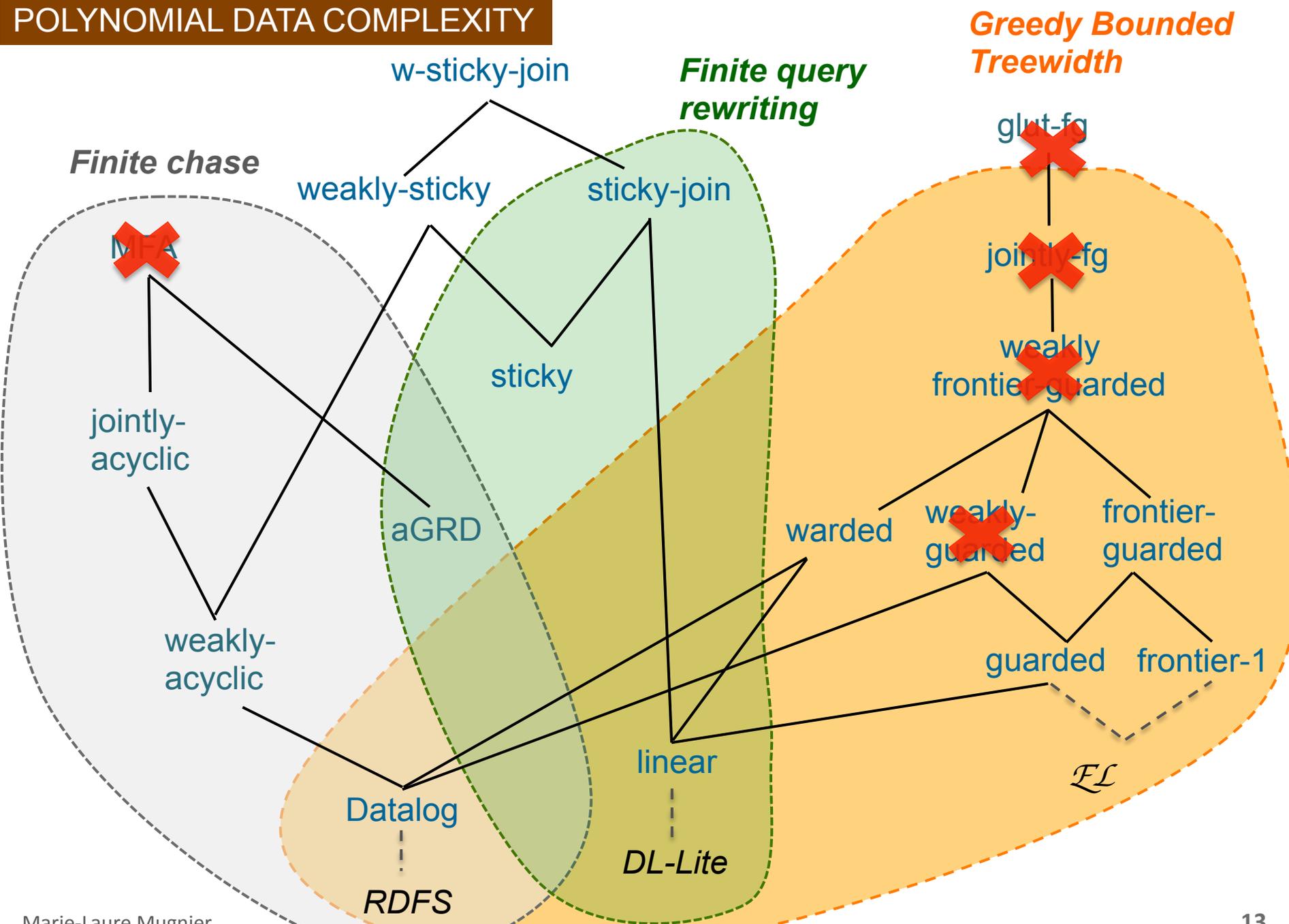
For **any**  $F$ ,  $Answers(q, (F, \mathcal{R})) = Answers(rew(q, \mathcal{R}), F)$

