Temporalizing Ontology-Based Data Access

Franz Baader

Stefan Borgwardt, Marcel Lippmann

Veronica Thost

TECHNISCHE UNIVERSITÄT DRESDEN

DRESDEN concept

DFG Deutsche Forschungsgemeinschaft

HAEC

Goal

Achieve higher-level Situation Awareness

Projection — Projecting how events may evolve over time, guess future behaviour

↑

Comprehension — Higher-level composite view of situation, at semantic level of human comprehension

forrest

↑

Perception — Sensors, Radar, Pictures, Videos, Eyewitness reports by humans

lots of trees

[Endsley; 1995]

Levels of Situation Awareness

Projection — danger of myocardial infarction

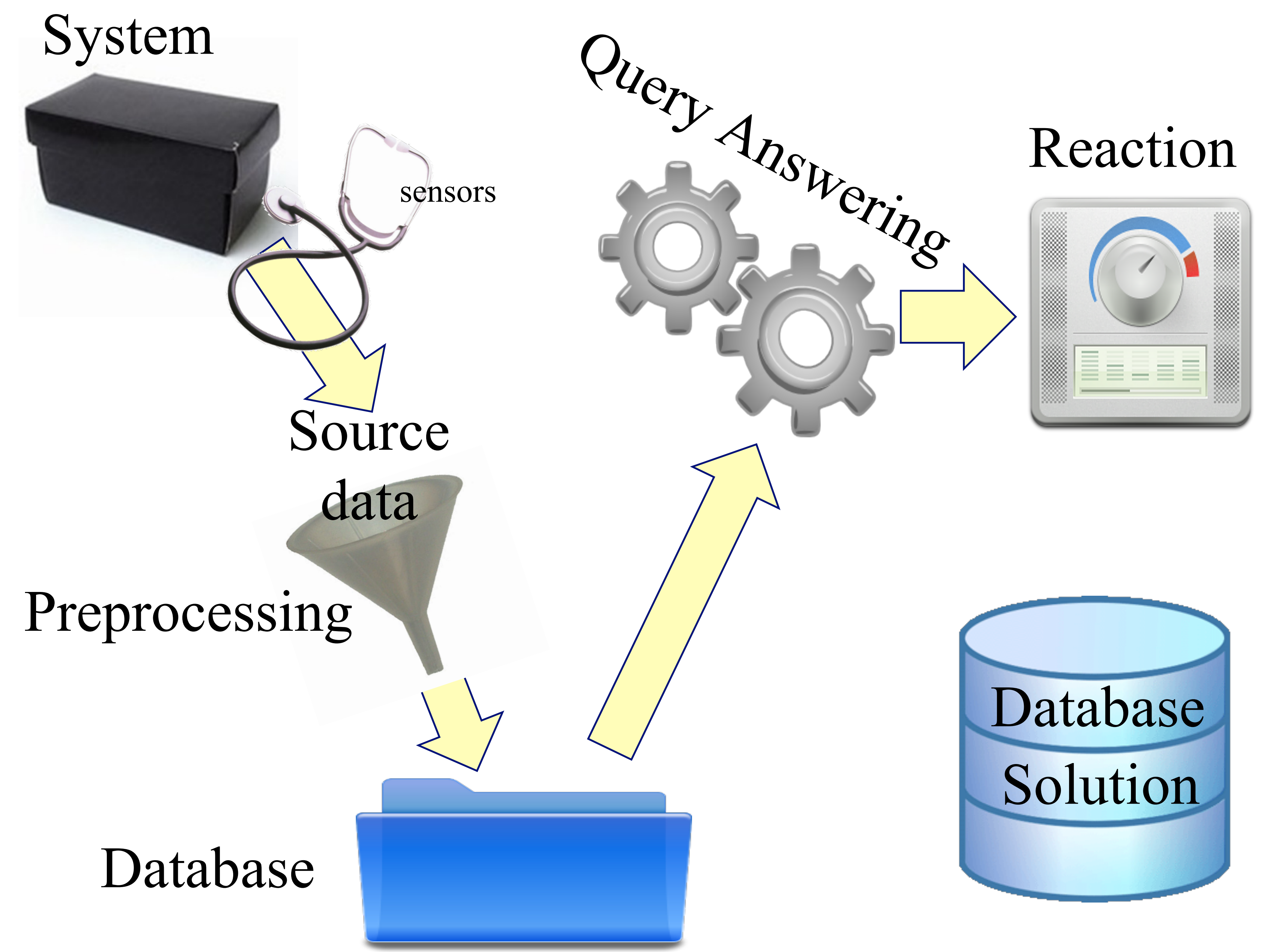
↑

Comprehension — medical ontology, patient record; hypertension, history of hypertension

