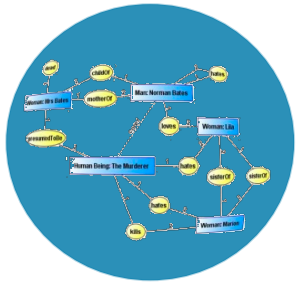
Laboratoire d'Informatique de Robotique et de Microélectronique de Montpellier

LIRMM



GraphIK

INRIA Project-team created in 2010

http://www.lirmm.fr/graphik

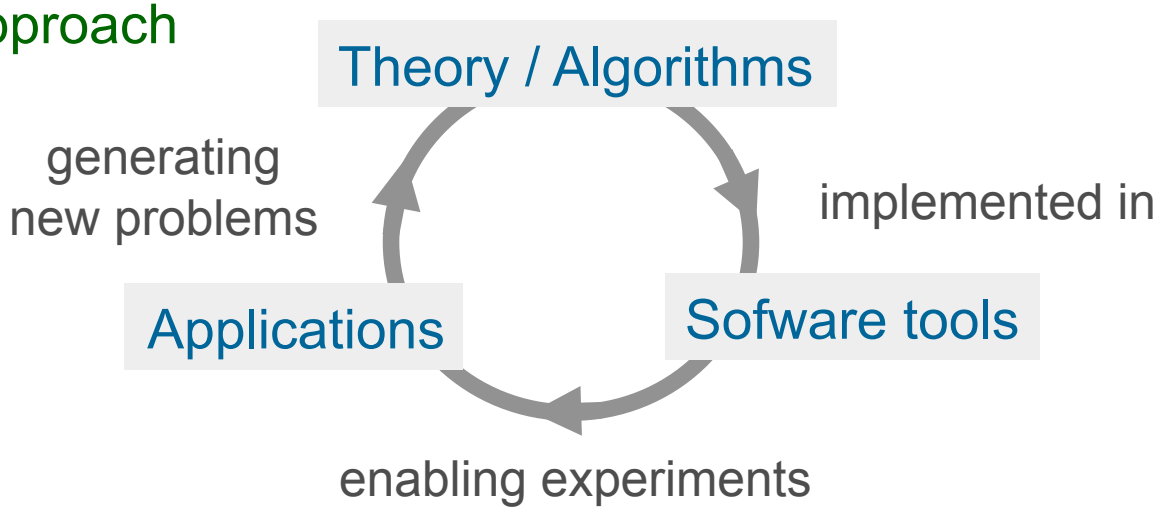
Jean-François Baget

baget@lirmm.fr

Athens, may 21-22, 2012

Thematic positioning / Approach

- Knowledge Representation and Reasoning (KR)
  - Logic-based but also graph-based approach to KR
  - Tradeoff expressivity of formalisms–tractability of reasoning
- Some challenges
  - Querying (large) knowledge bases
  - Reasoning with rules
  - Dealing with heterogeneous / hybrid knowledge bases
  - Representing and processing « imperfect » knowledge
- A three pole approach



GraphIK

Researchers	**2**
University members	**6**

[2010: 4]

Close Collaborator
Patrice Buche (IR INRA, HDR)

Engineers	**1,5**
PhD Students	**5 (+1)**

[2010: 2]

Expected PhD students
PhD with INA (CIFRE)

Team assistant

Annie Aliaga

Jean-François Baget (CR INRIA)
Michel Chein (Emeritus Pr UM2)
Madalina Croitoru (MdC UM2)
Jérôme Fortin (MdC UM2)
Souhila Kaci (Pr UM2) – Sept. 2011
Michel Leclère (MdC UM2)
Marie-Laure Mugnier (Pr UM2)
Rallou Thomopoulos (CR INRA)

Alain Gutierrez (IE CNRS)
Luc Menut (AI INRA, 50%)

