

Postdoc or Research Engineer Position: Practical Algorithms for Imperfect Data

Context

Accessing the relevant information in real-world data to support informed decision making is difficult, time-consuming, and error-prone due to the need to integrate data across multiple heterogeneous sources. Moreover, even if this first hurdle is overcome, a perhaps even more daunting challenge arises: how to obtain reliable insights from imperfect data? It is widely acknowledged that real-world data is plagued with quality issues, such as incompleteness (missing information) and errors (false or outdated information).

The ontology-mediated query answering (OMQA) paradigm [1,2,3] addresses the first challenge by facilitating access to (potentially heterogeneous) data sources through the use of ontologies that specify a convenient user-friendly vocabulary for query formulation (which abstracts from the way the data is stored) and capture domain knowledge that can be exploited at query time, via automated reasoning, to obtain more complete query results. While OMQA systems are growing in maturity [4], they too often fail to address the data quality issue, aside from issuing warnings when inconsistencies are discovered.

To widen the applicability of the OMQA approach, it is essential to equip OMQA systems with appropriate mechanisms for handling imperfect data: how to obtain meaningful answers to queries posed over imperfect data, and how best to generate a high-quality version of the data? While these questions have begun to be explored [5,6], we are still far from having robust and widely applicable techniques for handling data quality in OMQA.

The INTENDED Chair's overall aim is to develop intelligent, knowledge-based methods for handling imperfect data. Examples of ongoing research directions within the project include: the integration of qualitative and quantitative preferences into repair-based semantics [7,8], the development of logic-based approaches to entity resolution [9,10], and the definition of holistic frameworks for jointly tackling various data quality issues [11]. We are also interested in implementing the developed algorithms, in particular, using SAT solvers [12] and answer set programming (ASP) systems [13].

This position is part of the INTENDED Chair on Artificial Intelligence, funded by the French National Research Agency (ANR) and the University of Bordeaux.

Position

The primary focus of this position will be to take the project's foundational contributions from theory into practice by implementing, optimising, and testing various algorithms and prototypes that we are developing within the project. These include in particular SAT- and ASP-based implementations of various repair-based semantics for querying inconsistent data [12], which make it possible to extract various kinds of answers from inconsistent data, taking into account user preferences. The ultimate goal will be to provide an easy-to-use tool to showcase and enable easy testing and comparison of the developed algorithms. Example of concrete tasks include: utilizing (off-the-shelf or adapting) various existing reasoning systems to compute minimal inconsistent subsets and query supports, constructing encodings and interfacing with SAT and ASP solvers, adapting and developing new benchmark knowledge bases, and conducting rigorous experimental

evaluations of the implemented algorithms. Furthermore, depending on their qualifications and prior experience, the recruited person can take a leading role in developing novel algorithms and optimizations.

This position can be either for a **full-time postdoctoral researcher** or a **full-time research engineer**. In either case, the position is for a duration of **13 months, starting from September 1st, 2025**. The approximate monthly net salary is around 2400€ for a postdoctoral researcher, and 2280€ in the case of a research engineer (but there is possibility of a higher salary for applicants with prior postdoc / engineer experience). Note that the position does not include any mandatory teaching or supervisory activities, but there will be opportunities to engage in teaching and/or student supervision if desired.

Work Environment

The position will be based at LaBRI (Laboratoire Bordelais de Recherche en Informatique), a computer science research lab located on the University of Bordeaux's Talence campus. The lab can be easily reached from the city center of Bordeaux by tram or bike. Located in southwestern France, Bordeaux offers easy access to both the Atlantic coast and the Pyrenees mountains, while being only a two-hour train ride from Paris. Famed for its vineyards, local gastronomy, and well-preserved historic center (a UNESCO World Heritage Site), Bordeaux offers a high quality of life.

The recruited person will participate in the research team RATIO (Reasoning with data, knowledge and constraints) and will collaborate with Meghyn Bienvenu (leader of the chair) and other members of the INTENDED project. The chair provides significant resources to finance trips to conferences and to support research visits.

Candidate Profile

We are looking for someone who is excited about symbolic AI and declarative database methods and who is eager to see these techniques successfully applied in practice.

Applicants for a postdoctoral position should hold a PhD degree (or be in possession of a PhD degree by the end of June 2025 at the latest) and should have a strong research record in knowledge representation (in particular, ontologies), automated reasoning, databases (in particular, data quality), Semantic Web, or related fields, as demonstrated by publications in top venues. Prior experience in developing, implementing, and testing reasoning or database algorithms (SAT, ASP, OWL, RDF, Datalog...) would be a significant asset. Alternatively, the applicant should demonstrate significant programming experience.