Of course, there may be no finite  $rew(q, \mathcal{R})$

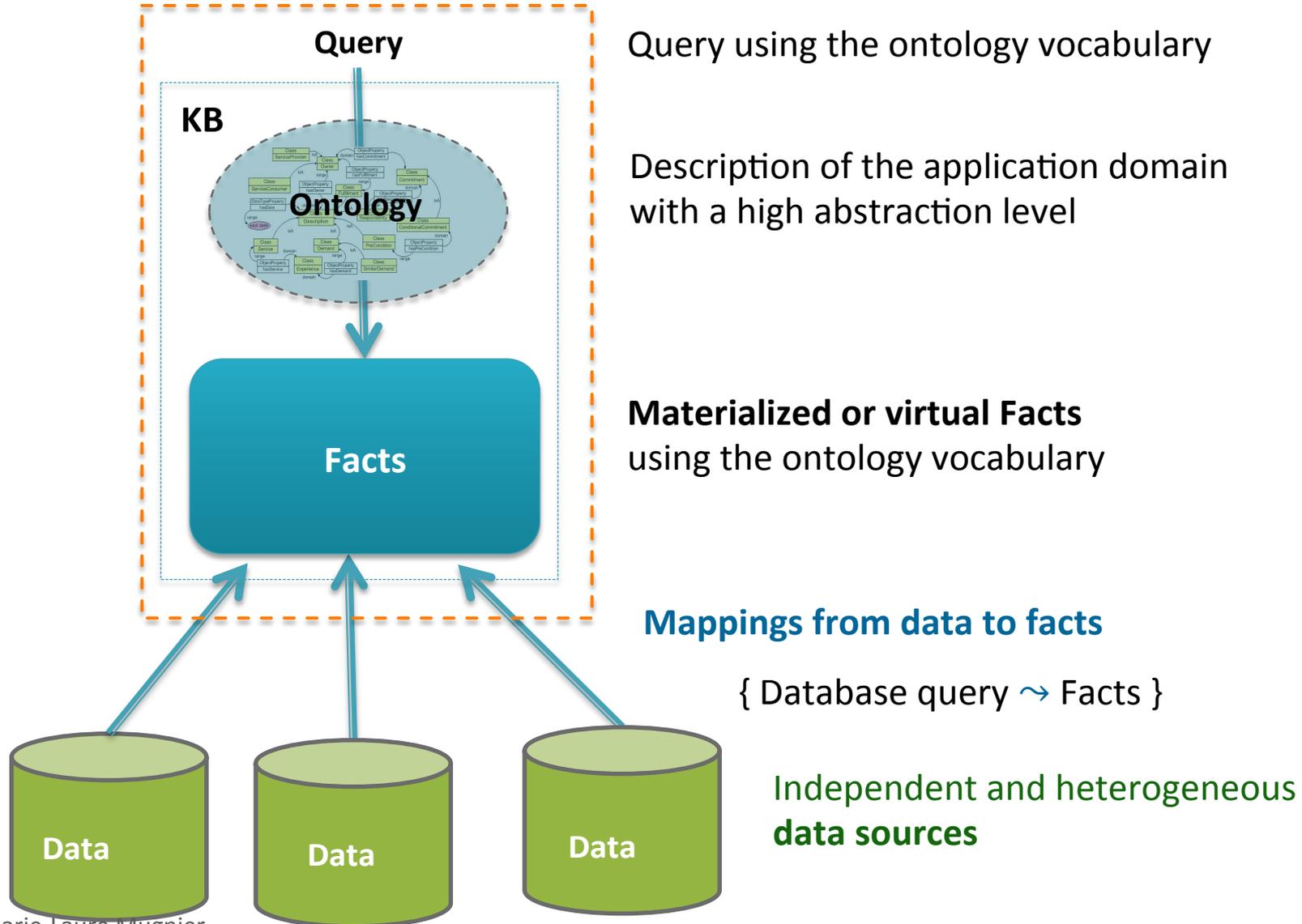
# (PARTIAL) DECIDABILITY