Léa Guizol (ABES/GraphIK, 1st year)
Mélanie König
(Ministry grant, 1st year)
Bruno Paiva Lima Da Silva
(Ministry grant, 2nd year)
Tjitze Rienstra
(Univ. Luxembourg, 2nd year)
Michaël Thomazo
(AMN Cachan, 2nd year)

Scientific axes

Axis 1: Decidability, complexity, algorithms
for languages fitting in classical logic

2010-12 focus Conjunctive queries with negation
Ontological Query Answering

Axis 2: Representing and processing imperfect knowledge

2010-12 focus Argumentation systems for multi-criteria decision making
Default rules (+ fuzzy values and types)
Applications to agrifood chains (INRA)

Axis 3: Integration of theoretical tools
into real knowledge-based systems

2010-12 focus Semantic data integration:
Application to ABES catalogs (National Bibliographic Agency)

International positioning

With respect to our **core competencies** (see Axis 1)

- KR groups with emphasis on
  - Computational aspects
  - Knowledge bases
- Data management groups using AI techniques

Dresden TU (Franz Baader's group)
Bremen (Theory of AI, Carsten Lutz's group)
Karlsruhe (Knowledge Management, Rudi Studer's group)
Vienna (Knowledge-Based Systems, Thomas Eiter's group)
Bozen-Bolzano (« KR meets Databases », Werner Nutt's group)
Roma La Sapienza (AI, e.g. Riccardo Rosati)
Oxford (Information Systems, Georg Gottlob's and Ian Horrocks' groups)
Our specificity: graph-based aspects, structural reasoning

Shift from Conceptual Graphs (→2007-2009) to logic/graph-based KR formalisms

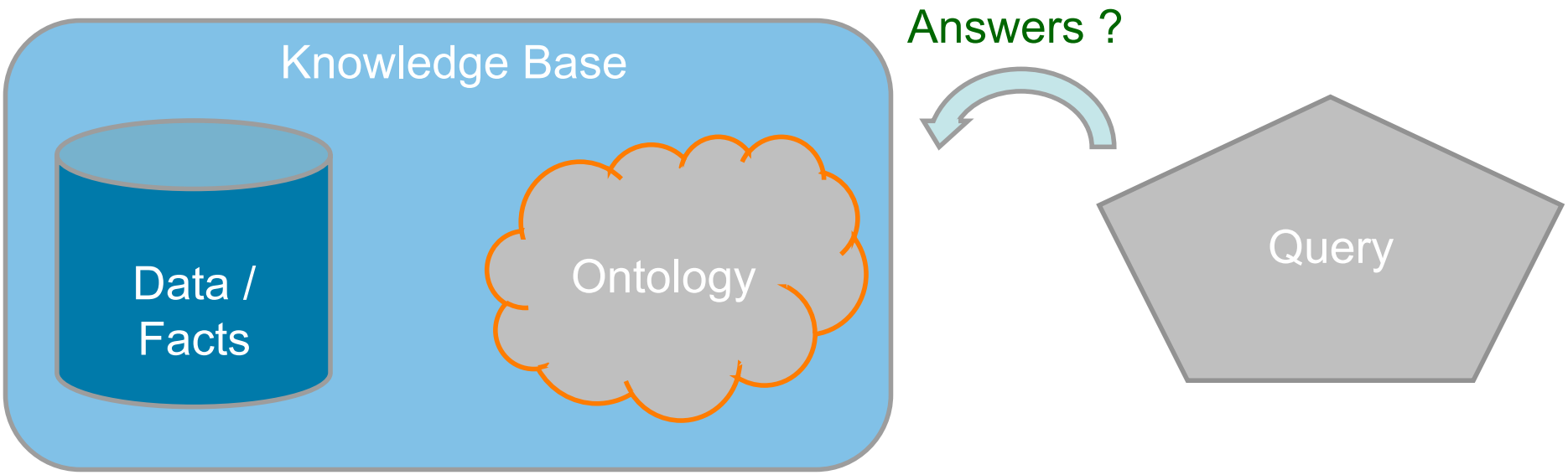
Close INRIA teams

- **LEO**
  - For data and knowledge integration
  - Common publications on data fusion in 2008 and 2010
  - Common submission to ANR program « Contint » (Sept. 2011) on the quality and interoperability of large document catalogs
- **EDELWEISS (→WIMMICS)**
  - For graph-based tools for querying heterogeneous knowledge bases
  - Latest common publication in 2008 (Color Griwes)
- **ZENITH**
  - Expected collaboration in the context of the Labex « Numev », (Numerical solutions and modelization for the environment and life), « Scientific Data Management » axis

