Knowledge Representation and Ontologies

Part 1: Modeling Information through Ontologies

Diego Calvanese

Faculty of Computer Science
Master of Science in Computer Science

A.Y. 2011/2012

Freie Universität Bozen
Libera Università di Bolzano
Free University of Bozen · Bolzano

Part 1

Modeling Information through Ontologies

Outline of Part 1

Outline of Part 1

Outline of Part 1

New challenges in information management

One of the key challenges in complex systems today is the management of information:

- The **amount** of information has increased enormously.
- The **complexity** of information has increased:
  structured $\rightsquigarrow$ semi-structured $\rightsquigarrow$ unstructured
- The underlying data may be of **low quality**, e.g., incomplete, inconsistent, not *crisp*.
- Information is increasingly **distributed** and **heterogeneous**, but nevertheless needs to be accessed in a uniform way.
- Information is consumed not only by humans, but also by machines.

Traditional data management systems are not sufficient anymore to fulfill today's information management requirements.

Addressing information management challenges

Several efforts come from the database area:

- New kinds of databases are studied, to manage semi-structured (XML), and probabilistic data.
- Information integration is one of the major challenges for the future or IT. E.g., the market for information integration software has been estimated to grow from $2.5 billion in 2007 to $3.8 billion in 2012 (+8.7% per year) [IDC. Worldwide Data Integration and Access Software 2008-2012 Forecast. Doc No. 211636 (2008)].

On the other hand, management of complex kinds of information has traditionally been the concern of **Knowledge Representation** in AI:

- Research in AI and KR can bring new insights, solutions, techniques, and technologies.
- **However, what has been done in KR needs to be adapted / extended / tuned to address the new challenges coming from today's requirements for information management.**

Description logics

Description Logics [Baader *et al.*, 2003] are an important area of KR, studied for the last 25 years, that provide the foundations for the structured representation of information:

- By grounding the used formalisms in logic, the information is provided with a **formal semantics** (i.e., a meaning).
- The logic-based formalization allows one to provide **automated support** for tasks related to data management, by means of **logic-based inference**.
- **Computational aspects** are of concern, so that **tools** can provide **effective support** for automated reasoning.

In this course we are looking into using description logics for data management.

Ontologies

Description logics provide the formal foundations for ontology languages.

Def.: Ontology

is a representation scheme that describes a **formal conceptualization** of a domain of interest.

The specification of an ontology usually comprises two distinct levels:

- **Intensional level**: specifies a set of conceptual elements and of constraints/axioms describing the conceptual structures of the domain.
- **Extensional level**: specifies a set of instances of the conceptual elements described at the intensional level.

Note: an ontology may contain also a **meta-level**, which specifies a set of modeling categories of which the conceptual elements are instances.