MAP



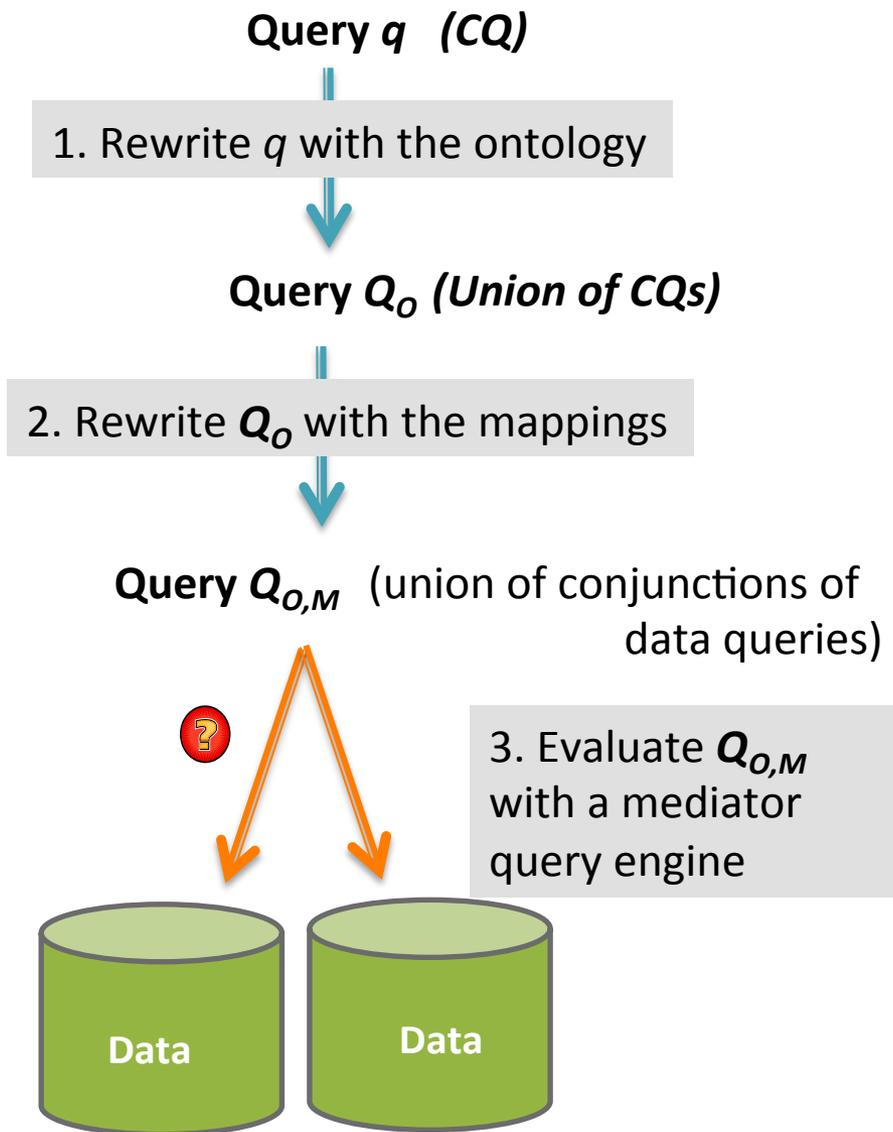
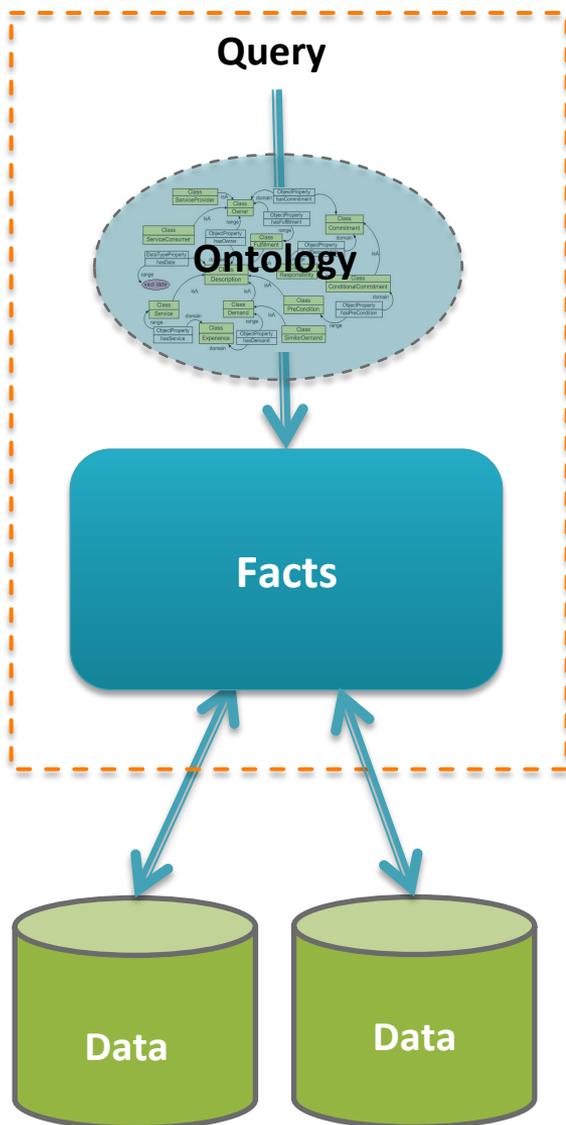
# POLYNOMIAL DATA COMPLEXITY