- Other KR related teams: **EXMO**, **ORPAILLEUR**

Axis 1: Decidability, Complexity, Algorithms …

Focus on Ontological Query Answering
also known as Ontology-Based Data Access



Challenge in databases: take the ontology into account

Challenge in KR: query large fact bases

Data / Facts

Relational Database

parentOf

A	B
A	C
C	*x*
…	…

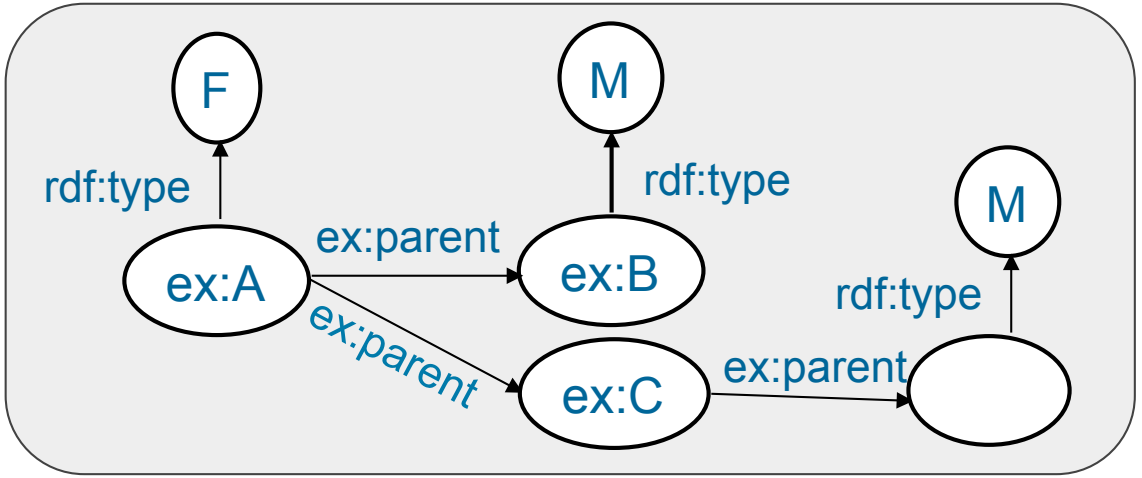
Male

B
x
…

Fem.

A
…

RDF (Semantic Web)

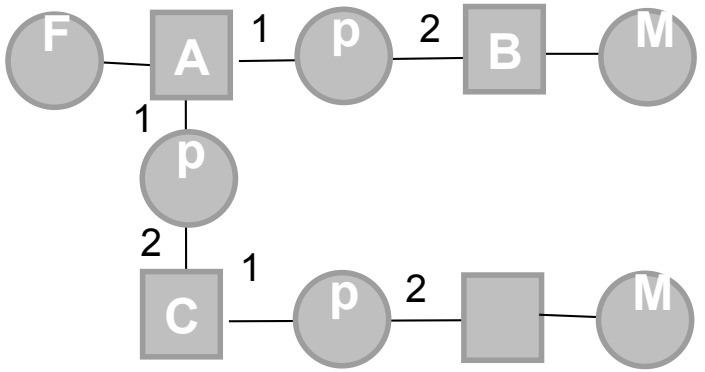


Etc.

