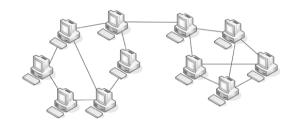
Decentralized query reformulation in DL-Lite (IJCAI 2009)



Marie-Christine Rousset

Université Grenoble & CNRS(LIG)

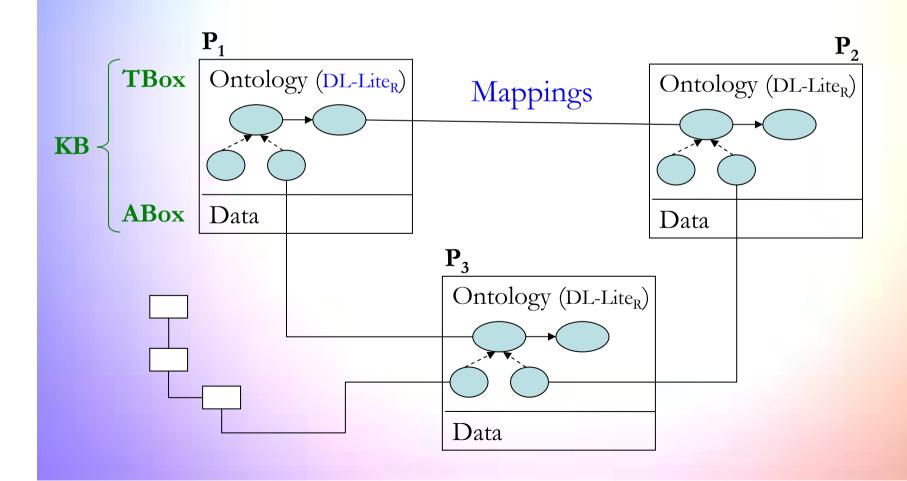
Joint work with Nada Abdallah and François Goasdoué

Université Paris-Sud & CNRS(LRI/IASI) – INRIA (Saclay/LEO)

Setting Decentralized data management for the Semantic Web

Decentralized: Dynamic network of collaborative peers

Semantic Web: Data are described with ontologies (OWL2)

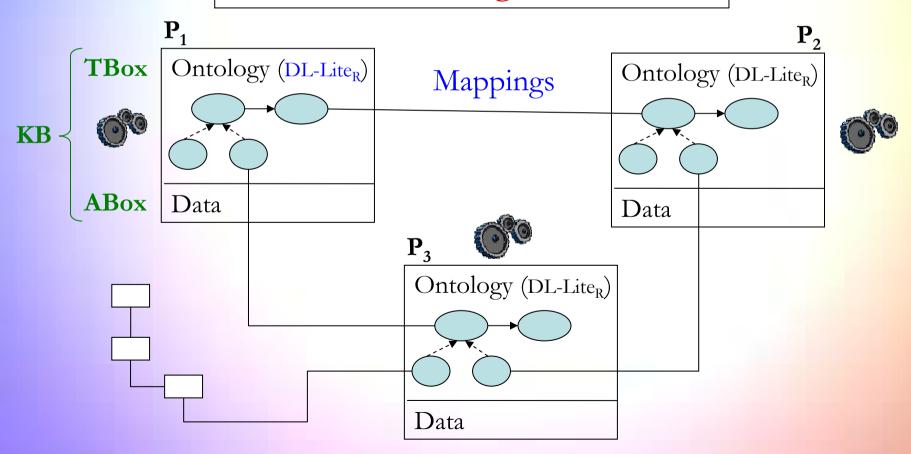


Context Decentralized Data management system for the Semantic Web

Decentralized: Dynamic Network of collaborative peers

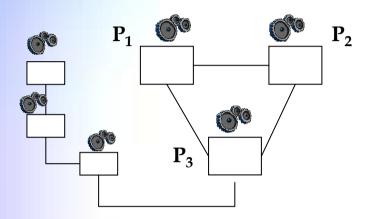
Semantic Web: Data are described with ontologies (OWL2)

Global knowledge unknown



Contributions

1. DL-Lite_R decentralized data model

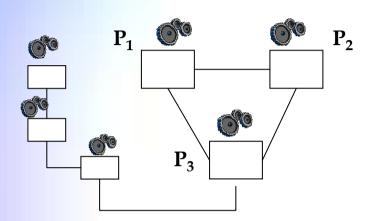


- 2. Decentralized Algorithms for:
 - ✓ Data consistency checking
 - ✓ Query Answering

Contributions

State of the art

1. DL-Lite_R decentralized data model



- 2. Decentralized Algorithms for:
 - ✓ Data consistency checking
 - ✓ Query Answering

• D.Calvenese and al. (JAR 07)

"Tractable reasoning and efficient query answering in description logics"

Centralized DL-Lite_R KB

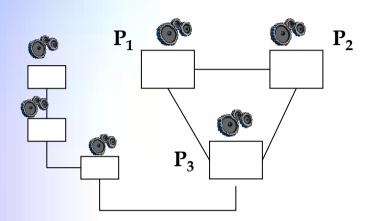
Centralized Algorithms for:

- ✓ Data consistency checking
- ✓ Query Answering

Contributions

State of the art

1. DL-Lite_R decentralized data model



- 2. Decentralized Algorithms for:
 - ✓ Data consistency checking
 - ✓ Query Answering by reformulation

• D.Calvenese and al. (JAR 07)

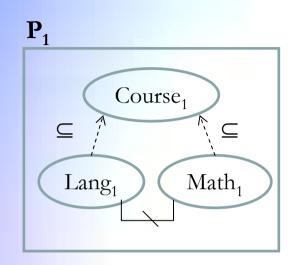
"Tractable reasoning and efficient query answering in description logics"

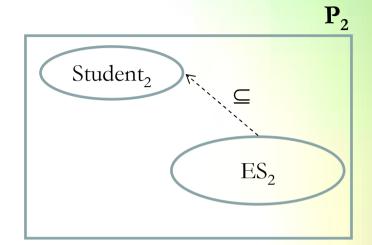
Centralized DL-Lite_R KB

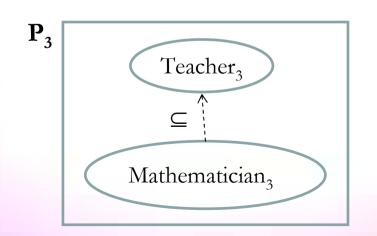
Centralized Algorithms for:

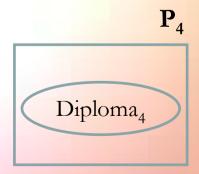
- ✓ Data consistency checking
- ✓ Query Answering by reformulation
- 3. Decentralized & Centralized Algorithms for:
 - ✓ View consistency checking
 - ✓ Query Answering using views

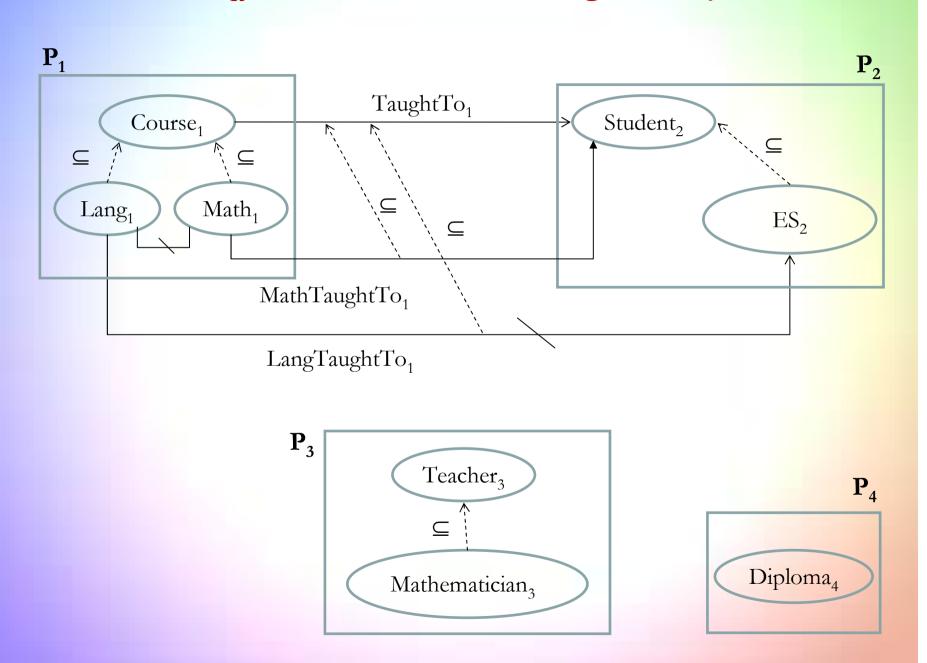
DL-Lite_R decentralized data management: illustration