Outline of Part 1

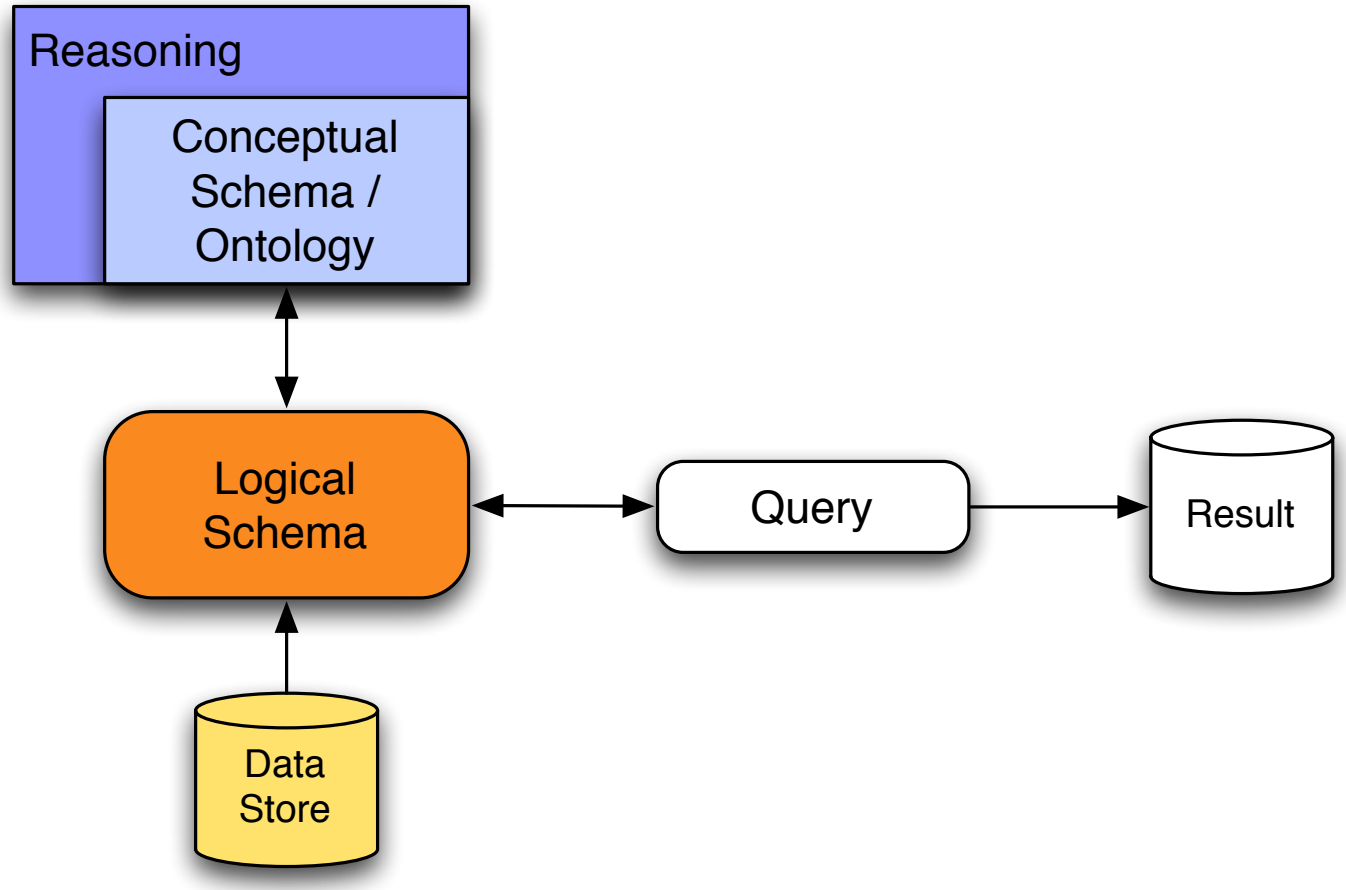
Conceptual schemas in information systems

Intensional information has traditionally played an important role in information systems.

Design phase of the information system:

1. From the requirements, a **conceptual schema** of the domain of interest is produced.
2. The conceptual schema is used to produce the logical data schema.
3. The data are stored according to the logical schema, and queried through it.