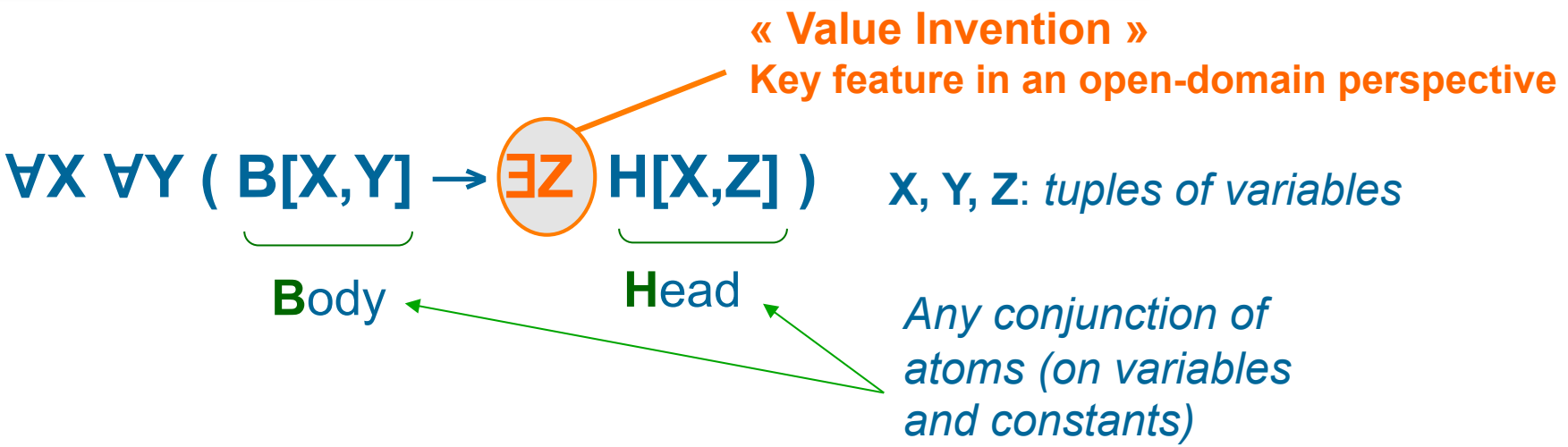
Abstraction in first-order logic

$$\exists x(\ \text{parentOf}(A,B) \wedge \text{parentOf}(A,C) \wedge \text{parentOf}(C,x) \wedge F(A) \wedge M(B) \wedge M(x)\ )$$

Or in **graphs / hypergraphs**



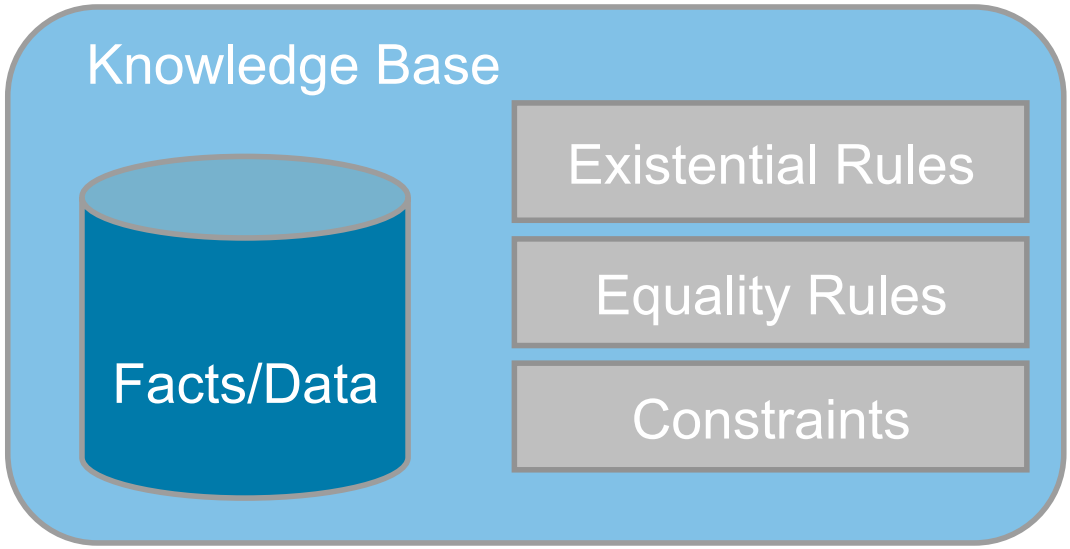
Ontological knowledge expressed by existential rules



