

RESEARCH CENTRE

**Inria Branch at the University of
Montpellier**

IN PARTNERSHIP WITH:

Université de Montpellier, INRAE

2023

ACTIVITY REPORT

Project-Team

BOREAL

Knowledge Representation and Rule-Based Languages for Reasoning on Data

IN COLLABORATION WITH: Laboratoire d'informatique, de robotique et
de microélectronique de Montpellier (LIRMM), Ingénierie des
Agropolymères et Technologies Emergentes (IATE)

DOMAIN

Perception, Cognition and Interaction

THEME

**Data and Knowledge Representation and
Processing**

The Inria logo is a stylized, cursive script in red, positioned in the bottom right corner of the page.

Contents

Project-Team BOREAL	1
1 Team members, visitors, external collaborators	2
2 Overall objectives	3
3 Research program	3
4 Application domains	4
4.1 Agronomy and agroecology	4
5 Highlights of the year	5
5.1 Awards	5
6 New software, platforms, open data	5
6.1 New software	5
6.1.1 InteGraal	5
6.1.2 TreeForce	6
7 New results	6
7.1 Knowledge-Based Data Management with existential rules	6
7.2 Foundations of rule-based reasoning and query answering	8
7.2.1 Terminating conditions for disjunctive rules	8
7.2.2 Rewriting with disjunctive existential rules and mappings	8
7.2.3 Decidability of bounded treewidth infinite core chase	9
7.2.4 Effects of rule normalization	9
7.3 Algorithms for querying specific data sources	9
7.3.1 Fine grained complexity of query answering	9
7.3.2 Querying JSON document stores	10
7.4 Applications of logics and rule-based languages	10
7.4.1 Systems for reasoning on heterogeneous and federated data	10
7.4.2 Knowledge Representation formalisms for testing power plants	10
7.4.3 Food packaging in agronomy	11
7.4.4 Human-robot interactions	11
8 Bilateral contracts and grants with industry	11
8.1 Bilateral Grants with Industry	11
9 Partnerships and cooperations	11
9.1 International initiatives	11
9.2 International research visitors	12
9.2.1 Visits of international scientists	12
9.2.2 Visits to international teams	13
9.3 National initiatives	13
9.4 Collaborations with Industry	14
10 Dissemination	14
10.1 Promoting scientific activities	15
10.1.1 Scientific events: organisation	15
10.1.2 Scientific events: selection	15
10.1.3 Invited talks	15
10.1.4 Scientific expertise	16
10.1.5 Research administration	16
10.2 Teaching - Supervision - Juries	17
10.2.1 Teaching	17

10.2.2 Supervision	17
10.3 Popularization	17
11 Scientific production	18
11.1 Major publications	18
11.2 Publications of the year	19
11.3 Cited publications	20

Project-Team BOREAL

Creation of the Project-Team: 2022 June 01

Keywords

Computer sciences and digital sciences

- A3.1. – Data
- A3.2. – Knowledge
- A7.1.3. – Graph algorithms
- A7.2. – Logic in Computer Science
- A9. – Artificial intelligence
- A9.1. – Knowledge
- A9.8. – Reasoning

Other research topics and application domains

- B3.5. – Agronomy
- B6.5. – Information systems

1 Team members, visitors, external collaborators

Research Scientists

- Federico Ulliana [Team leader, University of Montpellier, Associate Professor Detachement, HDR]
- Jean-Francois Baget [Inria, Researcher]
- Pierre Bisquert [INRAE, Researcher]
- Nofar Carmeli [Inria, Researcher]
- David Carral Martinez [Inria, Researcher]

Faculty Members

- Michel Chein [University of Montpellier, Emeritus, HDR]
- Madalina Croitoru [University of Montpellier, Professor, HDR]
- Michel Leclère [University of Montpellier, Associate Professor]
- Marie-Laure Mugnier [University of Montpellier, Professor, HDR]

PhD Students

- Akira Charoensit [Inria, from Apr 2023]
- Eleazar Mbaiornom [LIRMM, from Feb 2023, (until December 2023)]
- Guillaume Perution-Kihli [Inria, (until december 2023)]
- Olivier Rodriguez [Inria, (until December 2023)]
- Mohamed Aziz Sfar Gandoura [EDF, CIFRE, (until December 2023)]

