

# Internship proposal

## Querying inconsistent data

March 12, 2019

### Supervisors

- Cristina Sirangelo (IRIF, Université Paris Diderot) and
- Diego Figueira (CNRS, LABRI, Bordeaux)

### Contact

Cristina Sirangelo <http://www.irif.fr/~cristina>  
E-mail: [cristina@irif.fr](mailto:cristina@irif.fr)

Diego Figueira: <http://www.labri.fr/perso/dfigueir/>  
E-mail: [dfigueira@gmail.com](mailto:dfigueira@gmail.com)

**Keywords** Database theory, logic, complexity

**Internship topic** The internship topic is in the area of foundations of data management, focusing on querying inconsistent data.

Data management systems nowadays need to handle large data sets often containing inconsistent or incomplete information, and yet guarantee the relevance of query answers. New techniques are needed for making data access more reliable, and new tools for making data querying more efficient in this context.

The main focus of the internship will be the study of consistent query answering. In this scenario the data set may have inconsistencies relative to integrity constraints (e.g., key and foreign key constraints, functional dependencies, inclusion constraints, etc.). Queries issued on the database need to be answered despite the inconsistencies. The goal is to provide only consistent answers, i.e. answers that can be found after virtually repairing the database to remove inconsistencies. As many repairs may exist for a given database, the widely accepted notion is the one of certain answers, i.e. answers that can be found in all possible database repairs.

The challenge is then to provide efficient solutions for computing certain answers directly from the inconsistent database, as going through all possible repairs is clearly unfeasible. Particularly attractive is the possibility of obtaining

certain answers by issuing a first-order query on the inconsistent database, or a query expressible in other logics guaranteeing good computational properties.

First-order rewritability and PTIME computability of consistent query answers have been studied for various classes of query languages and integrity constraints (see [1] for an in-depth survey). In particular for conjunctive queries and large classes of constraints, a dichotomy conjecture states that computing certain answers can either be done in polynomial time or is coNP-complete. This conjecture has only been solved for restricted classes of queries and constraints, including key constraints and self-join free queries [2, 3].

The internship will start by investigating the rewritability question in the case of key constraints and arbitrary conjunctive queries, which is unexplored as of today. Depending on the results, the internship may go on investigating the more general dichotomy conjecture in the same class of constraints, as well as possibly larger classes (these include functional dependencies, equality-generating dependencies and GAV (Global-as-view) constraints).

**Information** The student will be hosted by IRIF, Paris Diderot University, and will be able to spend some time at LABRI, Bordeaux. We encourage applications from students potentially interested in pursuing the internship work in a PhD under our supervision.

**Requirements** The student should have a background in theoretical computer science. Some prior knowledge of database theory and systems is a plus, but not required.

## References

- [1] Leopoldo E. Bertossi. *Database Repairing and Consistent Query Answering*. Synthesis Lectures on Data Management. Morgan & Claypool Publishers, 2011.
- [2] Paraschos Koutris and Jef Wijsen. Consistent query answering for self-join-free conjunctive queries under primary key constraints. *ACM Transactions on Database Systems*, 42(2):9:1–9:45, 2017.
- [3] Paraschos Koutris and Jef Wijsen. Consistent query answering for primary keys and conjunctive queries with negated atoms. In *ACM Symp. on Principles of Database Systems (PODS)*, pages 209–224, 2018.