∀x ∀y (siblingOf(x,y) → ∃ z (parentOf(z,x) ∧ parentOf(z,y)))

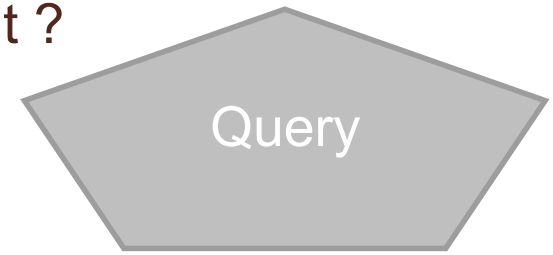
- Same as high-level integrity constraints in databases called Tuple Generating Dependencies
- See also the recent framework Datalog+
- Same as the logical translation of conceptual graph rules
- Generalize the core of new description logics used for query answering ($\mathcal{EL}$, DL-Lite, Horn DLs)

Logical / graphical framework



Entailment ?

$\vdash$



Basic problem

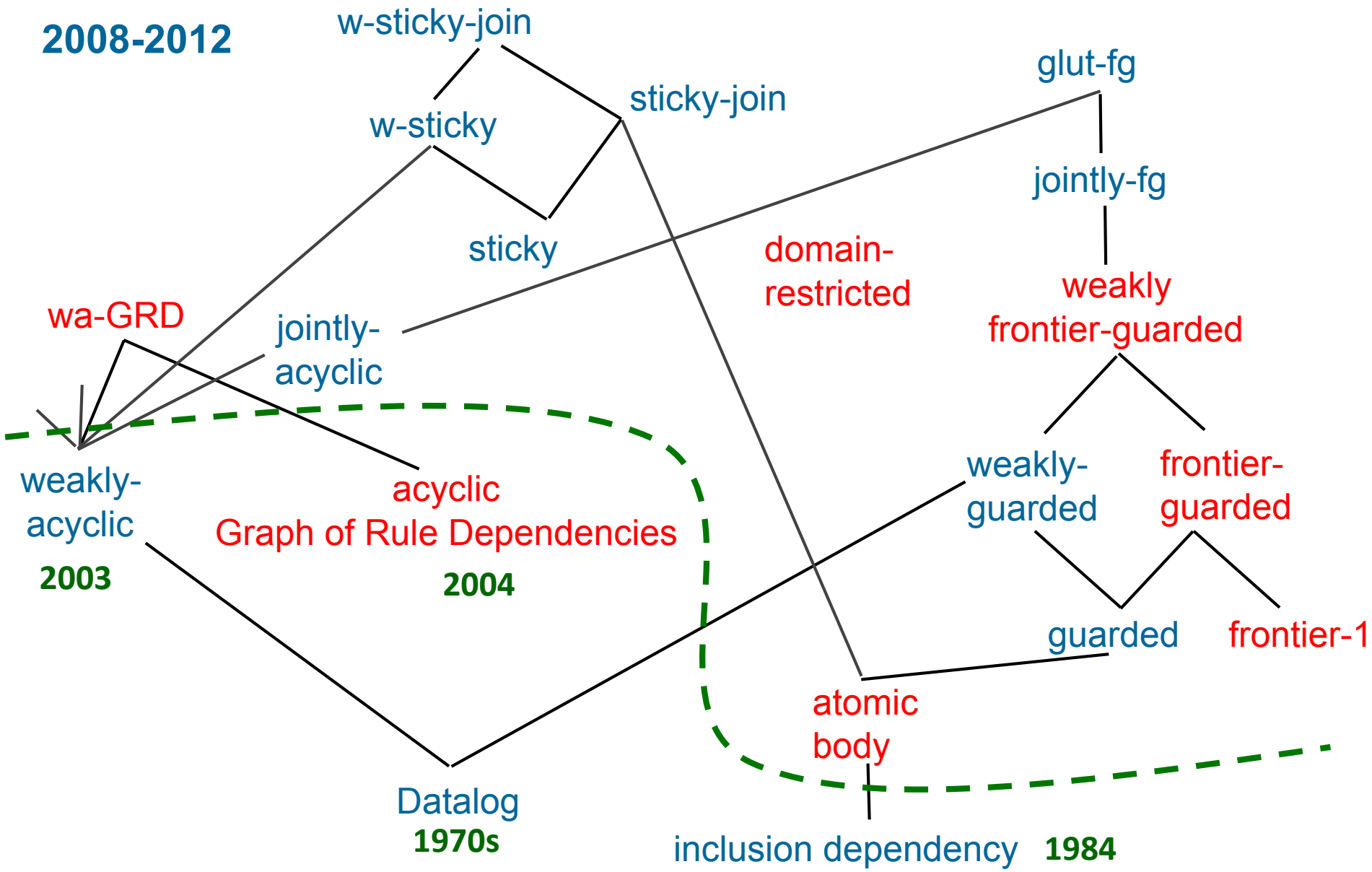
Input: Knowledge Base *K* and Conjunctive Query *Q*