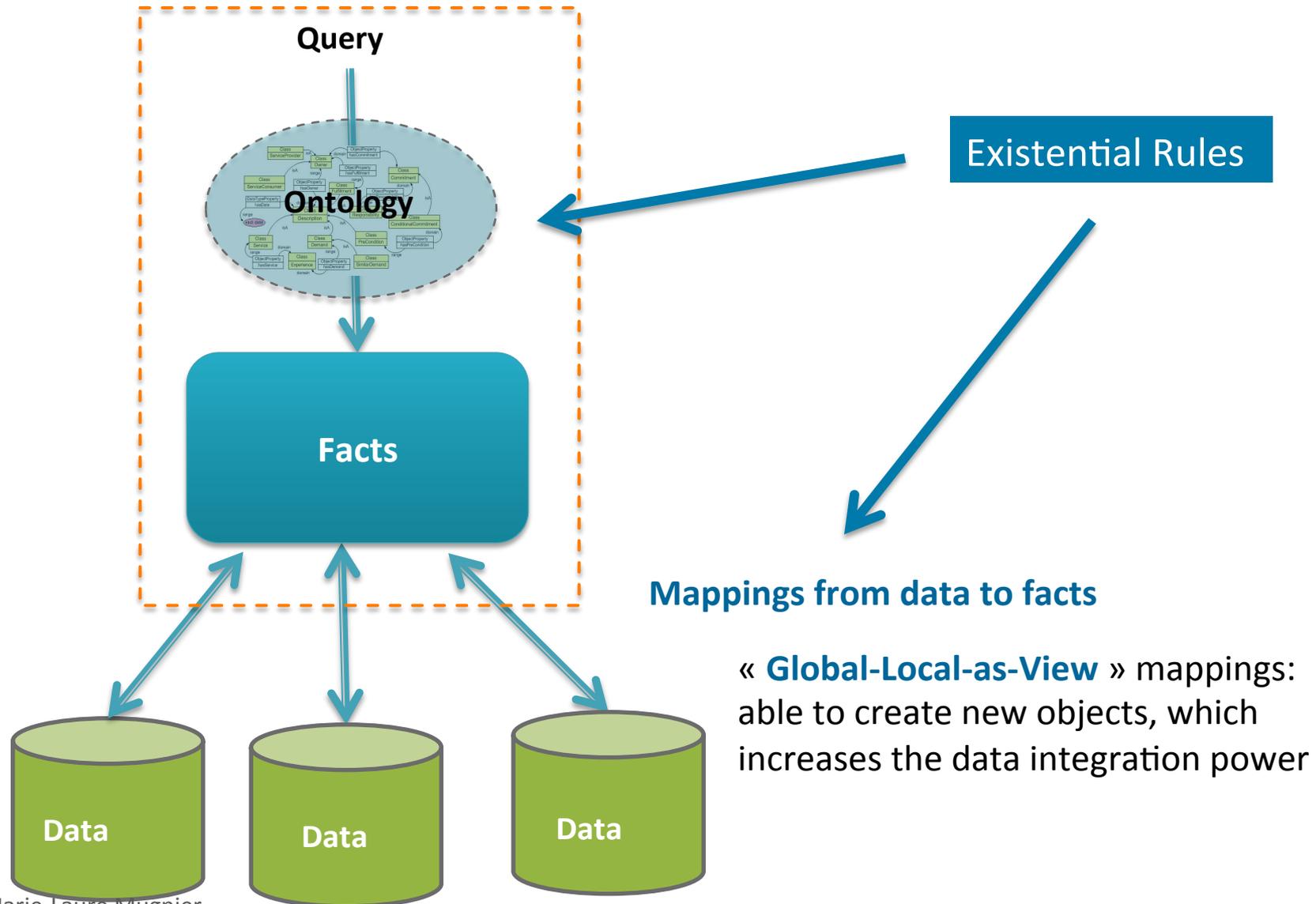
## Conceptual level



# OBDA WITH VIRTUAL FACTS



# EXISTENTIAL RULES AS A UNIFORM LANGUAGE FOR OBDA



# CONCLUSION

---

- Existential rules are able to express **complex structures** and create **new objects**
- These features can be exploited for both expressing **ontological knowledge** and **integrating data**
- A wide range of rule classes offer various **expressivity/complexity** tradeoffs
- The framework has been **extended** in several ways:
  - other rules: negative constraints and Equality Generating Dependencies
  - existential rules extended to stratified negation and disjunctive heads
- Efficient **systems** are available

# SOME SYSTEMS

---

## Query answering with existential rules

**VLog** fast chase-based engine <https://iccl.inf.tu-dresden.de/web/VLog/en>

**RDFox** fast chase-based engine <https://www.cs.ox.ac.uk/isg/tools/RDFox/>

**Vadalog** fast chase-based engine for warded Datalog+/- [Commercial]

**Graal** toolkit with fast query rewriting, rule set analysis, chase algorithms, ...  
<http://graphik-team.github.io/graal/>

**and many other usable tools** developed for other purposes:

**Llunatic** data exchange and data cleaning (chase on tuple-generating dependencies)

**DLV** Answer Set Programming (DLV<sup>∃</sup>: for shy existential rules) [Commercial]

...

