THÉORIE DES BASES DE CONNAISSANCES

HMIN312M

Marie-Laure MUGNIER (Univ. Montpellier)

Jean-François BAGET (Inria)

Equipe GraphIK (LIRMM & Inria)

https://team.inria.fr/graphik/

KNOWLEDGE REPRESENTATION AND REASONING (KR)

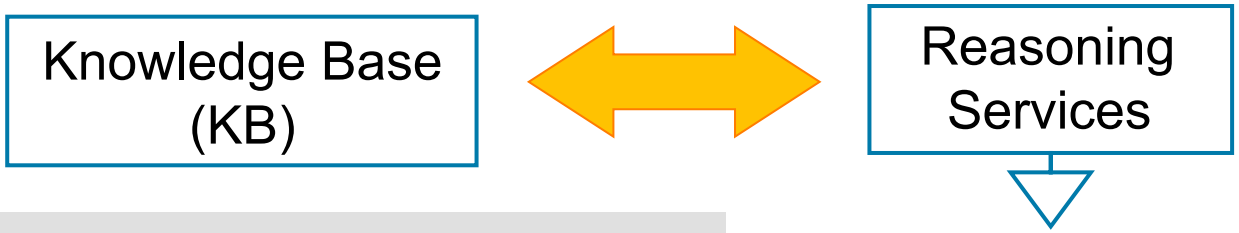
- A field historically at the heart of **Artificial Intelligence**
- Study **formalisms (or languages)** to
  - **represent** various kinds of human knowledge
  - do **reasoning** on these representations
- along the tradeoff **expressivity / tractability of reasoning**

→ KR languages based on **computational logic**

Mainly first-order logic (FOL)

Major conferences: IJCAI, AAAI, KR

KNOWLEDGE BASED SYSTEMS



- **General knowledge on the application domain**
  « Cats are Mammals »

  Ontology
- **Factual Knowledge**
  Description of specific individuals, situations, ...

  Félix is a Cat

  Factbase, Database

Knowledge expressed in a KR language

Fundamental tasks

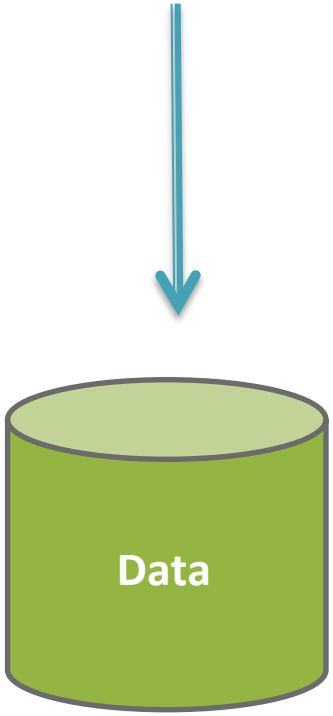
- **Checking the consistency** of the KB
- **Computing answers** to a query over the KB

...

Reasoning algorithms associated with the KR language

Reasoning on data (ex: Medical Records)

Query (SQL, SPARQL, MongoDB ...)



« find all patients affected by a lung disease due to a bacteria »

??

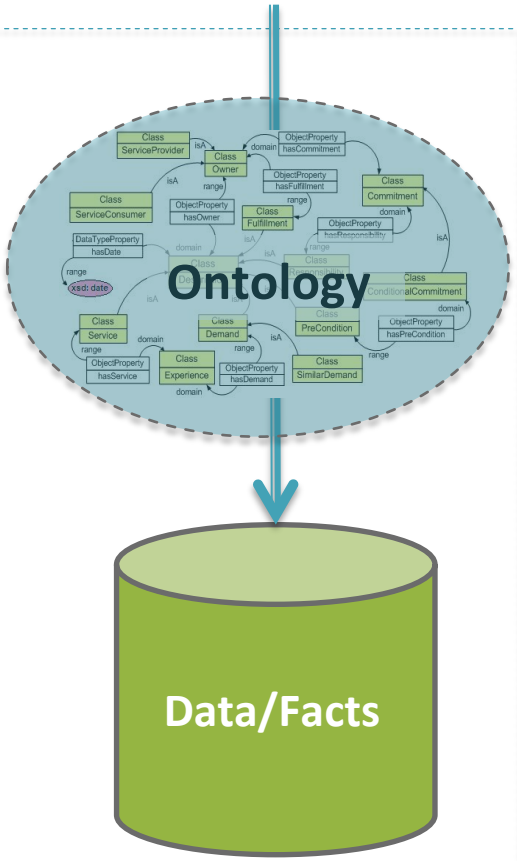
ID Patient	Diagnosis
P	« legionella »

Database (relational, RDF, NoSQL, ...)