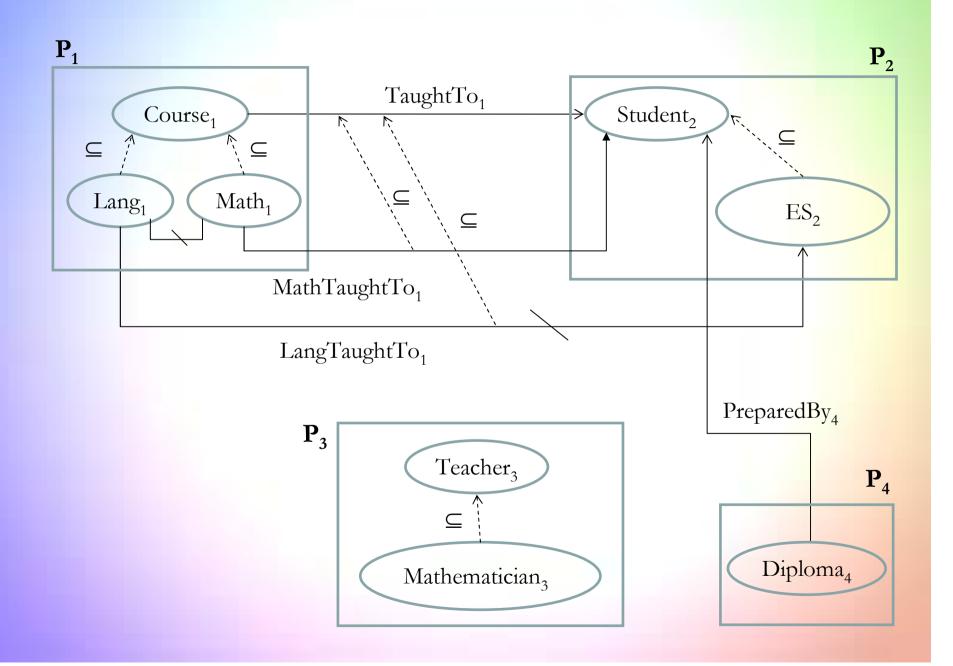


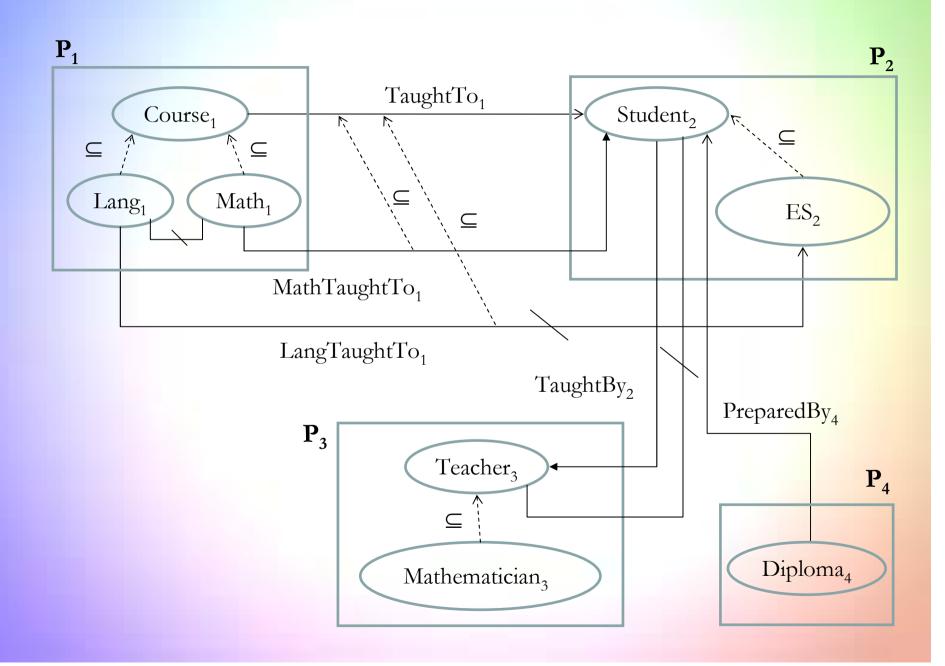


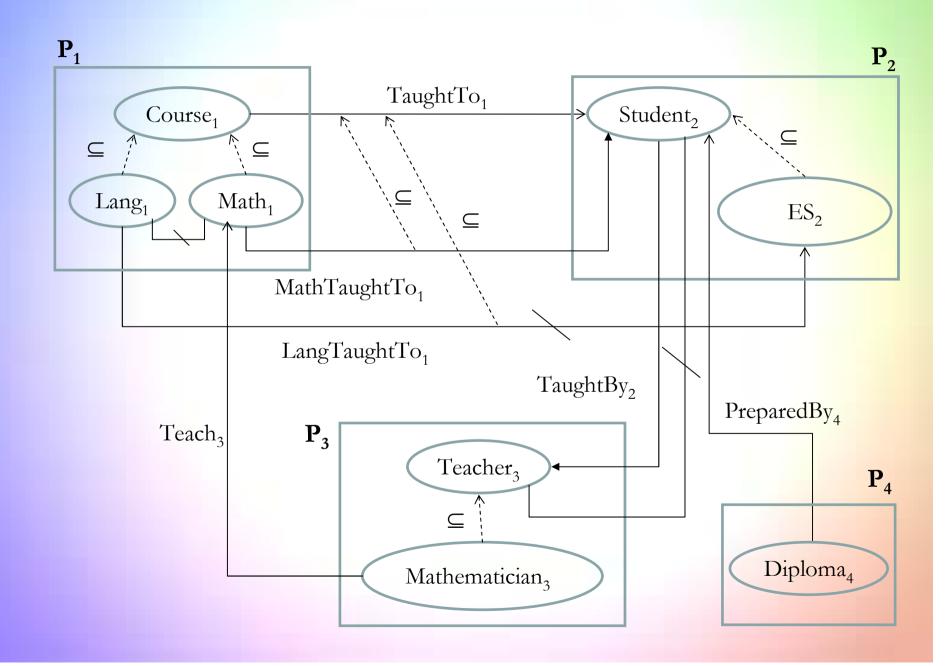












 \mathbf{P}_{1} $\exists \operatorname{TaughtTo}_{1} \subseteq \operatorname{Student}_{2}$ $\exists \text{LangTaughtTo}_1 \subseteq \neg \text{ES}_2$ $\exists \text{ TaughtTo}_1 \subseteq \text{Course}_1$ Lang ₁⊆ Course₁ \exists MathTaughtTo₁ \subseteq Student₂ $Math_1 \subseteq Course_1$ $\exists LangTaughtTo_1 \subseteq Lang_1$ \exists MathTaughtTo₁ \subseteq Math₁ Lang $_1 \subseteq \neg$ Math₁ \mathbf{P}_2 LangTaughtTo₁ \subseteq TaughtTo₁ $ES_2 \subseteq Student_2$ $MathTaughtTo_1 \subseteq TaughtTo_1$ $\exists \text{ TaughtBy}_2 \subseteq \text{Student}_2$ TaughtTo₁(Algebra, John) ES₂ (Mike) TaughtBy, (Mike, Nada) $\exists \text{ TaughtBy}_2^- \subseteq \text{Teacher}_3$ $Student_2 \subseteq \neg Teacher_3$ \mathbf{P}_3 $\exists PreparedBy_4^- \subseteq Student_2$ Mathematician $_3\subseteq$ Teacher $_3$ P_4 $\exists \text{ Teach}_3 \subseteq \text{Mathematician}_3$ $\exists \operatorname{Teach}_{3} \subseteq \operatorname{Course}_{1}$ $\exists \text{ PreparedBy}_4 \subseteq \text{Diploma}_4$ Teach₃ (John, Analysis)

PreparedBy₄(CS, Peter)

Data consistency Checking

Consistency problem:

A data management system is consistent iff its KB (Tbox + Abox) is satisfiable.