Conceptual schemas used at design-time



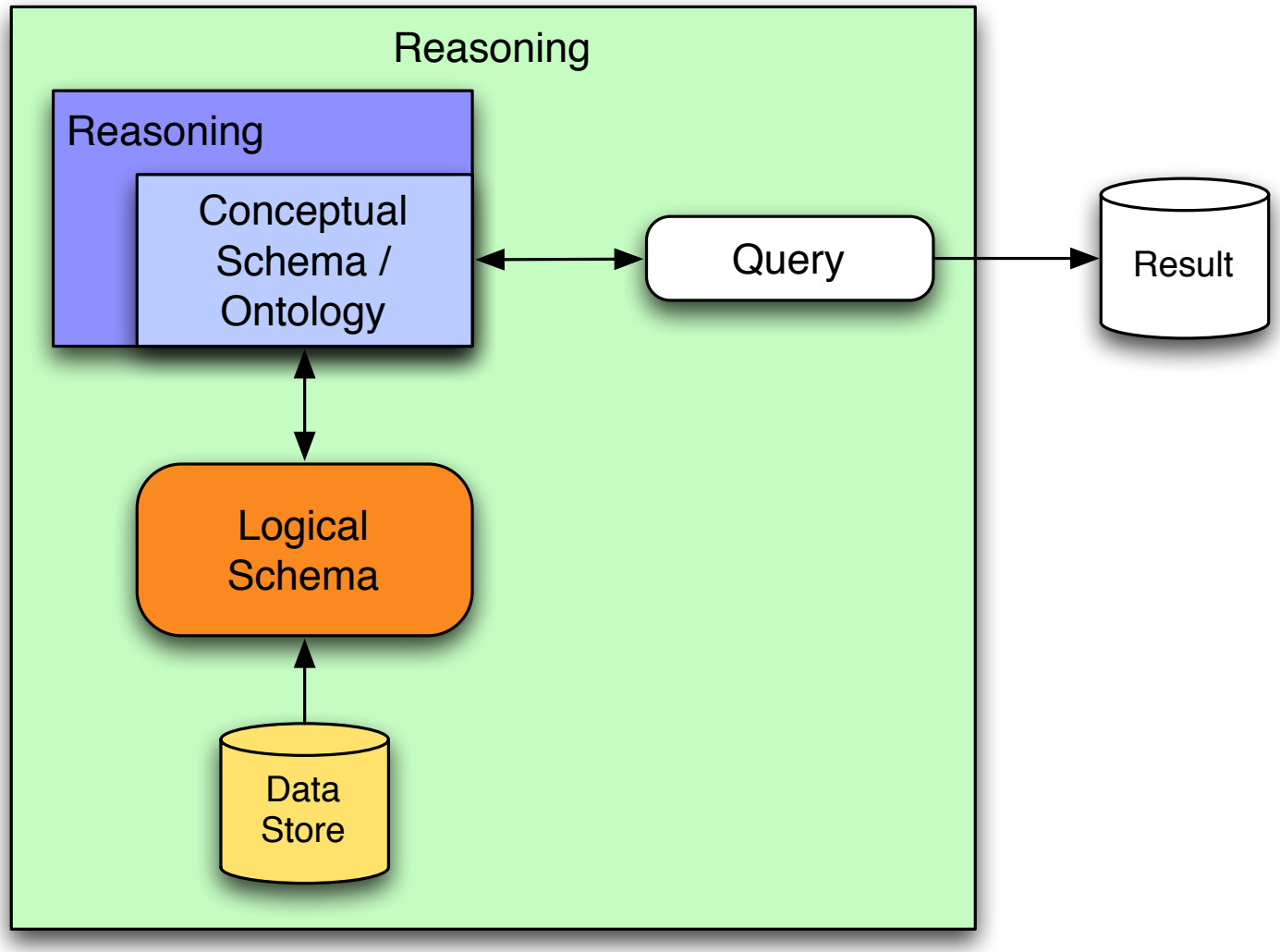
Ontologies in information systems

The role of ontologies in information systems goes beyond that of conceptual schemas.