↑

Perception — systolic pressure is high, red face

Dresden



System
sensors
Query Answering
Reaction
Source data
Preprocessing
Database
Database Solution

Conjunctive queries

important subclass of SQL queries:
select-project-join queries

A conjunctive query (CQ)

is an existentially quantified conjunction of atoms:

$$\exists y.\mathit{Male}(x) \land \mathit{history}(x,y) \land \mathit{Hypertension}(y)$$

Complexity of CQs:
deciding whether there is an answer tuple is

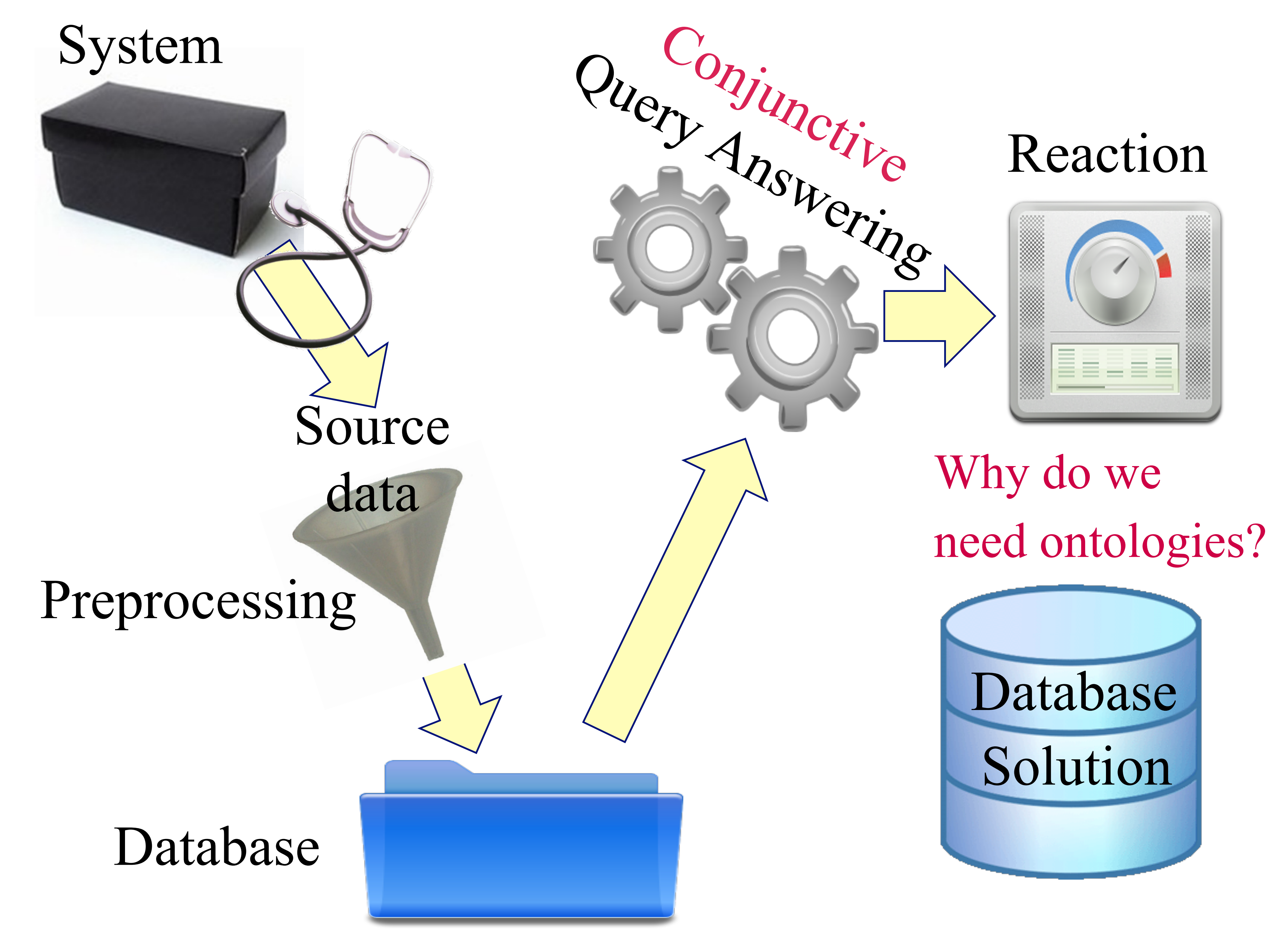
- NP-complete w.r.t. combined complexity.
- In AC^0 w.r.t. data complexity.