- In DL-Lite_R:
 - ✓ Tbox is always satisfiable
 - ✓ Tbox + Abox may be insatisfiable
- Example:

TBox
$$\exists TaughtTo_1 \subseteq Student_2$$
 $Student_2 \subseteq \neg Teacher_3$ $\exists TaughtTo_1 \subseteq Student_3 \subseteq Teacher_3$ $\exists Teach_3 \subseteq Mathematician_3$ $\exists Teach_3 \subseteq Mathematician_3$ $\exists TaughtTo_1 (Algebra, John)$ $\exists TaughtTo_1 (Algebra, John$

Data consistency Checking

Centralized case

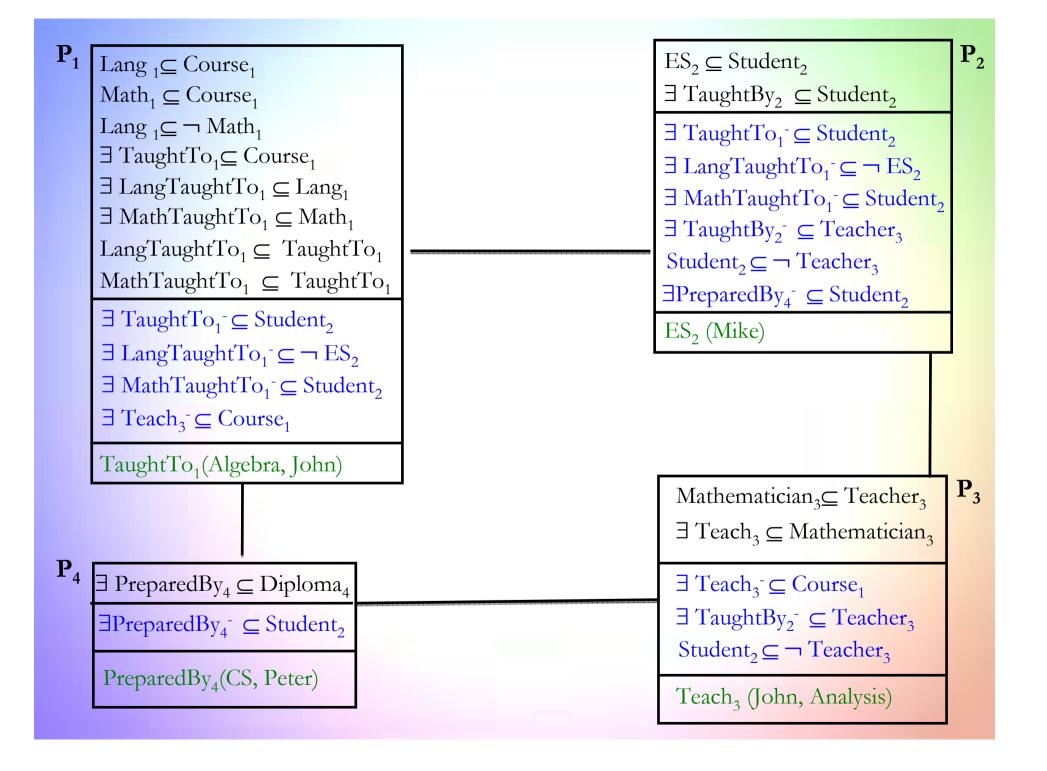
Existing Algorithm Consistent:

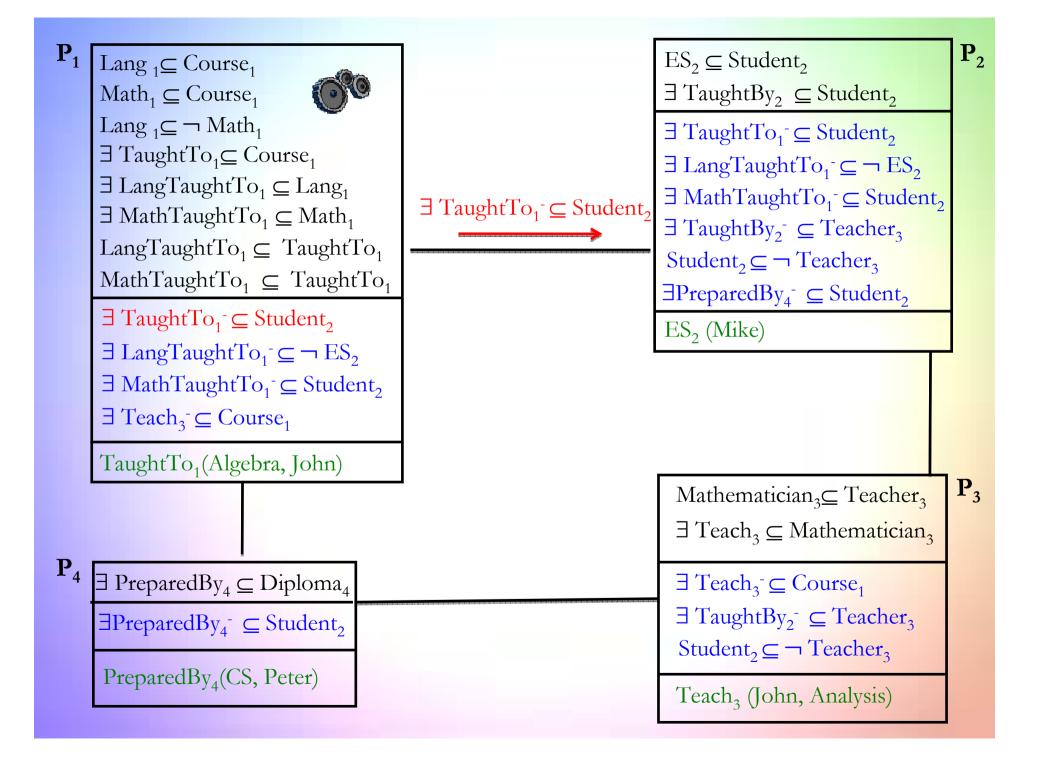
- D.Calvenese and al. (JAR 07): "Tractable reasoning and efficient query answering in description logics"
- > Computes all the Negative Inclusions entailed by the Tbox.
- Checks if the computed negative inclusions are violated by the Abox.

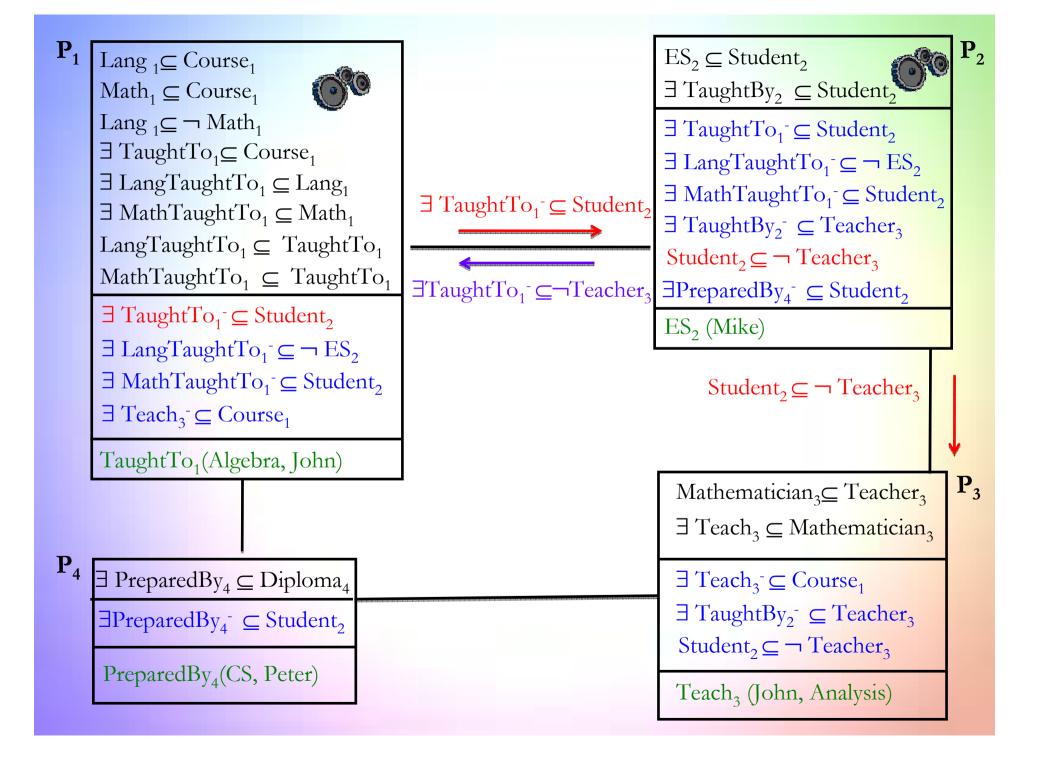
Checking if the DL-Lite_R formulae that must be disjoint according to the TBox, indeed have disjoint instances in the Abox