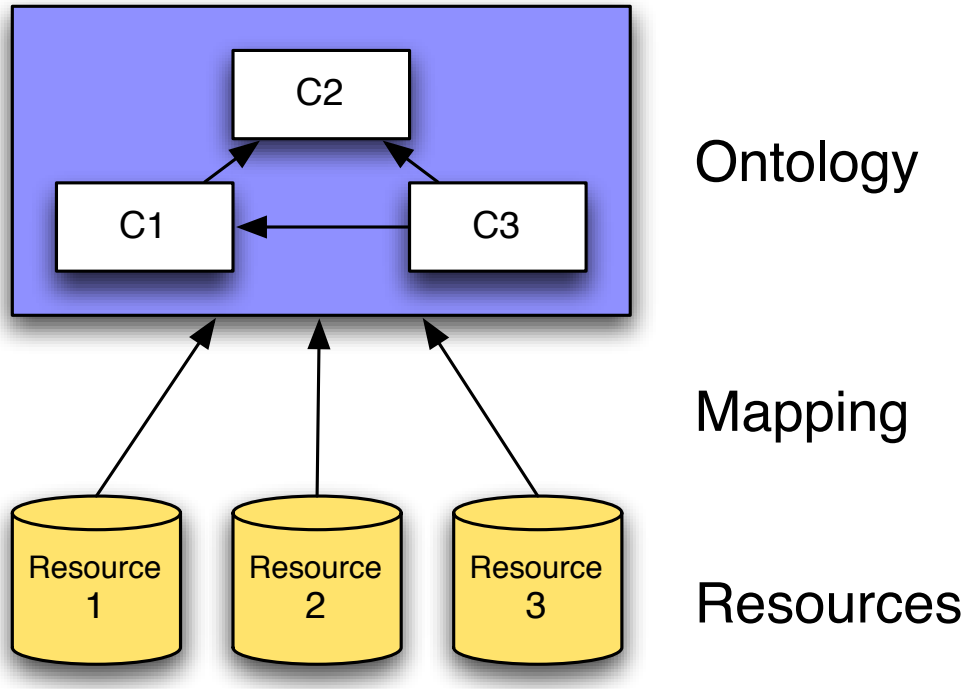
Ontologies affect the whole life-cycle of the information system:

- Ontologies, with the associated reasoning capabilities and inference tools, can provide support at design time.
- The use of ontologies can significantly simplify **maintenance** of the information system's data assets.
- The ontology is used also to support the interaction with the information system, i.e., at run-time.
  $\rightsquigarrow$ **Reasoning** to take into account the constraints coming from the ontology has to be **done at run-time**.

Ontologies used at run-time



Ontologies at the core of information systems



The usage of all system resources (data and services) is done through the domain conceptualization.