Adding an Ontological Layer

Query

« find all patients affected by a lung disease due to a bacteria »



A legionella is bacterial pneumonia
A bacterial pneumonia is a pneumonia
A pneumonia is a **lung disease**
A lung disease is a disease
A bacterial pneumonia is a bacterial disease
A bacterial disease is caused by a **bacteria**
If *x* is caused by *y* then *x* **is due to** *y*
If the diagnosis of a patient *x* contains a disease y then *x* is **affected by** *y*

ID Patient	Diagnosis
P	« legionella »

Knowledge Base

Ontology-Mediated Query Answering

$q(x) = \exists y \exists z$ (Patient(x) $\wedge$ isAffectedBy(x,y) $\wedge$ LungDisease(y) $\wedge$ dueTo(y,z) $\wedge$ Bacteria(z))

« find all patients affected by a lung disease due to a bacteria »

Factbase = { Patient(P), Diagnosis(P,M), Legionella(M) }

« The diagnosis for the patient P is legionella »

$\forall x$ (Legionella(x) $\rightarrow$ LungDisease(x) $\wedge$ BacterialDisease(x)) *[shortcut]*
hence LungDisease(M) and **BacterialDisease(M)**
$\forall x$ (LungDisease(x) $\rightarrow$ Disease(x))
hence Disease(M)

$\forall x$ (BacterialDisease(x) $\rightarrow$ $\exists y$ (hasCausativeAgent(x,y) $\wedge$ Bacteria(y)))
hence hasCausativeAgent(M,b) and **Bacteria(b)**

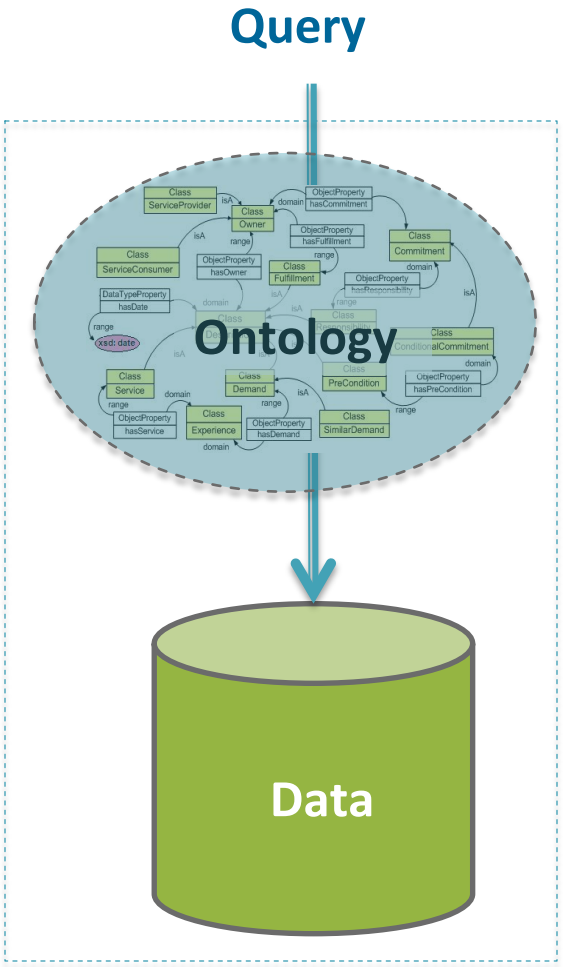
$\forall x \forall y$ (hasCausativeAgent(x,y) $\rightarrow$ dueTo(x,y))
hence dueTo(M,b)

$\forall x \forall y$ ((Diagnosis(x,y) $\wedge$ Disease(y)) $\rightarrow$ isAffectedBy(x,y))
hence isAffectedBy(P,M)

Answer : x = P

ONTOLOGY-MEDIATED QUERY ANSWERING (OMQA)