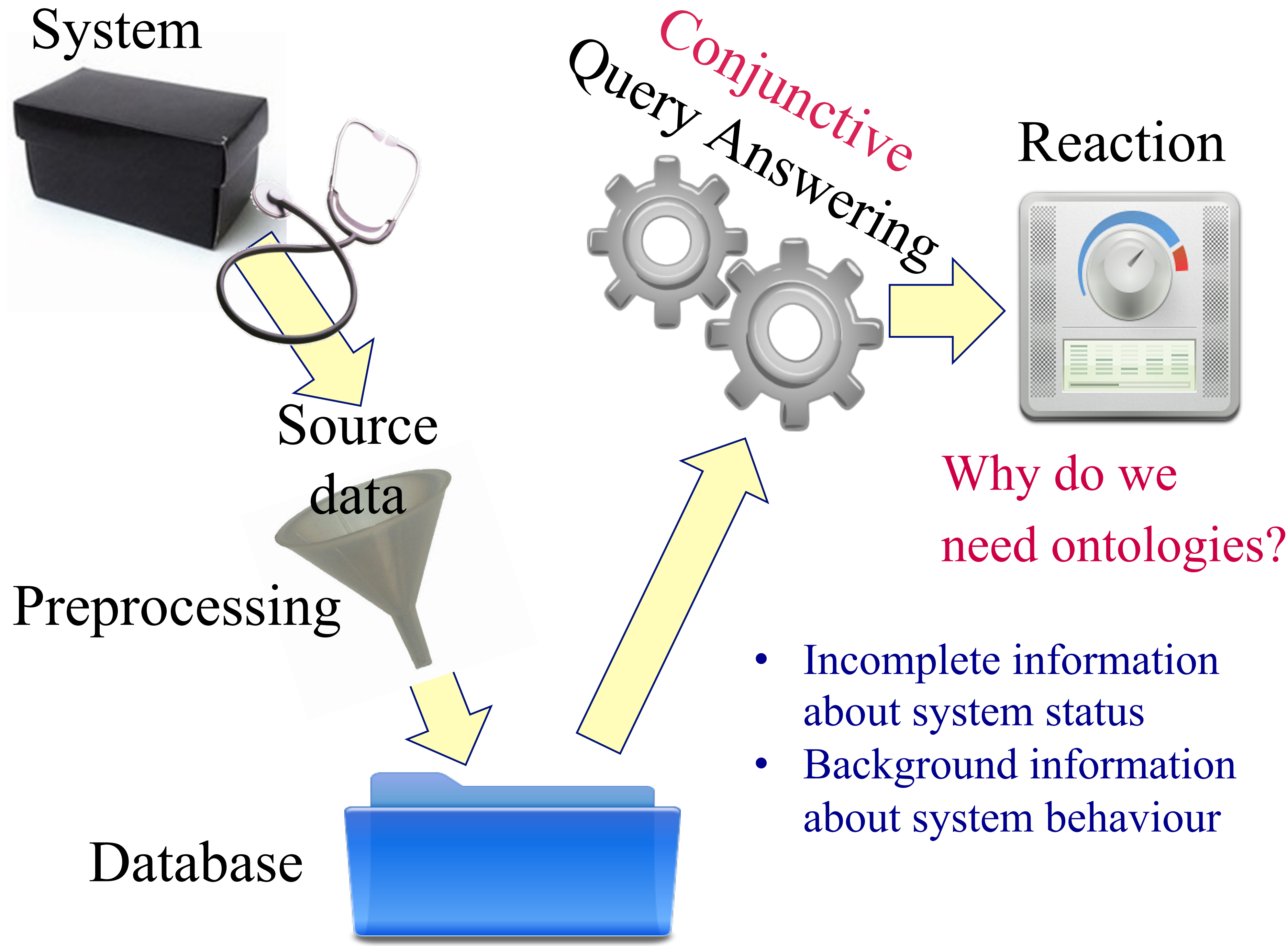
complexity measured w.r.t. size of data and query

