## Query Processing in Data Integration Systems

Diego Calvanese

Free University of Bozen-Bolzano

BIT PhD Summer School – Bressanone July 3–7, 2006



### Structure of the course

- Introduction to data integration
  - Basic issues in data integration
  - Logical formalization
- Query answering in the absence of constraints
  - Global-as-view (GAV) setting
  - Local-as-view (LAV) and GLAV setting
- Query answering in the presence of constraints
  - The role of integrity constraints
  - Global-as-view (GAV) setting
  - Local-as-view (LAV) and GLAV setting
- Concluding remarks



Part 4: Conclusions

### Part I

## Concluding remarks



#### Outline

- 1 The role of global integrity constraints
- Query answering in GAV with constraints
  - Incompleteness and inconsistency in GAV systems
  - Query answering in GAV under inclusion dependencies
  - Rewriting CQs under inclusion dependencies in GAV
  - Query answering in GAV under IDs and KDs
  - Query answering in GAV under IDs, KDs, and EDs
- 3 Query answering in LAV with constraints
  - LAV systems and integrity constraints
  - Query answering in (G)LAV under inclusion dependencies
  - Query answering in (G)LAV under IDs and EDs
  - LAV systems and key dependencies



#### Outline

- 1 The role of global integrity constraints
- 2 Query answering in GAV with constraints
  - Incompleteness and inconsistency in GAV systems
  - Query answering in GAV under inclusion dependencies
  - Rewriting CQs under inclusion dependencies in GAV
  - Query answering in GAV under IDs and KDs
  - Query answering in GAV under IDs, KDs, and EDs
- Query answering in LAV with constraints
  - LAV systems and integrity constraints
  - Query answering in (G)LAV under inclusion dependencies
  - Query answering in (G)LAV under IDs and EDs
  - LAV systems and key dependencies



## Further issues and open problems

- Further forms of constraints, e.g.,
  - KDs with restricted forms of key-conflicting IDs
  - ontology languages, description logics, RDF
    [— & al. PODS'98, & al., KR'06]
- Semistructured data and XMI
  - constraints (DTDs, XML Schema, ...)
  - query languages (transitive closure)
- Finite models vs. unrestricted models [Rosati, PODS'06]
- Data exchange and materialization



### Acknowledgements

- Andrea Calì
- Giuseppe De Giacomo
- Domenico Lembo
- Maurizio Lenzerini
- Riccardo Rosati
- Moshe Y. Vardi



### References I



S. Abiteboul and O. Duschka.

Complexity of answering queries using materialized views.

In Proc. of PODS'98, pages 254-265, 1998.



A. Calì, D. Calvanese, G. De Giacomo, and M. Lenzerini.

On the expressive power of data integration systems.

In Proc. of ER 2002, 2002.



A. Calì, D. Calvanese, G. De Giacomo, and M. Lenzerini.

On the role of integrity constraints in data integration.

*IEEE Bull. on Data Engineering*, 25(3):39–45, 2002.



A. Calì, D. Lembo, and R. Rosati.

Query rewriting and answering under constraints in data integration systems.

In Proc. of IJCAI 2003, pages 16-21, 2003.



### References II



D. Calvanese, G. De Giacomo, D. Lembo, M. Lenzerini, and R. Rosati.

Data complexity of query answering in description logics.

In Proc. of KR 2006, 2006.



D. Calvanese, G. De Giacomo, and M. Lenzerini.

On the decidability of query containment under constraints.

In Proc. of PODS'98, pages 149-158, 1998.



D. Calvanese, G. De Giacomo, M. Lenzerini, and M. Y. Vardi.

View-based query processing: On the relationship between rewriting, answering and losslessness.

In Proc. of ICDT 2005, volume 3363 of LNCS, pages 321-336. Springer, 2005.



D. Calvanese and R. Rosati.

Answering recursive queries under keys and foreign keys is undecidable.

In *Proc. of KRDB 2003*. CEUR Electronic Workshop Proceedings, http://ceur-ws.org/Vol-79/, 2003.



# References III



 $O.\ M.\ Duschka,\ M.\ R.\ Genesereth,\ and\ A.\ Y.\ Levy.$ 

Recursive query plans for data integration.

J. of Logic Programming, 43(1):49-73, 2000.



A. Fuxman and R. J. Miller.

First-order query rewriting for inconsistent databases.

In Proc. of ICDT 2005, volume 3363 of LNCS, pages 337-351. Springer, 2005.



G. Grahne and A. O. Mendelzon.

Tableau techniques for querying information sources through global schemas.

In Proc. of ICDT'99, volume 1540 of LNCS, pages 332-347. Springer, 1999.



A. Y. Halevy.

Answering queries using views: A survey.

VLDB Journal, 10(4):270-294, 2001.



References IV

Query answering in LAV with constraints

#### Part 4: Conclusions



M. Lenzerini.

Data integration: A theoretical perspective.

In Proc. of PODS 2002, pages 233-246, 2002.



N. Leone, T. Eiter, W. Faber, M. Fink, G. Gottlob, G. Greco, E. Kalka, G. lanni,

D. Lembo, V. Lio, B. Nowicki, R. Rosati, M. Ruzzi, W. Staniszkis, and G. Terracina.

Boosting information integration: The INFOMIX system.

In Proc. of SEBD 2005, pages 55-66, 2005.



A. Y. Levy, A. Rajaraman, and J. J. Ordille.

Query answering algorithms for information agents.

In Proc. of AAAI'96, pages 40-47, 1996.



### References V



#### R. Rosati.

On the decidability and finite controllability of query processing in databases with incomplete information.

In Proc. of PODS 2006, 2006.