Technical Staff

- Florent Tornil [Inria, Engineer]

Interns and Apprentices

- Paul Fontaine [University of Montpellier, Intern, from May 2023 until Jul 2023]
- Sarah Michel [University of Montpellier, Intern, from Jun 2023 until Jun 2023]

Administrative Assistant

- Maeva Jeannot [Inria]

External Collaborators

- Patrice Buche [INRAE, HDR]
- Maxime Buron [University of Clermont Auvergne]

2 Overall objectives

Current information systems are grounded on the exploitation of data coming from an increasing number of heterogeneous sources. Today, coping with the variety of data requires novel paradigms for effectively accessing and querying information that adapt to the different types of sources, as well as declarative high-level languages to drive the data processing and data quality tasks.

BOREAL is a team working at the crossroads of *knowledge representation and reasoning* and *database theory*. The team focuses on the study of foundational and applied issues of *reasoning* in a context of *data variety*. More specifically, the team aims at deriving a better understanding of the logical fragments that are at the foundations of the frameworks used for exploiting corporate and Web data - and in particular *rule-based languages*. This will pave the way to novel automated-reasoning and graph-based techniques that can be put at service of data-centric applications exploiting heterogeneous and federated data. The team also aims at combining solid foundational and algorithmic work with software development and applications, with an emphasis on the field of agronomy.

3 Research program

The BOREAL team pursues a *knowledge-based data management* (KBDM) approach for tackling the grand challenges posed by data variety, with an important focus on the framework of *existential rules*. The idea of knowledge-based data management is to orchestrate the access to a complex information system made by federated databases through a three-layer architecture - also common to data-integration and ontology-based data access (OBDA). Under this prism, a set of *heterogeneous data sources* is connected to a *knowledge base* via a layer of *mappings*. The idea of KBDM is to define the business logic for data-centric applications at the knowledge base level, and then automatically translate the data-services towards the heterogeneous sources - through reasoning. This approach paves the way to a more principled use of complex information systems, with benefits to both data scientists, data curators, and administrators. What really characterizes the KBDM approach is the leveraging on formalized *domain-specific knowledge*, for abstracting on heterogeneous data and achieving high-quality of data-integration, and on expressive *rule-base languages* like existential rules (and extensions thereof), to drive the effective exploitation of data through reasoning.

Our project focuses on a set of issues related to knowledge-based data management we now describe.

Foundations of rule languages A great deal of the power of a KBDM system comes from its rule base. A prominent research direction for the team is the analysis and design of rule languages for reasoning on data. It is well understood that enriching a language with novel features can sensibly increase the complexity of the reasoning tasks. Our goal is hence to identify rules featuring decidable query answering and static analysis, and at the same time find good tradeoffs between their expressivity and complexity, so as to devise novel and practically useful rule-based frameworks.

Algorithms and optimizations for query answering Reasoning-driven data management needs optimization to effectively exploit large data. We target the design of efficient and scalable algorithms for query answering. Our goal is to devise novel hybrid approaches that combine materialization and virtualization strategies and account for the interplay between the components of the KBDM system (data, mappings, rules). Our ambition is also to build new bridges between knowledge representation and data-management by exploring the range of possibilities opened by the reuse of existing database technology to develop new reasoning systems.

Fine-grained complexity of query answering The query answering problem is at the heart of many reasoning tasks in KBDM. From a complexity analysis point of view, since the database to query can be voluminous, it is not always enough to know that a certain task can be done in polynomial time. Hence, an important goal for us is to study the fine-grained complexity (that is, to find the degree of the polynomial that bounds the number of operations required) as well as the enumeration complexity of the query answering problem. The aim of this research direction is to obtain the theoretical knowledge required for practical query optimization.

Architectures for knowledge-based data integration The realm of possibilities in heterogeneous data integration leads to the offspring of a family of KBDM architectures, one for each applicative context. Our goal is to study architectures inspired from emerging practical use-cases, including *federations* of independent sources as well as *multi-level* architectures where KBDM systems are stacked to progressively distill information and achieve high-value data. We also focus on the type of mappings required to cope with heterogeneity, because data may differ along several dimensions such as its format, refinement, dynamicity, and certainty; this is required to build a unified view of a complex information system.