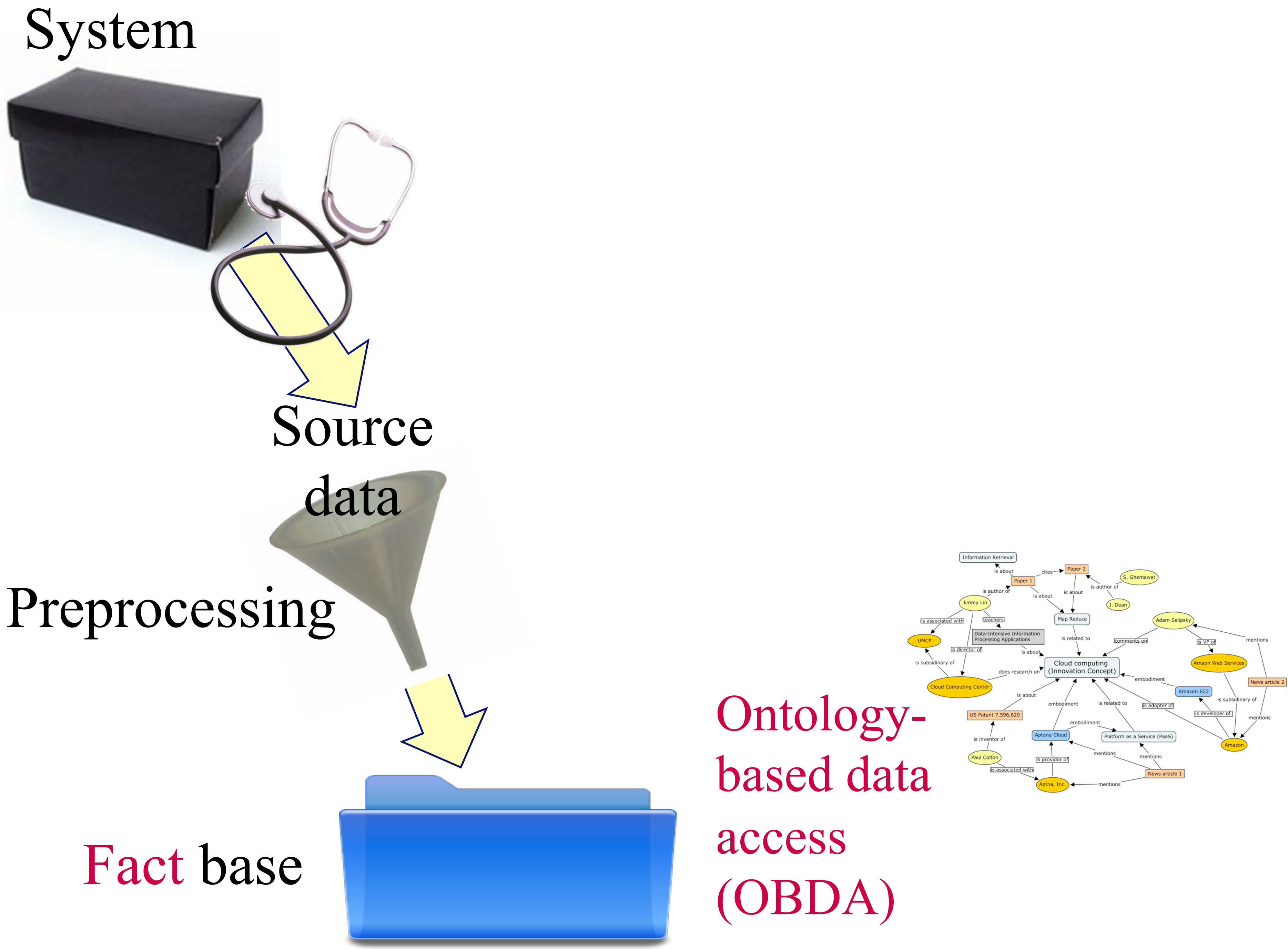
complexity measured w.r.t. size of data only