Applicants for a research engineer position must hold a Master's or engineering degree (or equivalent). They must demonstrate solid programming skills and should have at least basic familiarity with logic and symbolic AI techniques (e.g. via relevant coursework and/or programming projects).

The working language (for either type of position) can be either French or English. For the postdoc position, strong English language skills and an ability to take active part in writing up research papers are expected.

How to Apply

Interested candidates should contact Meghyn Bienvenu and provide:

- a detailed CV, containing the applicant's full publication list (if applicable), relevant coursework, and detailing their programming experience,
- university transcripts for Master's / engineering degree,
- names of 2-3 references,
- a brief description of how the project topics relate to their prior experience and interests.

Applications should be submitted by email to both of the following email addresses: meghyn.bienvenu@u-bordeaux.fr and meghyn.bienvenu@cnrs.fr

The deadline for applications is: **April 20, 2025**. Applicants will receive an email to confirm receipt of their application within a few working days. Online interviews with selected candidates will be scheduled between mid-April and mid-May.

Bibliography

- [1] A. Poggi, D. Lembo, D. Calvanese, G. De Giacomo, M. Lenzerini, and R. Rosati. **Linking Data to Ontologies**. Journal of Data Semantics, 10:133–173, 2008.
- [2] G. Xiao, D. Calvanese, R. Kontchakov, D. Lembo, A. Poggi, R. Rosati, and M. Zakharyashev. **Ontology-Based Data Access: A Survey**. In Proc. of the 27th International Joint Conference on Artificial Intelligence (IJCAI), pages 5511–5519, 2018.
- [3] M. Bienvenu, M. Ortiz: **Ontology-Mediated Query Answering with Data-Tractable Description Logics**. Tutorial Notes of the 11th International Reasoning Web Summer School (LNCS 9203), pages 218-307, 2015.
- [4] D. Calvanese, B. Cogrel, S. Komla-Ebri, R. Kontchakov, D. Lanti, M. Rezk, M. Rodriguez-Muro, G. Xiao. **Ontop: Answering SPARQL Queries over Relational Databases**. Semantic Web Journal, vol. 8, no. 3, pp. 471-487, 2017.
- [5] M. Bienvenu, C. Bourgaux: **Inconsistency-Tolerant Querying of Description Logic Knowledge Bases**. Tutorial Notes of the 12th International Reasoning Web Summer School (LNCS 9885), pages 156- 202, 2016.
- [6] M. Bienvenu: **A Short Survey on Inconsistency Handling in Ontology-Mediated Query Answering**. Künstliche Intelligenz 34(4): 443-451, 2020.
- [7] M. Bienvenu, C. Bourgaux: **Querying and Repairing Inconsistent Prioritized Knowledge Bases: Complexity Analysis and Links with Abstract Argumentation**. In Proc. of the 17th International Conference on Principles of Knowledge Representation and Reasoning (KR), pages 141-151, 2020.
- [8] M. Bienvenu, C. Bourgaux, R. Jean. **Cost-Based Semantics for Querying Inconsistent Weighted Knowledge Bases**. Proc. of the 21st International Conference on Principles of Knowledge Representation and Reasoning (KR), 2024.
- [9] M. Bienvenu, G. Cima, V. Gutierrez-Basulto. **LACE: A Logical Approach to Collective Entity Resolution**. Proc. of the 41st ACM SIGMOD-SIGACT-SIGAI Symposium on Principles of Database Systems (PODS), 2022.
- [10] M. Bienvenu, G. Cima, V. Gutierrez-Basulto, Y. Ibanez-Garcia. **Combining Global and Local Merges in Logic-based Entity Resolution**. Proc. of the 20th International Conference on Principles of Knowledge Representation and Reasoning (KR), 2023.

- [11] M. Bienvenu, G. Cima, V. Gutierrez-Basulto. **REPLACE: A Logical Framework for Combining Collective Entity Resolution and Repairing**. Proc. of the 32nd International Joint Conference on Artificial Intelligence (IJCAI), 2023.
- [12] M. Bienvenu, C. Bourgaux: **Querying Inconsistent Prioritized Data with ORBITS: Algorithms, Implementation, and Experiments**. Proc. of the 19th International Conference on Principles of Knowledge Representation and Reasoning (KR), 2022.
- [13] Z. Xiang, M. Bienvenu, G. Cima, V. Gutierrez-Basulto, Y. Ibanez-Garcia: **ASPEn: ASP-Based System for Collective Entity Resolution**. Proc. of the 21st International Conference on Principles of Knowledge Representation and Reasoning (KR), 2024.