Quality of knowledge-based data integration Knowledge-based data management systems can result in high quality data for users and applications. Yet, they also need mechanisms to assist data curators to constantly evaluate and improve all of their components towards the ultimate goal of matching the desired data integration level. Our aim is to investigate explanation mechanisms able to justify answers to queries and to point out inconsistencies in the data. We are also interested in techniques for deriving, within a knowledge-base, equivalent formulations of queries that are expressed outside of it, at the source level; these are critical for the verification of mappings and rules.

4 Application domains

4.1 Agronomy and agroecology

Agronomy is today more and more at the center of important debates around questions of environmental impact related to the practice of intensive agriculture, especially at large scale. Through our research collaborations with INRAE (National Research Institute for Agriculture, Food and Environment) and DFKI (German Institute for Artificial Intelligence) our goal is to contribute and to define new models, techniques, and applications, enabling a better exploitation of data generated in these fields so as to put it at the service of decision-making processes.

Agronomy is a strong expertise domain in the area of Montpellier. And indeed, BOREAL is a joint team with INRAE, and the team has established closed collaborations with two Montpellier research laboratories (UMR, “Unités Mixte de Recherche”), namely IATE and ABSys. These collaborations can also reach a larger extent, for example, in the context of the #DigitAg (Institute Convergences Agriculture Numérique, Section 9.3) our team participated to the joint Inria-INRAE “White Book” on digital agriculture which can be considered a manifesto of the current challenges posed by digital agriculture [30].

A major issue for IATE (Engineering of Agro-polymers and Emerging Technologies) is to model the transformation of products in agrifood chains (i.e., the chain of all processes leading from some raw material, such as plants, to the final products, including waste treatment). This modeling has several objectives. It provides better understanding of the processes from start to finish, which aids in decision making, with the aim of improving the quality of the products and decreasing the environmental impact (e.g., reducing waste, choosing right food packaging). There is a need for tools for making easier for data scientists to integrate and analyze the heterogeneous data resulting from agrifood chains.

A major issue for ABSys (Biodiversified Agrosystems) is the study of sustainable farming systems. It is now established that the restoration of sustainable farming systems requires the adoption of agroecological practices supporting the reintroduction of biodiversity in agroecosystems. Indeed, an agroecosystem should provide not only cash crops but also ecosystem services that support the durability of the farming systems itself. This leads to more complex agroecosystems including a higher number of plant species. There is thus a crucial need for tools that would assist users in the design of such new agroecosystems, from researchers in agronomy to agricultural advisors and farmers.

Beside INRAE, our team also collaborates with two DFKI teams located in Osnabrück and Kaiserslautern in the context of a bilateral project Inria-DFKI (“R4Agri”, Section 9). From an applicative perspective, the major issue targeted by this project is the development of monitoring tools based on reasoning which can equip robotic or mechanic devices used in agricultural farms. This can be used to enhance the agricultural processes but also to enforce regulations, for instance by assessing that the spraying of chemicals remains at a safe distance from river borders. In this context, there is a need for tools allowing one to interpret and analyze the number of types of sensor data that are generated.

5 Highlights of the year

- The team had 8 papers in top venues (A*) of knowledge representation and reasoning and databases.
- The team welcomed a new Inria researcher, Nofar Carmeli, expert in the theory of data management.
- The team intensified its involvement in the database community, as witnessed by its international publications in major venues and the 5 papers presented at the French National Database Conference “BDA 2023” held in Montpellier.
- The team thanks Madalina Croitoru for her involvement in the Graphik team and in the creation of Boreal. Beginning January 2024, Madalina Croitoru will be moving to the IDH team at LIRMM to pursue her interests in human-robot interactions. The PhD candidates she supervises, Eleazar Mbaionom and Mohamed Aziz Sfar Gandoura, will also move from Boreal to the IDH team.