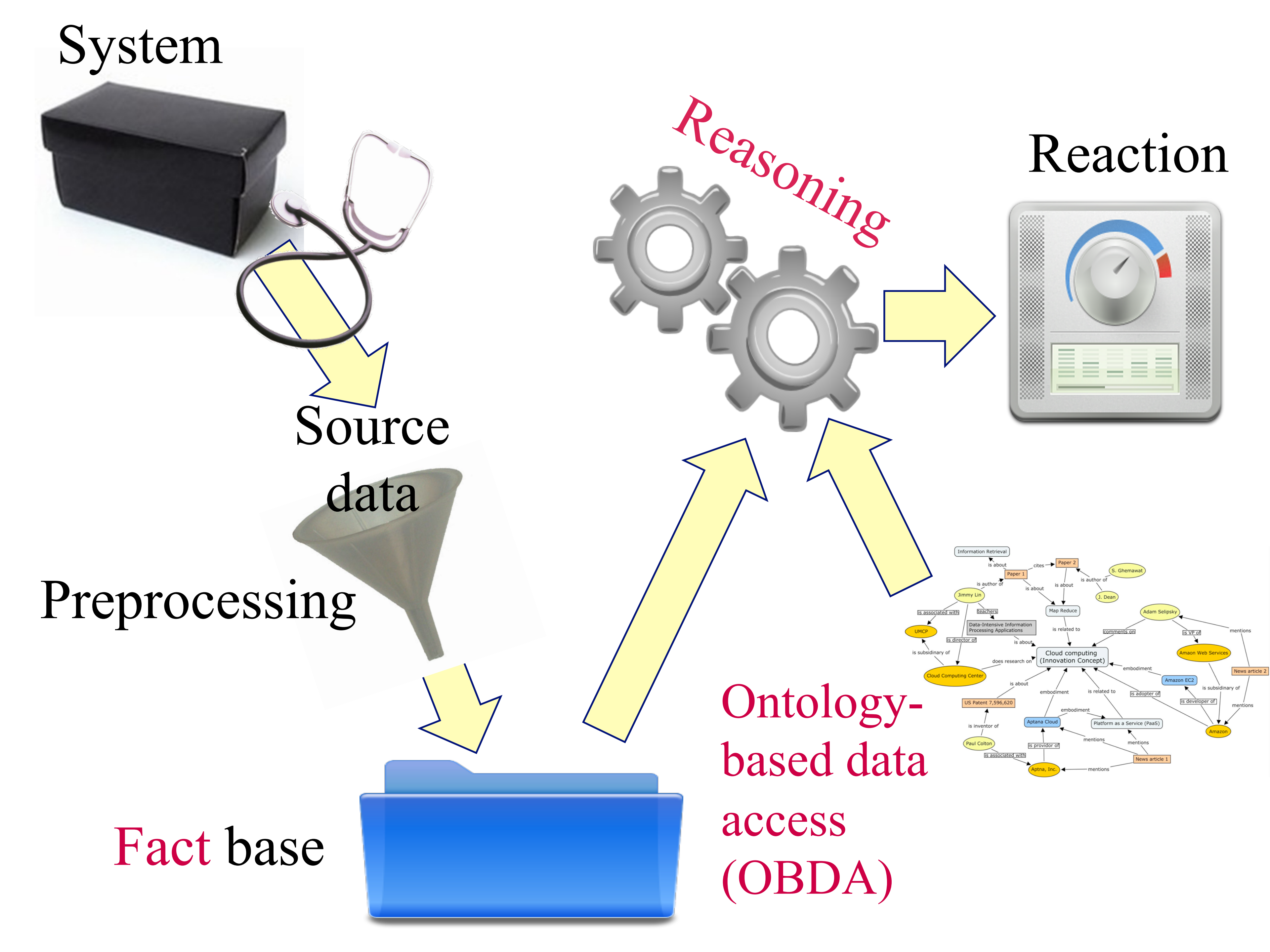
$$\mathrm{AC}^0 \subset \mathrm{LogSpace} \subseteq \mathrm{P}$$

Dresden


System
Conjunctive
Query Answering
Reaction
Source
data
Preprocessing
Why do we
need ontologies?
Database
Solution
Database


System
Conjunctive
Query Answering
Reaction
Source
data
Preprocessing
Database
Why do we
need ontologies?
• Incomplete information about system status
• Background information about system behaviour


System
Source data
Preprocessing
Fact base
Ontology-based data access (OBDA)



System
Reasoning
Reaction
Source data
Preprocessing
Fact base
Ontology-based data access (OBDA)
Information Retrieval
Paper 1
Paper 2
S. Ghemawat
J. Dean
Jimmy Lin
Map Reduce
UMCP
Data-Intensive Information Processing Applications
Adam Selipsky
Cloud computing (Innovation Concept)
Amaon Web Services
Cloud Computing Center
Amazon EC2
News article 2
US Patent 7,596,620
Aptana Cloud
Platform as a Service (PaaS)
Amazon
Paul Colton
News article 1
Aptna, Inc.

Ontology-based data access

Generalizes answering queries over databases in two directions:

- Presence of an ontology $\mathcal{O}$:
  predicates used in the queries are constrained by ontology axioms

- Incompleteness:
  Queries evaluated over a fact base $\mathcal{A}$ rather than a database
  (no closed-word assumption)

We want to compute certain answers of $\phi(x_1, \ldots, x_n)$ over $\mathcal{A}$ w.r.t. $\mathcal{O}$:

- a tuple $(a_1, \ldots, a_n)$ of individuals occurring in $\mathcal{A}$ such that
- $(a_1, \ldots, a_n)$ is an answer tuple of $\phi(x_1, \ldots, x_n)$ w.r.t. $\mathcal{D}$
- for all databases $\mathcal{D}$ that satisfy the axioms in $\mathcal{O}$ and the facts in $\mathcal{A}$.