**OBDA systems** are still restricted to lightweight description logics or (extended) RDFS:  
**OnTop, Mastro, UltraWrap**<sup>OBDA</sup> ...

# FURTHER READING ON THE EXISTENTIAL RULE FRAMEWORK

---

## Introductions and surveys

M.-L. Mugnier (2011). [Ontological Query Answering with Existential Rules](#) (Keynote Talk). Int. Conf. on Web Reasoning and Rule Systems (RR 2011), LNCS 6902, pp 2-23, Springer.

G. Gottlob, G. Orsi, A. Pieris, M. Simkus (2012). [Datalog and its extensions for semantic web databases](#). Reasoning Web Summer School, LNCS 7487, pp 54–77, Springer.

M.-L. Mugnier, M. Thomazo (2014). [An introduction to ontology-based query answering with existential rules](#). Reasoning Web Summer School, LNCS 8714, pp 245–278, Springer.

## Seminal papers (journal versions)

J.-F. Baget, M. Leclère, M.-L. Mugnier, E. Salvat (2011). [On Rules with Existential Variables: Walking the Decidability Line](#), Artificial Intelligence Journal 175, pp. 1620-1654.

A. Cali, G. Gottlob, G. Lukasiewicz (2012). [A general Datalog-based framework for tractable query answering over ontologies](#). Journal of Web Semantics 14, pp 57-83.

## Latest results

See conferences in AI (IJCAI, KR), databases (PODS, ICDT, VLDB) and the semantic web (ISWC)