5.1 Awards

The paper "Scalable Reasoning on Document-Stores by Instance-Aware Query Rewriting" by Olivier Rodriguez, Federico Ulliana and Marie-Laure Mugnier obtained a best paper award at the 2023 French National Database Conference BDA “Gestion de Données – Principes, Technologies et Applications ” [26]

6 New software, platforms, open data

Since 2021, **InteGraal** is the main (Java) platform developed by the team to reason on heterogeneous data with existential rules. Our need to efficiently access tree data such as JSON data successively led us to the development of the **TreeForce** library for reasoning on NoSQL document stores.

This year, InteGraal has been used in the context of our collaborations with INRAE as well as in the context of the bilateral project INRIA-DFKI “R4Agri” focusing on reasoning on agricultural data [23]. InteGraal has been used to realize a proof-of-concept in our collaboration with industrial partners (see Section 9). InteGraal has been used in the context of the internship of Paul Fontaine (L3 UM) and Sarah Michel (L2 UM) as well as for experimental analysis in the PhD thesis of Akira Charoensit (funded by INRIA-DFKI “R4Agri”) and Olivier Rodriguez (funded by ANR “CQFD”), see Section 10.2.2. In general, InteGraal is a federating tool for the team, which follows a monthly-based software development cycle for its advancement.

6.1 New software

6.1.1 InteGraal

Name: InteGraal : Knowledge-Representation and Reasoning for Data Integration

Keywords: Knowledge Bases, Data integration, Knowledge representation, Automated Reasoning, Heterogeneous Data, Knowledge Graphs

Scientific Description: InteGraal is a tool for integrating and reasoning on heterogeneous and federated data. The tool embodies algorithms and techniques developed at the crossroads between the fields of knowledge representation and reasoning and data management. From the historic point of view, this tool is the result of a complete re-engineering of the Graal tool, whose API and functionalities have been completely updated. Also, with respect to Graal, the tool is very much oriented towards data integration.

Functional Description: InteGraal has been designed in a modular way, in order to facilitate software reuse and extension. It should make it easy to test new scenarios and techniques, in particular by combining algorithms. The main features of Graal are currently the following: (1) internal storage to store data by using a SQL or RDF representation (Postgres, MySQL, HSQL, SQLite, Remote

SPARQL endpoints, Local in-memory triplestores) as well as a native in-memory representation (2) data-integration capabilities for exploiting federated heterogeneous data-sources through mappings able to target systems such as SQL, RDE, and black-box (eg. Web-APIs) (3) algorithms for query-answering over heterogeneous and federated data based on query rewriting and/or forward chaining (or chase)

Release Contributions: 2023. Mappings for integrating heterogeneous data. Compilation-based query rewriting. Command line interface. 2022: First release, software deposit with Apache 2 licence. 2021: Functional specification, design and development of a major improved version of the tool. Started refactoring of the API, and of several modules for knowledge base representation, data storage, query answering and forward-chaining reasoning (chase). Started the development of new modules for handling heterogeneous data: mappings and federations.

News of the Year: This year we added a major module including mappings for integrating heterogeneous data. We also refactored and improved compilation-based query rewriting from the previous Graal tool. We added a command line interface to interact with the platform. Moreover, the tool has been used in several internship and thesis of team, as well as for collaborations with our research partners and industrial actors.

URL: <https://gitlab.inria.fr/rules/integraal>

Authors: Florent Tornil, Guillaume Perution-kihli, Clément Sipieter, Federico Ulliana, Jean-Francois Baget, Pierre Bisquert, David Carral Martinez, Michel Leclère, Marie-laure Mugnier

Contact: Federico Ulliana

6.1.2 TreeForce

Keywords: JSon, Databases, Knowledge Bases, Automated Reasoning, Rewriting, NoSQL, Data integration, Knowledge representation, Heterogeneous Data

Functional Description: TreeForce is a java tool for reasoning on tree data. It leverages on query rewriting techniques and NoSQL document oriented key-value stores. This library can be seen as a general toolbox for implementing reasoning techniques tailored for tree-shaped data and rules. It is composed of two main modules. The first includes generic data structures and algorithms for trees and tree-automata. The second includes automata-based query rewriting techniques as well as efficient evaluation techniques for large sets of rewritings.

Release Contributions: 2023. ArangoDB wrapper. Code improvement. 2022. Novel instance-aware rewriting and evaluation algorithms. Introduced summarization, partitioning and parallelization techniques. 2021: First version of TreeForce. Automata for unordered tree languages. Automata-based query-rewriting algorithms. MongoDB wrapper.