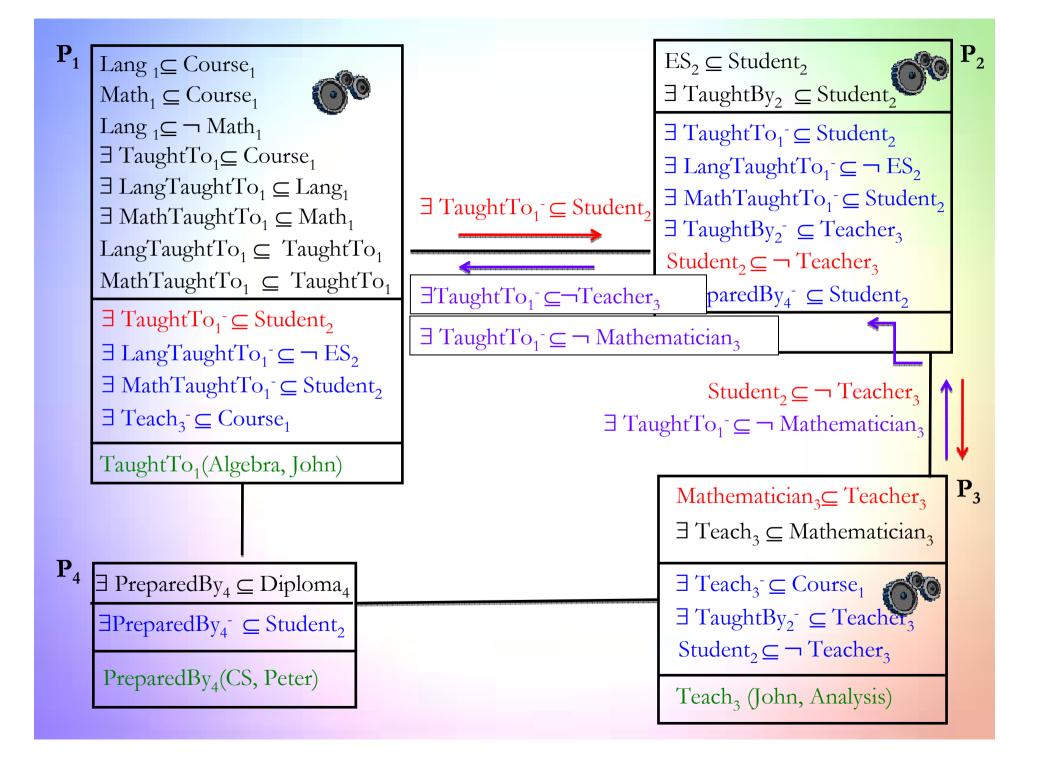
Data consistency Checking

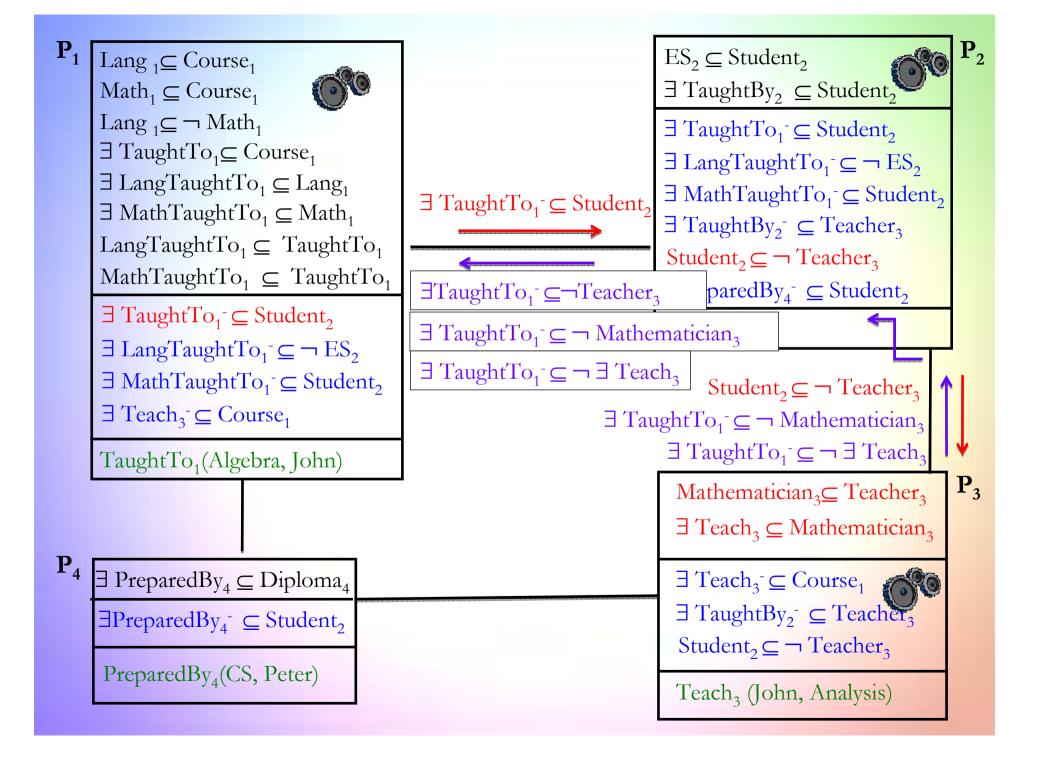
- Decentralized case
 - Decentralized Data consistency Checking w.r.t a peer
 - ✓ Each peer propagates over the network each of its positive or negative inclusions and collects in return all the negative inclusions of the network the entailment of which uses the Tbox of the peer.
 - o All Negative Inclusions entailed by the network are not necessarily computed.
 - ✓ Each peer checks if the negative inclusions that he has collected are violated by the Data by sending queries to the corresponding peers.