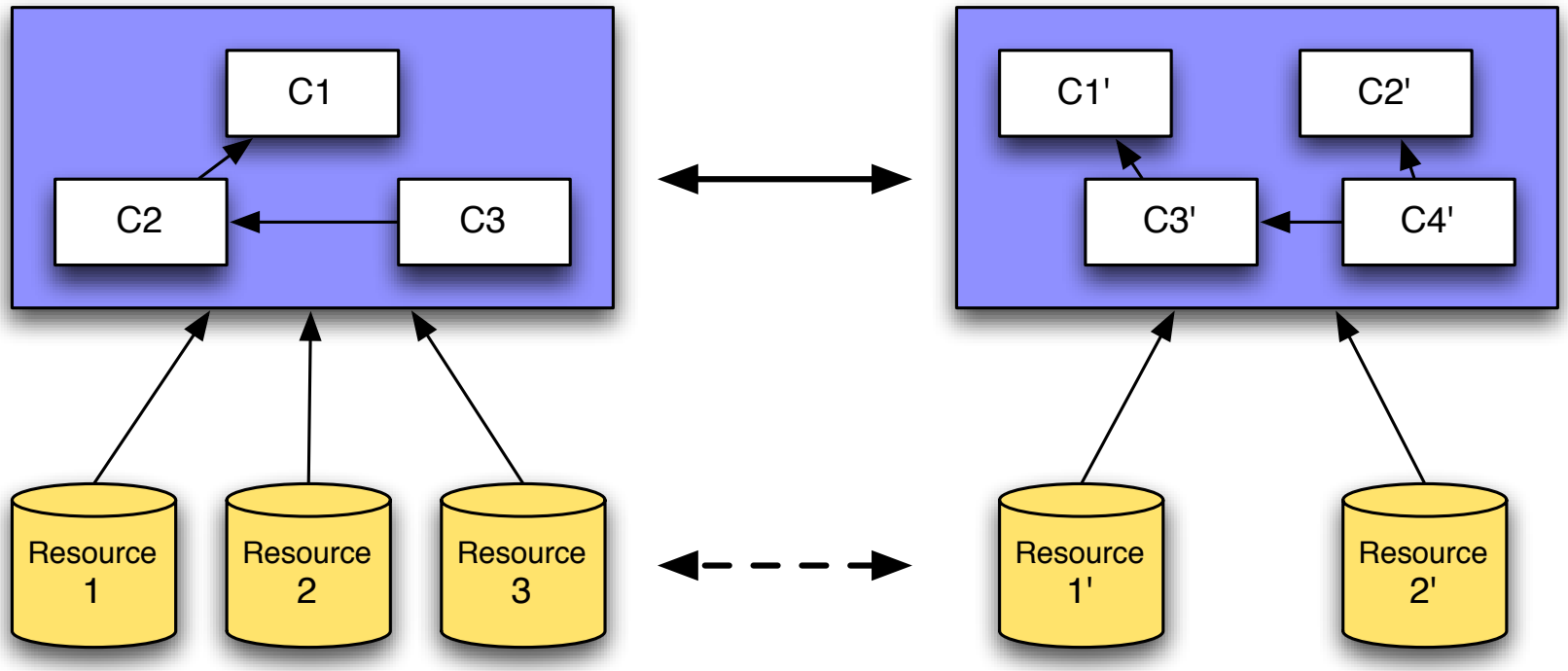
Ontology mediated access to data

Desiderata: achieve **logical transparency** in access to data:

- **Hide** to the user where and how data are stored.
- Present to the user a **conceptual view** of the data.
- Use a **semantically rich formalism** for the conceptual view.

This setting is similar to the one of Data Integration. The difference is that here the ontology provides a rich conceptual description of the data managed by the system.

Ontologies at the core of cooperation



The cooperation between systems is done at the level of the conceptualization.

Outline of Part 1

Issues in ontology-based information management

1. Choice of the formalisms to adopt
2. Efficiency and scalability
3. Tool support

Issue 1: Formalisms to adopt

1. Which is the right ontology language?
   - many proposals have been made
   - differ in expressive power and in complexity of inference
2. Which languages should we use for querying?
   - requirements for querying are different from those for modeling
3. How do we connect the ontology to available information sources?
   - mismatch between information in an ontology and data in a data source

In this course:

- We present and discuss variants of ontology languages, and study their logical and computational properties.
- We study the problem of querying data through ontologies.
- We discuss problems and solutions related to the impedance mismatch between ontologies and data sources.

Issue 2: Efficiency and scalability

- How can we handle large ontologies?
  - We have to take into account the **tradeoff** between **expressive power** and **complexity** of inference.
- How can we cope with large amounts of data?
  - What may be good for large ontologies, may not be good enough for large amounts of data.
- Can we handle multiple data sources and/or multiple ontologies?

In this course:

- We discuss in depth the above mentioned **tradeoff**.
- We will also pay attention to the aspects related to **data management**.
- We do not deal with the problem of integrating multiple information sources. See the course on *Information Integration*.

Issue 3: Tools

- According to the principle that "there is no meaning without a language with a formal semantics", the formal semantics becomes the solid basis for dealing with ontologies.
- Hence every kind of access to an ontology (to extract information, to modify it, etc.), requires to **fully** take into account its semantics.
- We need tools that perform reasoning over the ontology that is **sound and complete** wrt the semantics.
- The tools have to be as "efficient" as possible.

In this course:

- We discuss the requirements, the principles, and the theoretical foundations for ontology inference tools.
- We also present and use a tool for querying data sources through ontologies that has been built according to those principles.