Ontology-based data access

$\mathcal{O}$

$$\exists \textit{systolic\_pressure}.\textit{High\_pressure} \sqsubseteq \exists \textit{finding}.\textit{Hypertension}$$

$$\exists \textit{finding}.\textit{Hypertension} \sqcap \exists \textit{history}.\textit{Hypertension} \sqsubseteq \exists \textit{risk}.\textit{Myocardial\_infarction}$$

$\mathcal{A}$

$$\textit{systolic\_pressure}(\textit{BOB}, \textit{P1})$$
$$\textit{High\_pressure}(\textit{P1})$$
$$\textit{history}(\textit{BOB}, \textit{H1})$$
$$\textit{Hypertension}(\textit{H1})$$
$$\textit{Male}(\textit{BOB})$$

$$\exists y.\textit{Male}(x) \wedge \textit{risk}(x, y) \wedge \textit{Myocardial\_infarction}(y)$$

BOB is a certain answer of this query over $\mathcal{A}$ w.r.t. $\mathcal{O}$.

Dresden

Ontology-based data access

Investigated in detail for the case where:

- the ontology is expressed using an appropriate description logic,
- the query is a conjunctive query.

For the expressive DL $\mathcal{ALC}$, deciding whether there is a certain answer is

- ExpTime-complete w.r.t. combined complexity.
- coNP-complete w.r.t. data complexity.

For the inexpressive DLs of the DL-Lite family, deciding whether there is a certain answer is

- NP-complete w.r.t. combined complexity.
- in AC^0 w.r.t. data complexity.

Ontology-based data access

Why temporalize OBDA?

To recognize certain situations, we need to consider the system status at different points in time.

> Example: recognize "history of hypertension" instead of assuming that it is directly stated in the patient record.

Set-up considered in this talk:

- instead of a single fact base consider a finite sequence of fact bases
- background knowledge is encoded in a global $\mathcal{ALC}$- or DL-Lite-ontology
- temporalized conjunctive queries obtained from propositional linear temporal logic (LTL) by replacing propositional variables by conjunctive queries

Ontology-based data access

temporalized OBDA

To recognize certain situations, we need to consider the system status at different points in time.

Example: recognize "history of hypertension" instead of assuming that it is directly stated in the patient record.

$$\underbrace{\mathit{Male}(x)}_{p} \wedge \bigcirc^{-}\Diamond^{-} \underbrace{(\exists y.\mathit{finding}(x,y) \wedge \mathit{Hypertension}(y))}_{q}$$

TCQ

$$p \wedge \bigcirc^{-}\Diamond^{-} q$$

- temporalized conjunctive queries obtained from propositional linear temporal logic (LTL) by replacing propositional variables by conjunctive queries

Dresden

Ontology-based data access

temporalized OBDA

To recognize certain situations, we need to consider the system status at different points in time.

> Example: recognize "history of hypertension" instead of assuming that it is directly stated in the patient record.

$$\underbrace{\mathit{Male}(x)}_{p} \wedge \bigcirc^{-}\Diamond^{-} \underbrace{(\exists y.\mathit{finding}(x,y) \wedge \mathit{Hypertension}(y))}_{q} \qquad \text{TCQ}$$

$$p \wedge \bigcirc^{-}\Diamond^{-} q$$

Rigid versus flexible concepts and roles:

- flexible symbols can change their interpretations over time,
- whereas rigid ones cannot.

Temporalized OBDA

our results for $\mathcal{ALC}$

Complexity of TCQ entailment: is a given tuple a certain answer for a given TCQ w.r.t. a sequence of fact bases and a global $\mathcal{ALC}$-ontology?

Combined complexity of TCQ entailment w.r.t. $\mathcal{ALC}$-ontology

- ExpTime-complete without rigid symbols
- co-NExpTime-complete w.r.t. rigid concepts
- 2-ExpTime-complete w.r.t. rigid concepts and roles

same as without temporal operators

Data complexity of TCQ entailment w.r.t. $\mathcal{ALC}$-ontology

- coNP-complete without rigid symbols
- coNP-complete w.r.t. rigid concepts
- co-NP-hard / in EXPTIME w.r.t. rigid concepts and roles

same as without temporal operators

Temporalized OBDA

How to show the results?

Hardness results:

- Data complexity: reduction from the atemporal case
- Combined complexity: reduction from unsatisfiability in $\mathcal{ALC}$-LTL [B., Ghilardi, Lutz; 2008]

Upper bounds: adapt satisfiability algorithms for $\mathcal{ALC}$-LTL.
Split problem into

- satisfiability problem for propositional LTL,
- entailment problem for atemporal conjunctive queries w.r.t. $\mathcal{ALC}$-ontologies.

More effort needed to deal with CQs instead of $\mathcal{ALC}$-axioms in LTL formula and to get tight data complexity results.