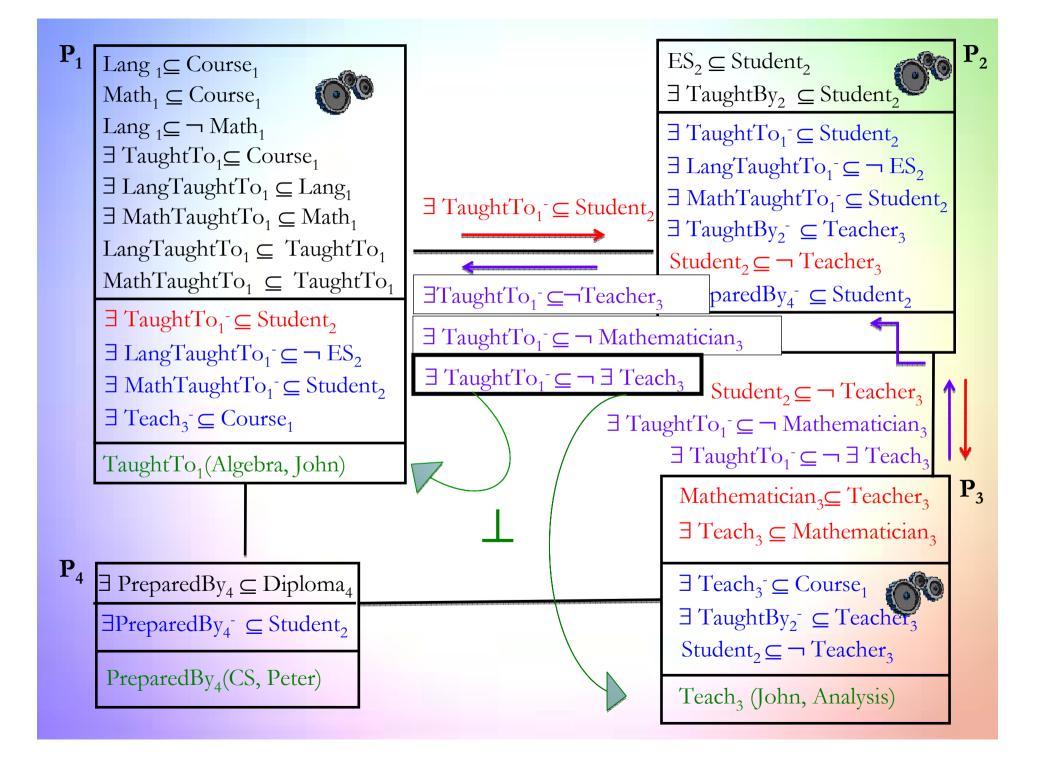


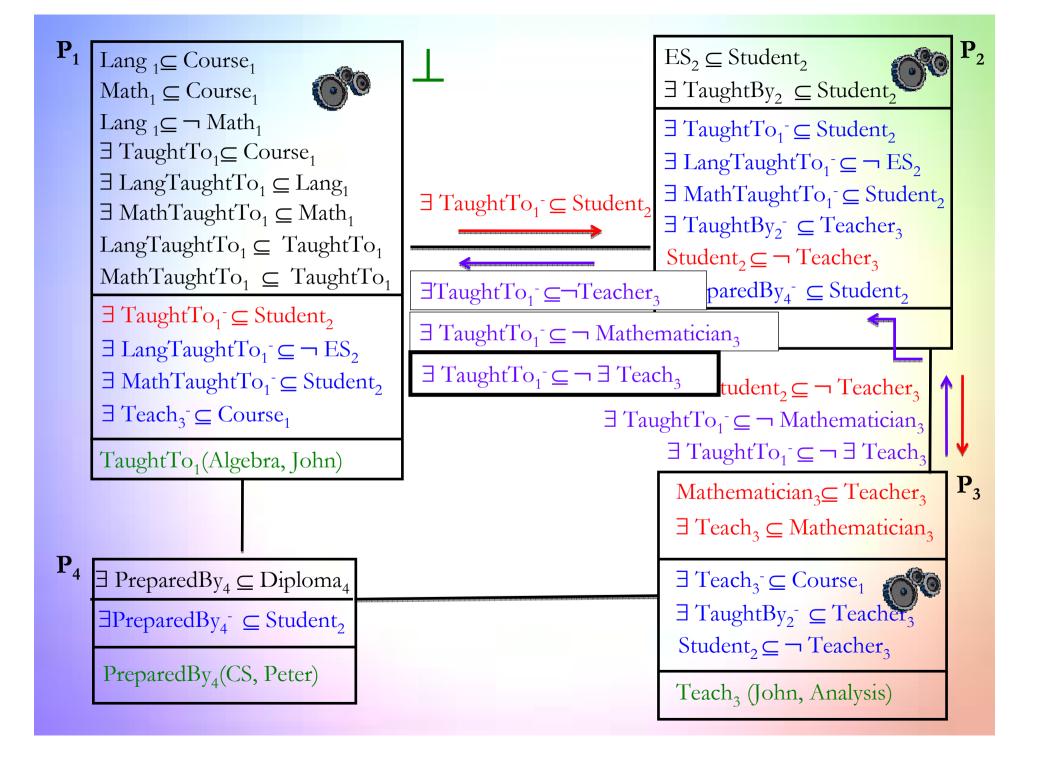


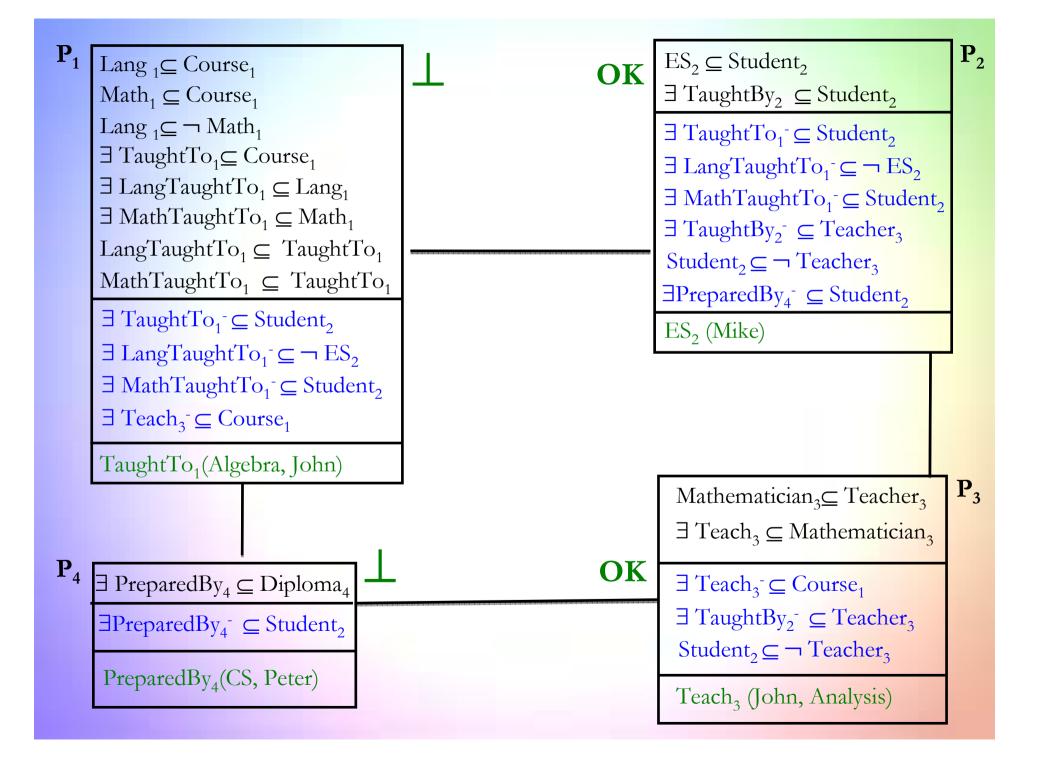












Query Answering

• Query Answering problem:

Given a conjunctive query q, compute the answers of q over the system.

• (Certain) Answer:

A tuple t of the constants appearing in KB such that:

KB = q(t)

• In centralized case:

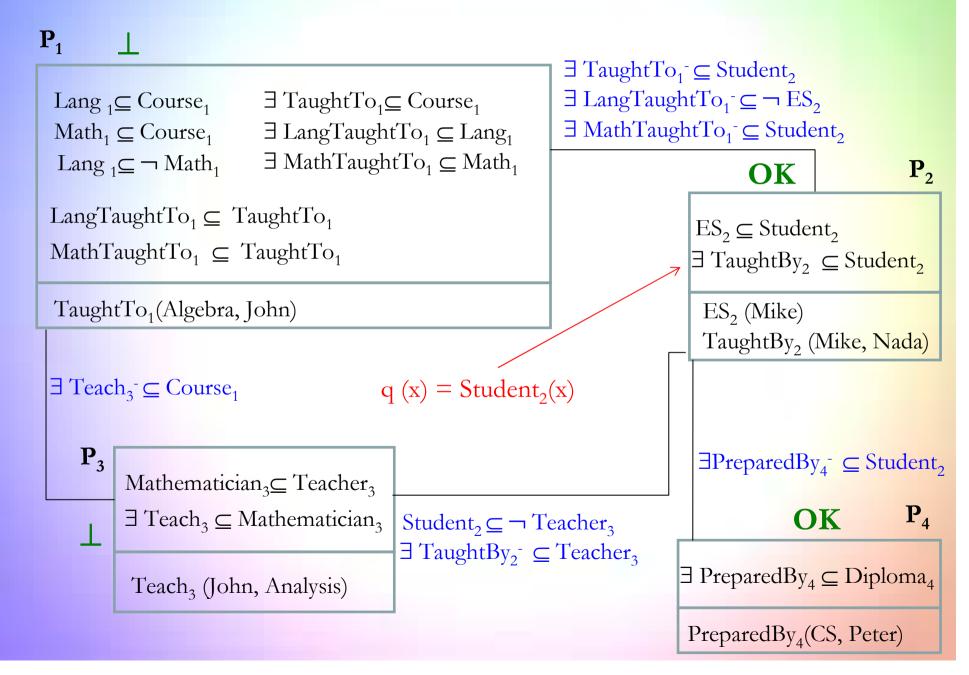
Algorithm PerfectRef: D.Calvenese and al. (JAR 07)

"Tractable reasoning and efficient query answering in description logics"