Question: is *Q* entailed by *K* (does *K* provide an answer to *Q*) ?

This problem is **undecidable** (semi-decidable)

Challenge: find decidable subclasses of rules with good expressivity / tractability tradeoff

The (ever growing?) map of decidable classes



Main achievements (2010-2011)

- « Walking the Decidability Line … »
  - Abstract criteria for decidability
  - New (and easily recognizable) decidable classes
  - Results on the combination of decidable classes
- « Walking the Complexity Lines … »
  - Analysis in terms of combined / data complexity
  - Some of our classes have polynomial data complexity

Main publications
[IJCAI 2009]
KR 2010 (Principles of Knowledge Representation)
Artificial Intelligence 2011
IJCAI 2011
Keynote talk at RR 2011 (Reasoning the Web)

- On inclusion of conjunctive queries with negation:
  Experimental results (DEXA'10 and 11)
  Complexity results (submitted to a journal)

Perspectives (four years)

- Deepen the analysis of this framework
  Build a unified framework
  Extend it
  Find smart ways of allowing some difficult-to-process but desired kinds of rules (e.g. equality rules)

- Build scalable Algorithms
  Polynomial data complexity is not enough

  Related work: Study and evaluate data storage techniques w.r.t. this framework

- Tools implementing the algorithms and validation on applications

- Collaborations: currently with Karlsruhe IT
  → Build a collaboration at a European level
  → Associate complementary competencies
  (description logics, rule-based languages, databases, graphs)

Axis 2: Coping with imperfect knowledge

- Start from application needs
  Local context: IATE (Agronomy Laboratory in Montpellier) / INRA
  Aim: integrate knowledge about agrifood chains
  and use it to support decision making

- Agrifood chain characteristics:
  - No global mathematical models
  - Few experimental results,
    importance of expert knowledge to meet this lack

- Identified difficulties:
  - Process experimental and expert knowledge
    with varying precision and fiability
  - Arbitrate in presence of conflicting viewpoints / criteria
    while being able to explain the decision

Focus on argumentation and preferences

- Case study: Controversy about the composition of French bread flour
  - Various types of criteria: environmental, economical, functional, sanitary…
  - Various actors: millers, bakers, consumers, Ministry of Health