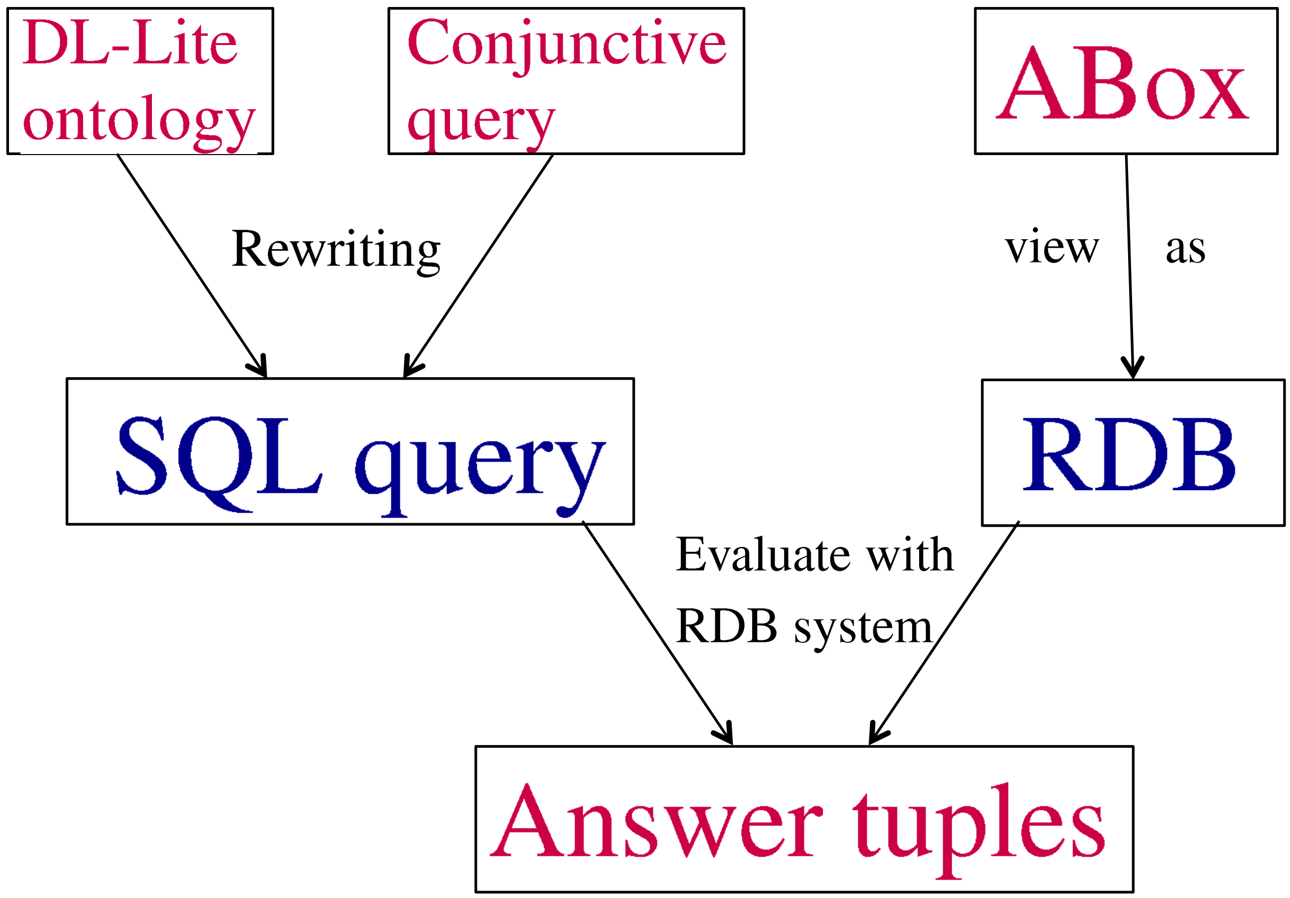
Temporalized OBDA

our results for DL-Lite

Show that the query rewriting approach for DL-Lite also works in the temporal case if the TCQs are negation free.