Data consistency checking

Query Answering

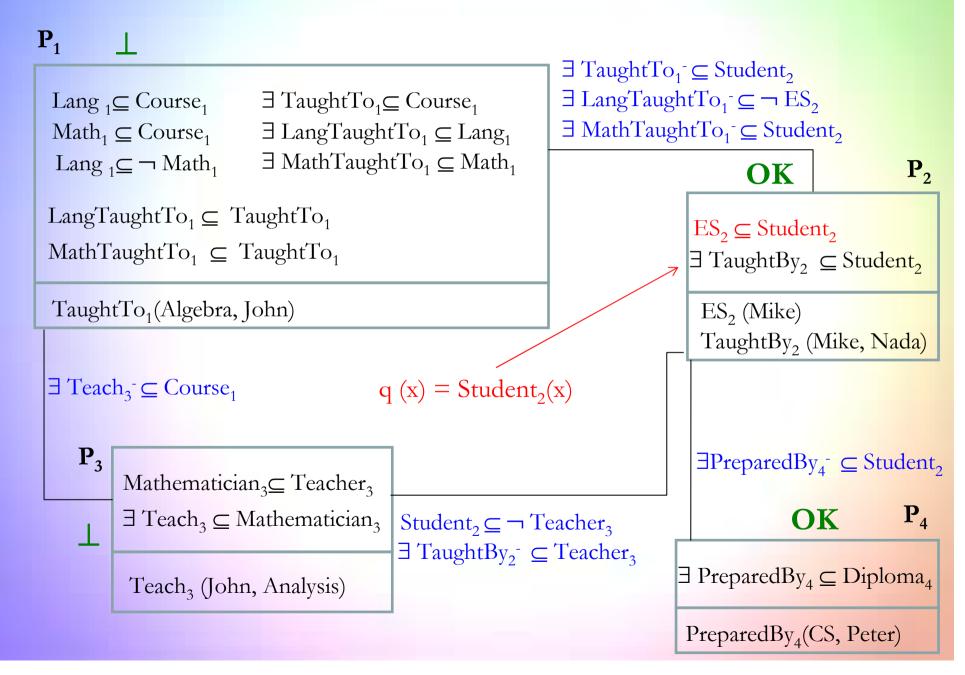
- Query Reformulation
- Reformulation Evaluation



- Contribution:
 - Decentralized algorithm for query answering.
 - Well Founded Answers even in globally inconsistent Networks
 - 1. Decentralized Query Reformulation
 - 2. Checking data consistency w.r.t the peers involved in each reformulation
 - We keep only sound reformulations
 - 3. Evaluation of the sound reformulations
 - Well founded answers

- Contribution:
 - Decentralized algorithm for query answering.
 - ➤ Query reformulation

$$q(x) = Student_2(x)$$



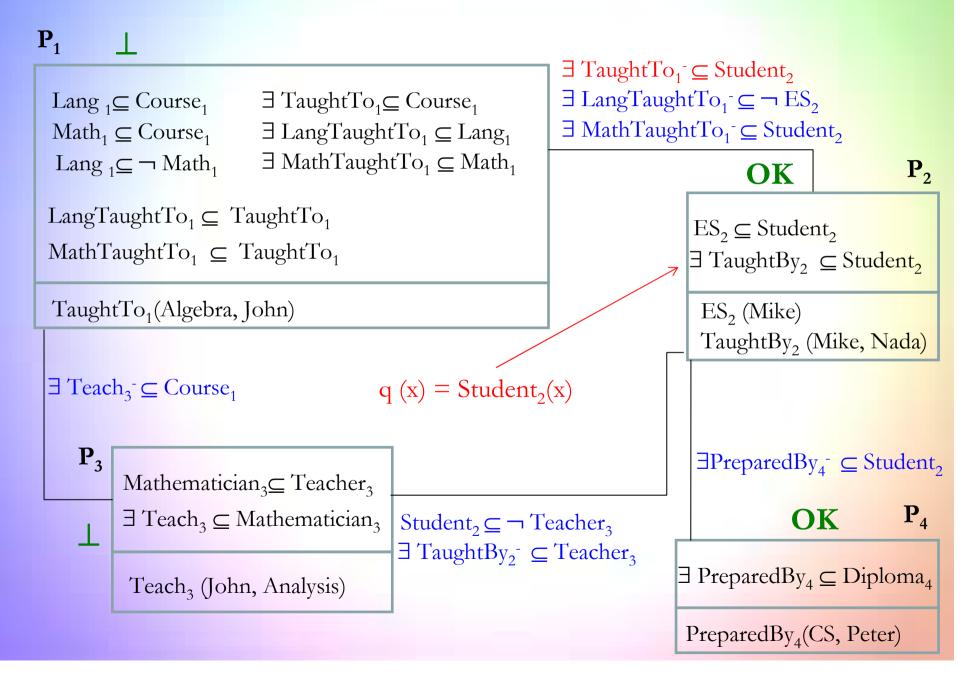
• Contribution:

Decentralized algorithm for query answering.

➤ Query reformulation

$$q(x) = Student_2(x)$$

$$q_1(x) = ES_2(x)$$

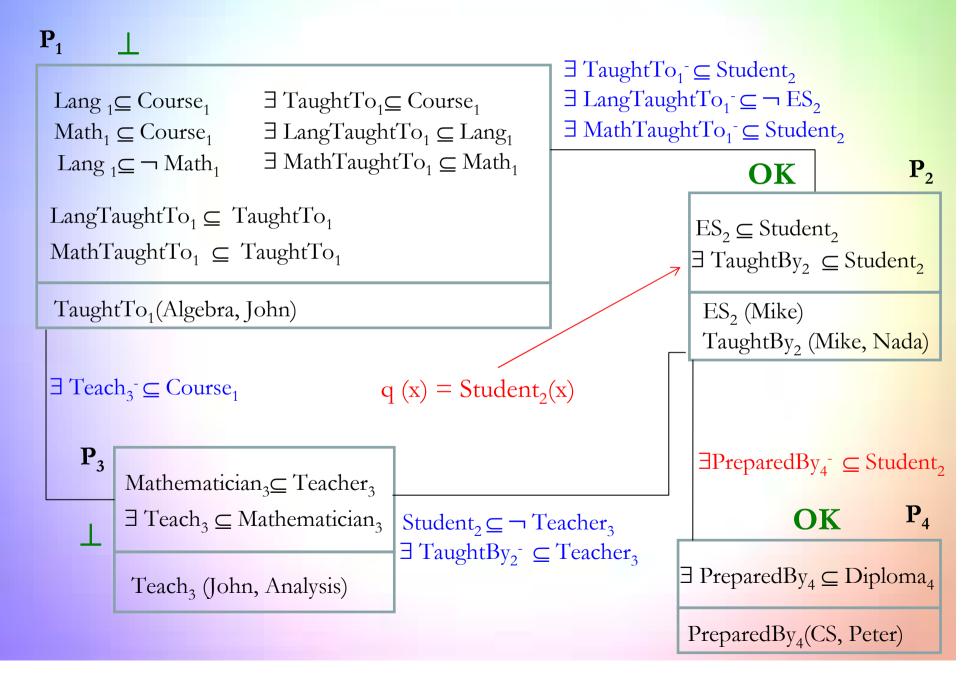


• Contribution:

Decentralized algorithm for query answering.

➤ Query reformulation

```
q(x) = Student_2(x)
q_1(x) = ES_2(x)
q_2(x) = \exists z TaughtTo_1(z, x)
```



• Contribution:

Decentralized algorithm for query answering.

➤ Query reformulation

```
q(x) = Student_2(x)
q_1(x) = ES_2(x)
q_2(x) = \exists z TaughtTo_1(z, x)
q_3(x) = \exists y PreparedBy_4(y, x)
```

• Contribution:

Decentralized algorithm for query answering.

➤ Query reformulation

```
q(x) = Student_2(x)
q_1(x) = ES_2(x)
q_2(x) = \exists z TaughtTo_1(z, x)
q_3(x) = \exists y PreparedBy_4(y, x)
```

Data consistency checking

• Contribution:

Decentralized algorithm for query answering.

> Query reformulation

```
q(x) = Student_2(x)
q_1(x) = ES_2(x)
q_2(x) = \exists z TaughtTo_1(z, x)
q_3(x) = \exists y PreparedBy_4(y, x)
```

Data consistency checking

Consistent algorithm on P₁, P₂ and P₄

On P_1 : \perp

On P₂: OK

On P₄: OK

• Contribution:

Decentralized algorithm for query answering.