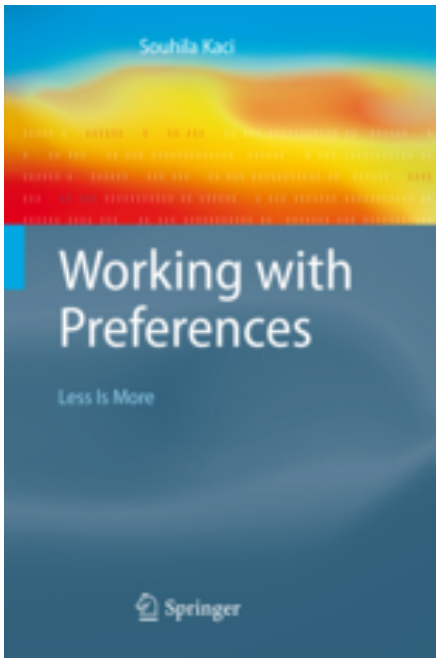
  Expertise in agronomy (cereal processing) provided by INRA/IATE

- Led us to argumentation systems (emerging field in AI)
  and preferences

  First results: PhD thesis
  (co-supervised with IRIT)

- Recruitement of a professor to lead this axis

  Expertise on preferences
  argumentation systems
  decision
  inconsistency handling …



Perspectives (four years)

- A framework for argumentation-based decision making
  Validation on agrifood chain analysis
  - Take into account preferences on arguments + different contexts or viewpoints (preference relations on arguments vary according to the context)
  - Aggregate contextual preferences
  - Integrate preferences in the decision making process
  - Represent the strength of an attack
  - Consider other interactions between arguments (e.g. « support »)
  - Define a logical framework for representing arguments and inferring interactions between them
  - Take into account the dynamicity of argumentative systems
- Use of preferences in argumentation and decision, but also in query answering, reasoning in presence of inconsistencies, …
  (in relationship with languages studied in Axis 1)

Axis 3: Integration of theoretical tools …

Focus on semantic data integration

- International Context: Linked Open Data

  → Many initiatives to interconnect public bibliographic data

  Go from specific formats to Semantic Web languages

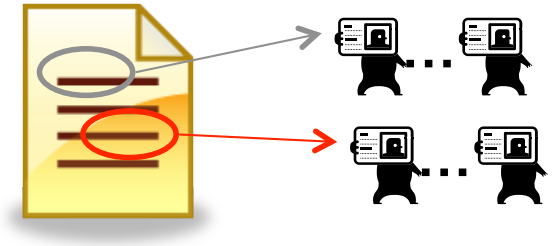
  e.g. WorldCat catalog
  Virtual International Authority File (VIAF) Initiative
  led by OCLC (Online Computer Library Center)

- ABES (Bibliographic Agency for French Universities): a strategic partner
  - Involved in WorldCat and VIAF
  - Will play a key role in building a French hub of bibliographic data

- SUDOC: main ABES base
  (collective catalog of French academic libraries)

  2000 libraries
  10 M bibliographic notices, 2 M authority notices (e.g. on authors)

Record linkage problem



ABES objective: Hub of authority notices for other bibliographic bases

Preliminary work (2010-2011)

1. Formalization of SUDOC
   Ontology in RDFS+OWL compatible with document description standards
   Required expertise in Library and Information Science (ABES)
2. Export SUDOC bases to Semantic Web formats
3. Prototype of an entity identification service (for author names)

→Link reliability as a requirement to solve linkage problems

Perspectives (four years)

Qualinca: « Quality and Interoperability of Large Catalogues of Documents»
Submitted to ANR Call Contint (Sept. 2011)

Leader: GraphIK

Partners: INRIA LEO, LIG, ABES,
INA (National Institute for Audiovisual Archives)

Objectives: Improve the quality of document bases by

- detecting and repairing linkage errors
- detecting and fusionning duplicate authority notices
- enriching authority notices to allow disambiguation by a person
- explicitly representing the reliability of a link

Expected results:

- A KR framework allowing to formalize quality problems
- Methods/tools to improve the quality of a base
  (integrate logical and numerical approaches)
- A prototype and validation of the proposed tools on ABES and INA bases