Adding an ontological layer on top of data



1- **Enrich** the vocabulary

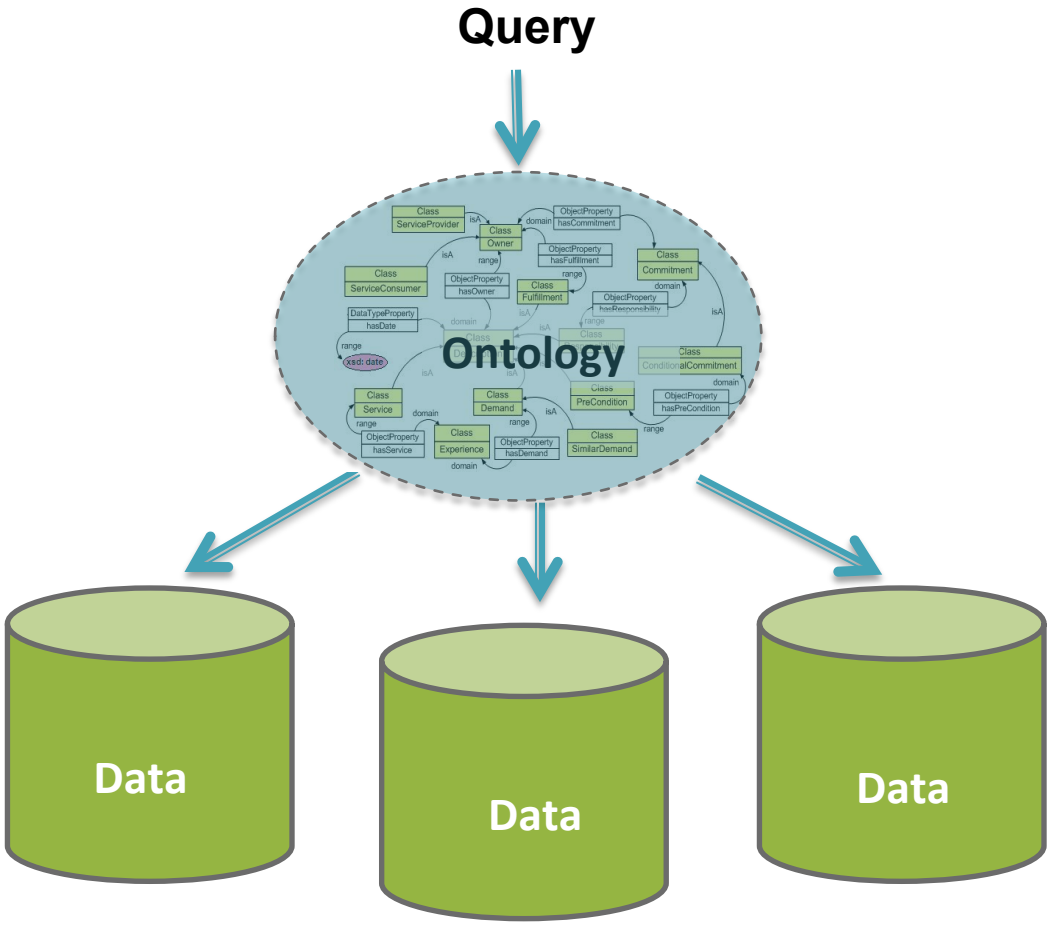
allowing to **abstract** from a specific data storage

2 - **Infer** new facts, not explicitely stored,

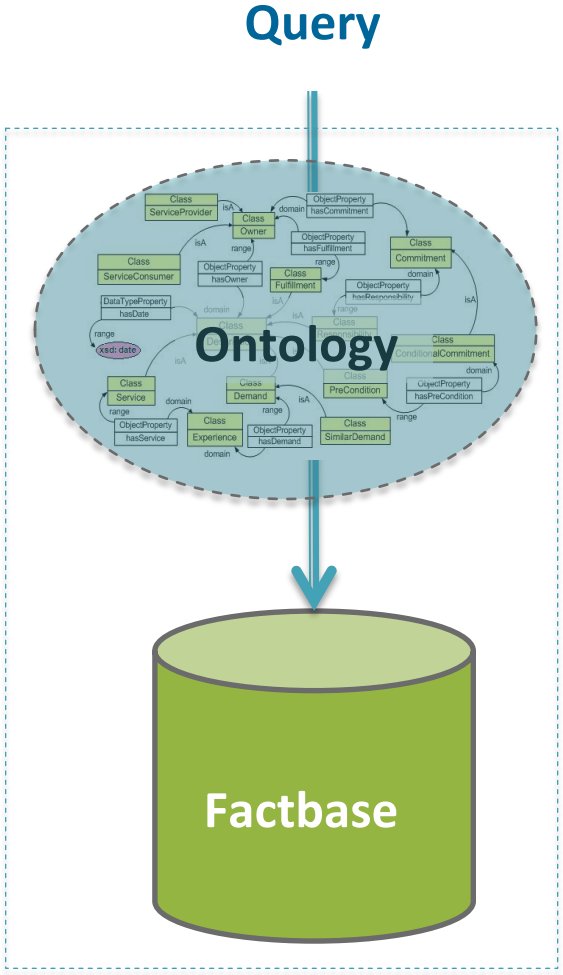
allowing for **incomplete data** representation

ONTOLOGY-MEDIATED QUERY ANSWERING (OMQA)

3 – provide a **unified view** of multiple sources



ONTOLOGY-MEDIATED QUERY ANSWERING (OMQA)



(Boolean) conjunctive query *q*

Set of formulas *O* in a suitable FOL fragment

Set of atoms (« facts ») *F*

Find all answers to *q*

that are *logically entailed* by (*O*, *F*)

Dans ce module

- Quels formalismes pour représenter ces connaissances ?

  → langages à base de règles

  - qui sont des « fragments » de la logique du premier ordre classique

    règles à la Datalog,

    contraintes négatives

    règles existentielles

  - qui sortent de la logique classique

    Datalog avec négation du monde clos

    Answer Set Programming (ASP)

    [sémantiques tolérantes aux inconsistances]

- Quelle est la complexité des raisonnements dans ces formalismes ?
- Quelles techniques algorithmiques ?