> Query reformulation

$$q(x) = Student_2(x)$$

$$q_1(x) = ES_2(x)$$

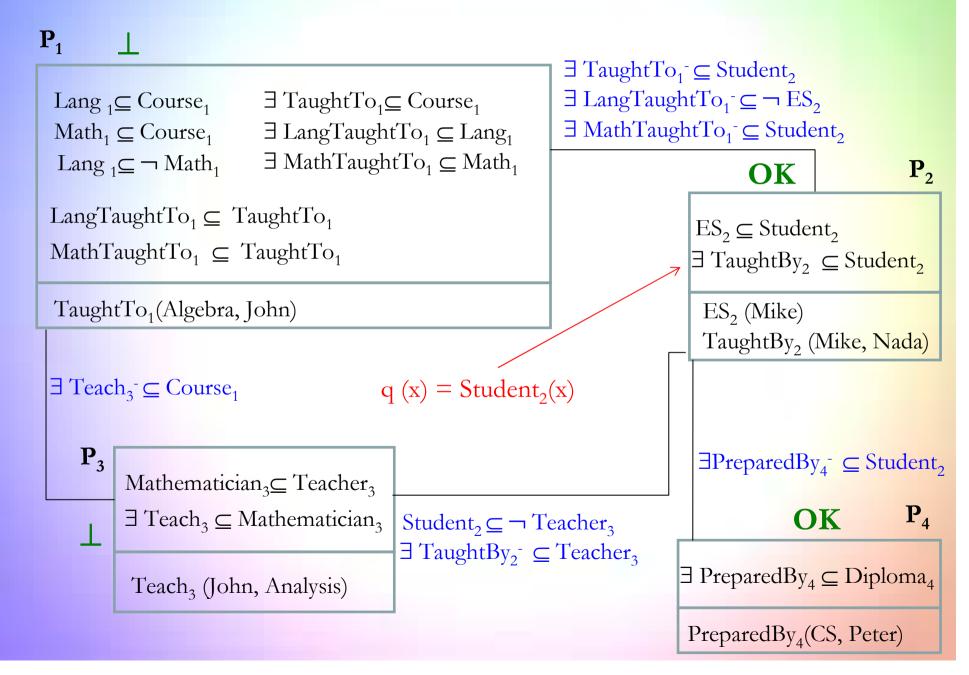
$$q_2(x) = \exists z TaughtTo_1(z, x)$$

$$q_3(x) = \exists y PreparedBy_4(y, x)$$

Data consistency checking

Consistent algorithm on P₁, P₂ and P₄

On P_1 : \perp On P_2 : OK On P_4 : OK Query Evaluation



• Contribution:

Decentralized algorithm for query answering.

> Query reformulation

$$q(x) = Student_2(x)$$

$$q_1(x) = ES_2(x)$$

$$q_2(x) = \exists z \text{ TaughtTo}_1(z, x)$$

$$q_3(x) = \exists y \text{ PreparedBy}_4(y, x)$$

Data consistency checking

Consistent algorithm on P₁, P₂ and P₄

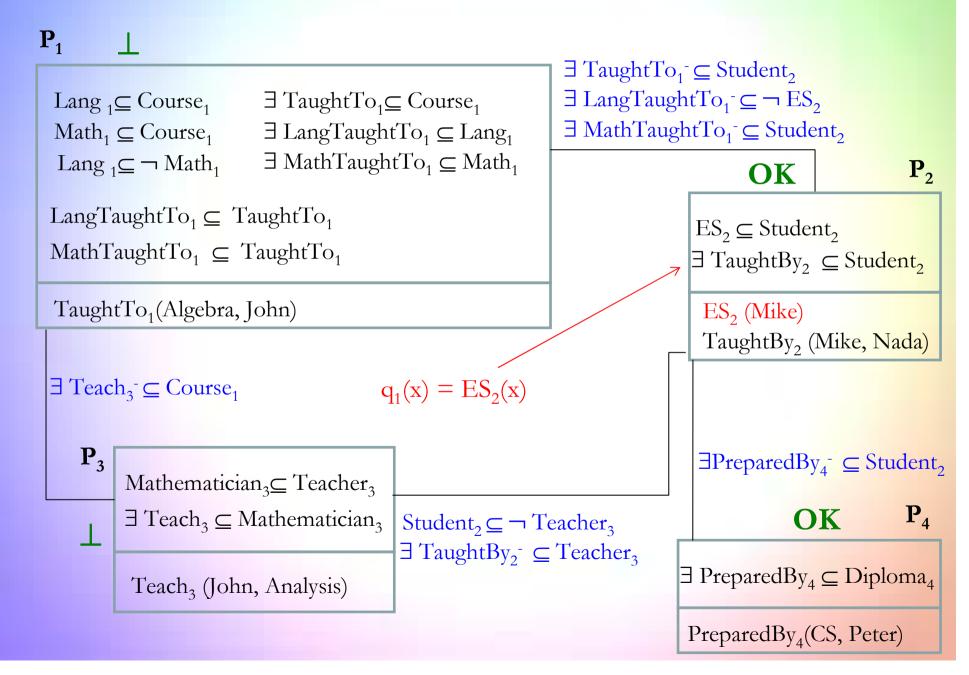
On P_1 : \perp

On P₂: OK

On P₄: OK

Query Evaluation

No answer



• Contribution:

Decentralized algorithm for query answering.

> Query reformulation

$$q(x) = Student_2(x)$$

$$q_1(x) = ES_2(x)$$

$$q_2(x) = \exists z \text{ TaughtTo}_1(z, x)$$

$$q_3(x) = \exists y \text{ PreparedBy}_4(y, x)$$

Data consistency checking

Consistent algorithm on P₁, P₂ and P₄

On P_1 : \perp

On P₂: OK

On P₄: OK

Query Evaluation

No answers

Well founded Answer= {Mike}

• Contribution:

Decentralized algorithm for query answering.

➤ Query reformulation

$$q(x) = Student_2(x)$$

$$q_1(x) = ES_2(x) \qquad \qquad \dots$$

$$q_2(x) = \exists z \text{ TaughtTo}_1(z, x)$$

$$q_3(x) = \exists y \text{ PreparedBy}_4(y, x)$$

Data consistency checking

Consistent algorithm on P₁, P₂ and P₄

On P_1 : \perp

On P₂: OK

On P₄: OK

Query Evaluation

No answers

Well founded Answer= {Mike}

```
\mathbf{P}_{1}
                                                                                                   \exists \operatorname{TaughtTo}_{1} \subseteq \operatorname{Student}_{2}
                                                                                                   \exists \text{ LangTaughtTo}_1 \subseteq \neg \text{ ES}_2
                                          ∃ TaughtTo<sub>1</sub>⊆ Course<sub>1</sub>
  Lang _1\subseteq Course_1
                                                                                                   \exists MathTaughtTo<sub>1</sub>^{-}\subseteq Student<sub>2</sub>
  Math_1 \subseteq Course_1
                                          \exists \text{ LangTaughtTo}_1 \subseteq \text{Lang}_1
                                          \exists MathTaughtTo<sub>1</sub> \subseteq Math<sub>1</sub>
   Lang _1 \subseteq \neg Math<sub>1</sub>
                                                                                                                                                                    \mathbf{P}_2
                                                                                                                                      OK
  LangTaughtTo_1 \subseteq TaughtTo_1
                                                                                                                            ES_2 \subseteq Student_2
  MathTaughtTo_1 \subseteq TaughtTo_1
                                                                                                                           \exists \text{ TaughtBy}_2 \subseteq \text{Student}_2
   TaughtTo<sub>1</sub>(Algebra, John)
                                                                                                                             ES<sub>2</sub> (Mike)
                                                                                                                             TaughtBy, (Mike, Nada)
  \exists \text{ Teach}_3 \subseteq \text{Course}_1 q_3(x) = \exists y \text{ PreparedBy}_4(y, x)
       \mathbf{P}_3
                                                                                                                             \exists PreparedBy_4^- \subseteq Student_2
                Mathematician₃⊆ Teacher₃
                                                                                                                                                                    \mathbf{P}_4
                \exists \operatorname{Teach}_3 \subseteq \operatorname{Mathematician}_3
                                                                                                                                               OK
                                                                    Student<sub>2</sub> \subseteq \neg Teacher<sub>3</sub>
                                                                    \exists \text{ TaughtBy}_{2} \subseteq \text{Teacher}_{3}
                                                                                                                         \exists \text{ PreparedBy}_4 \subseteq \text{Diploma}_4
                 Teach<sub>3</sub> (John, Analysis)
                                                                                                                          PreparedBy<sub>4</sub>(CS, Peter)
```

• Contribution:

Decentralized algorithm for query answering.