Query rewriting

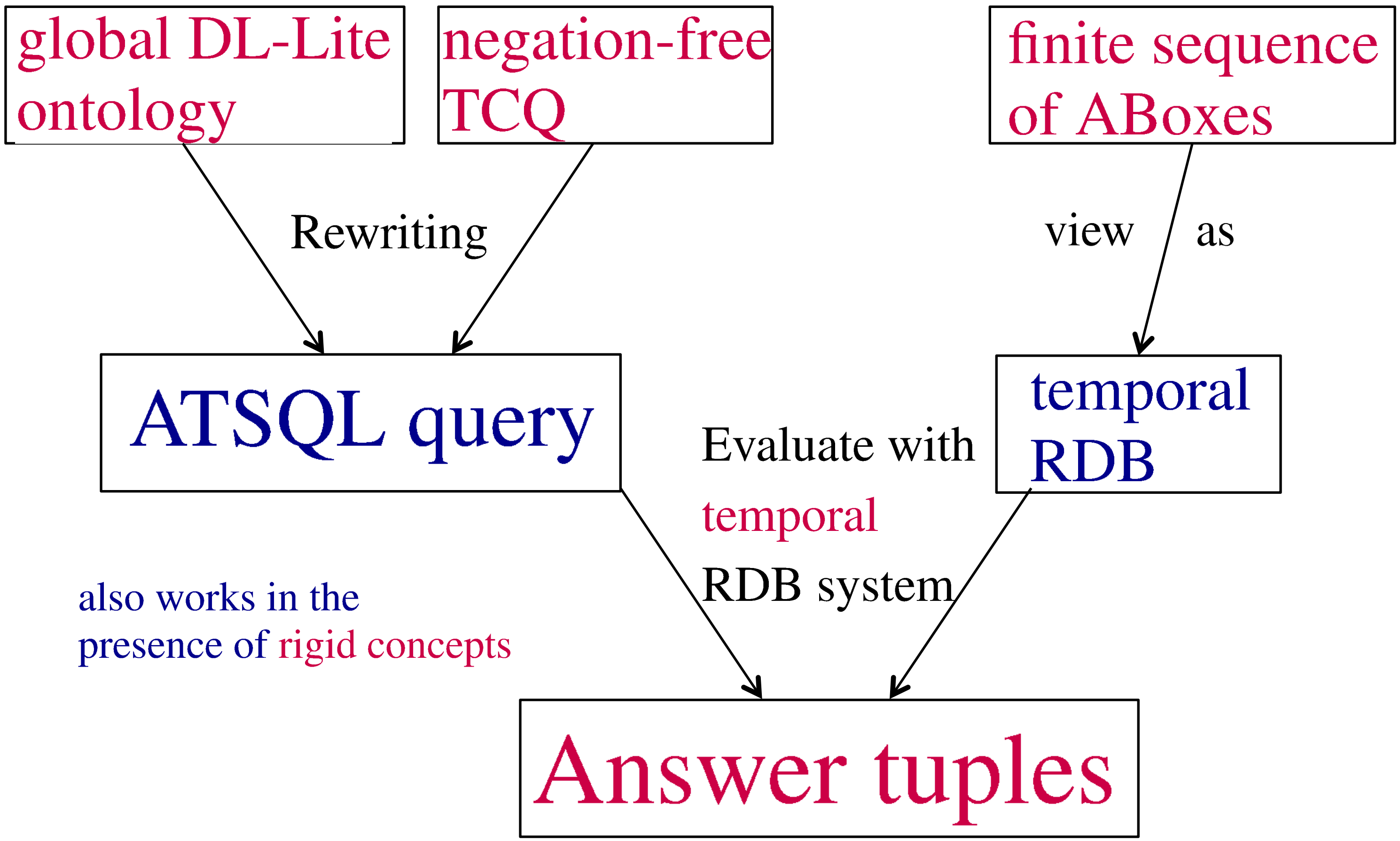
the atemporal case



Temporalized OBDA

our results for DL-Lite

Show that the query rewriting approach for DL-Lite also works in the temporal case if the TCQs are negation free.



also works in the presence of rigid concepts

Conclusion

Our results for $\mathcal{ALC}$ show:

- without rigid symbols, the complexity does not increase
- rigid symbols may increase the combined complexity, but enable a limited projection into the future

For DL-Lite, the rewriting approach works if

- TCQs do not contain negation
- only rigid concepts are allowed

Future work:

- for $\mathcal{ALC}$, close gap for data complexity for rigid concepts and roles
- for DL-Lite, investigate TCQs with negation and the case of rigid roles

Dresden

References

Franz Baader, Stefan Borgwardt, and Marcel Lippmann:

Temporalizing Ontology-Based Data Access.

In Proceedings of the 24th International Conference on Automated Deduction (CADE-24), volume 7898 of Lecture Notes in Artificial Intelligence. Lake Placid, NY, USA, Springer-Verlag, 2013.

Stefan Borgwardt, Marcel Lippmann, and Veronika Thost:

Temporal Query Answering in the Description Logic DL-Lite.

In Proceedings of the 9th International Symposium on Frontiers of Combining Systems (FroCoS 2013), volume 8152 of Lecture Notes in Computer Science. Nancy, France, Springer-Verlag, 2013.