> Query reformulation

$$q(x) = Student_2(x)$$

$$q_1(x) = ES_2(x) \qquad ----$$

$$q_2(x) = \exists z \text{ TaughtTo}_1(z, x)$$

$$q_3(x) = \exists y \text{ PreparedBy}_4(y, x) \longrightarrow$$

Data consistency checking

Consistent algorithm on P₁, P₂ and P₄

On P_1 : \perp

On P₂: OK

On P₄: OK

Query Evaluation

No answers

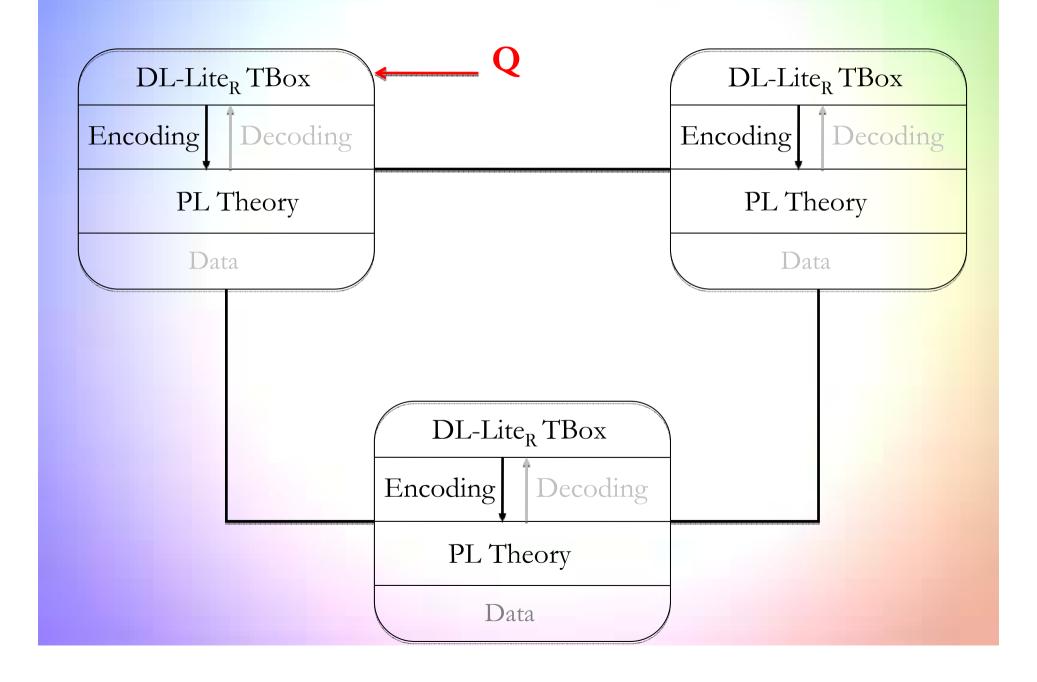
Well founded Answer= {Mike}

Well founded Answer= {Peter}

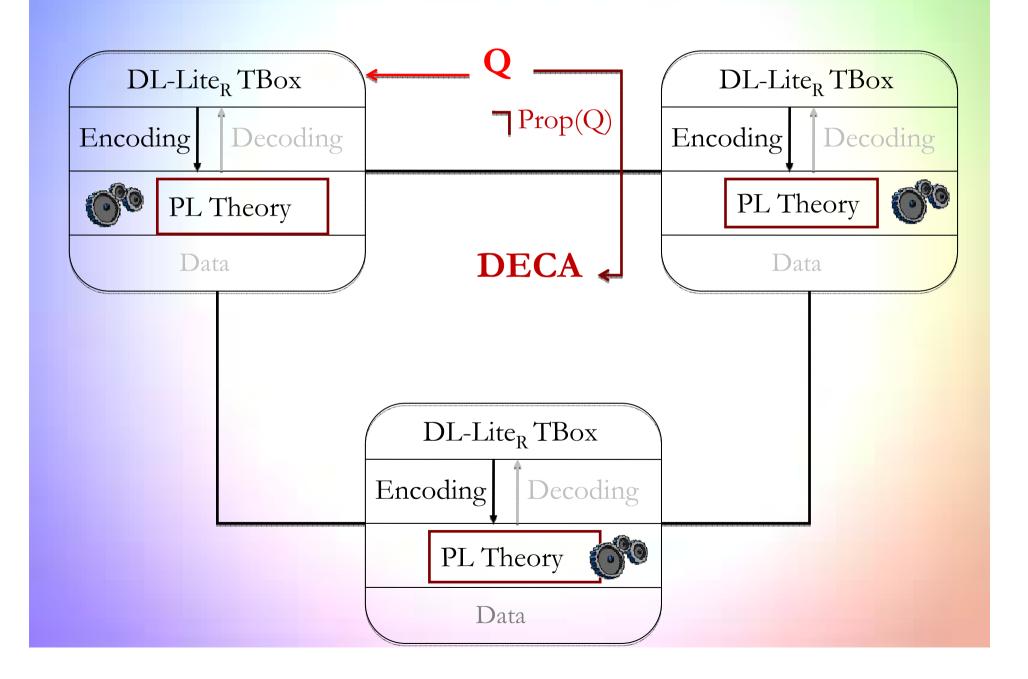
Decentralized query reformulation by using SomeWhere

- ✓ An existing Decentralized Infrastructure "SomeWhere" for reasoning in propositional logic and for which the experiments have demonstrated the scalability.
 - P. Adjiman and al. (IJCAI 05): Scalability study of peer-to-peer consequence finding'
- ✓ An existing Decentralized Consequence finding Algorithm in propositional logic (DECA)
- ✓ A propositional encoding of the DL-Lite_R TBox.

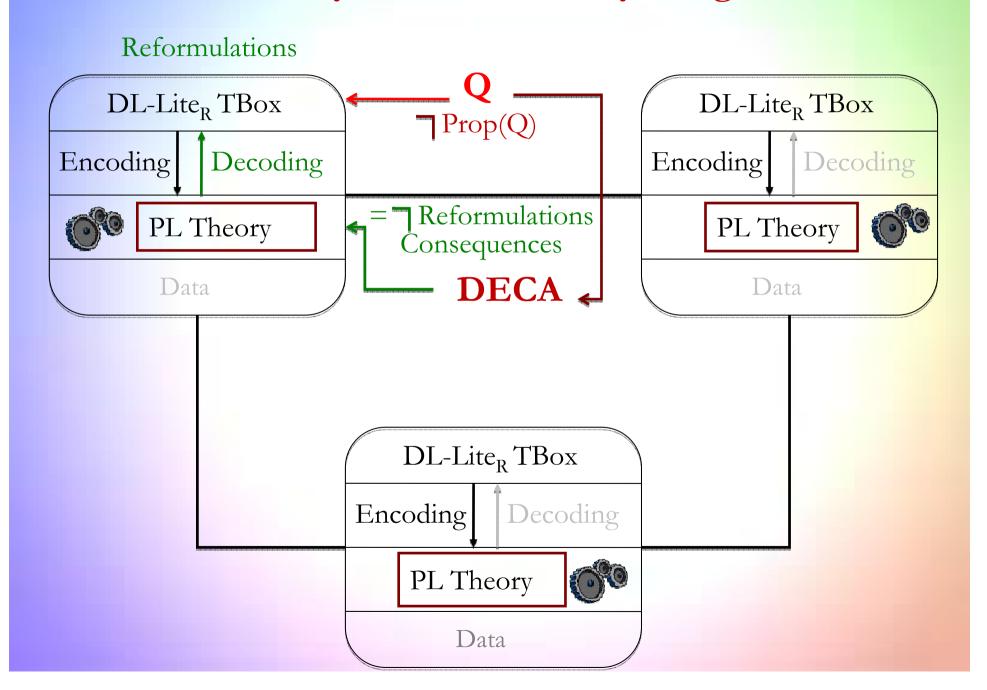
Decentralized Query Reformulation by using SomeWhere



Decentralized Query Reformulation by using SomeWhere

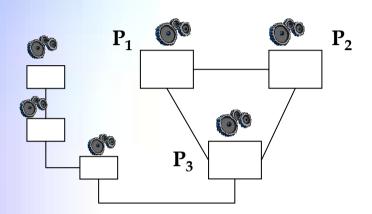


Decentralized Query Reformulation by using SomeWhere



Conclusion & future work

1. DL-Lite_R decentralized data model



- 2. Decentralized Algorithms for:
 - ✓ Data consistency checking
 - ✓ Query Answering
- 3. Decentralized & Centralized Algorithms for:
 - ✓ View consistency checking
 - **✓ Query Answering using views**

Other Applications

- Mapping discovery:
 Rémi Tournaire's PhD 2010
- Data reconciliation:
 Extension of Fatiha Sais' PhD
- Explanation of answers to queries

Feasibility of the approach with other dialects of DL-Lite and with Datalog+-