News of the Year: 2023. This year, we added support for querying ArangoDB and performed several code improvements.

Contact: Federico Ulliana

7 New results

Before presenting this year's results, we introduce some general notions about our main research focus, namely *Knowledge-Based Data Management with existential rules* (Section 7.1). This allows us to put into context our results on foundations and algorithms for rule-based reasoning (Sections 7.2 and 7.3).

7.1 Knowledge-Based Data Management with existential rules

This broad topic encompasses research areas such as ontology-mediated query answering (OMQA), data integration (DI), and ontology-based data access (OBDA) because of the expressivity of existential rule languages and the complexity of integration architectures it embraces.

Existential rules. Existential rules are first-order-logic formulas representing implications of the form $\forall X \forall Y [Body(X, Y) \rightarrow \exists Z Head(X, Z)]$ where *Body* and *Head* are positive conjunctions of atoms without functional symbols, and *Head* can have existentially quantified variables. These rules allow one to model complex relationships over the domains of interest, and at the same time dispose of a value invention mechanism through existentially quantified variables. This makes them suitable for many data and knowledge tasks on both open and closed domains. As a result, existential rules are ubiquitous in many fields. They are used to model dependencies, schema mappings, and expressive queries in databases. They are used as ontological languages as a valid complement to Description Logics, and at the same time as a generalization of so-called Horn Description Logics which lay at the foundations of important Semantic Web standards.

Rule-based query answering Given a query Q , a database D , and a set of rules R , query answering asks to determine whether $D, R \models Q$ (where \models denotes standard first-order logic entailment), that is if the query Q is a logical consequence of the knowledge base made by the database D and the rules R . In the field of knowledge representation and reasoning, rule-based query answering is studied for rules expressing ontologies and referred as ontology-mediated query answering (OMQA). Formalisms such as Description Logics and Existential Rules (a.k.a, Tuple-Generating-Dependencies, or Datalog \pm) are typically targeted for expressing ontologies. Overall, the main emphasis of this topic is in the study of *rule languages* and the role they play in query answering.

Rule-based query answering over heterogeneous and federated data In this context, the problem formulation remains similar, however the database D is replaced by a more complex notion of *federation* $(\mathcal{D}, \mathcal{M}, \mathcal{S})$ where \mathcal{D} is a collection of heterogeneous data sources, \mathcal{S} is a global integration schema, and \mathcal{M} is a set of mappings linking the datasources in \mathcal{D} to the global schema \mathcal{S} . This framework is at the foundations of data integration (DI) in databases and of ontology-based data access (OBDA) in knowledge representation and reasoning. OBDA focuses on global integration schemes and rules built on ontologies enabling query rewriting, while DI is more concerned with rules representing data-dependencies. Overall, both give more emphasis to *heterogeneous and federated data* in rule-based query answering.

Reasoning strategies for query answering The two prominent strategies for rule-based query answering are *materialization* (also known as saturation, or forward-chaining) and *virtualization* (also known as query rewriting, or backward chaining). Both can be seen as ways of reducing query answering (which involves reasoning) into classical query evaluation. Materialization amounts at storing the inferences enabled by rules, thereby obtaining an extended database, on which queries are evaluated. Query rewriting amounts at compiling relevant rules into the query, thereby obtaining a rewritten query (usually a union of queries), which is evaluated on the (unaltered) database. Both approaches have their own strengths, and at the basis of this duality is the fact that while materialization is independent of queries, rewriting is independent of the database. Hence, each strategy better suits certain applicative scenarios, and both can possibly be combined thereby resulting in hybrid approaches.

Contributions This year, we studied a number of theoretical, algorithmic, and applied questions of knowledge-based data management. Our main contributions cover the following topics:

- foundational issues (Section 7.2) related to the termination of reasoning strategies using existential rules or their extension (notably with disjunctive heads);
- algorithms for querying specific data sources (Section 7.3), such as relational databases or JSON document stores provided with rules;
- the application of knowledge representation languages to modelling and reasoning on complex systems (Section 7.4).

This work led to the publications presented next.

Complementing methodological work, it is worth noting that we also pursued an important team effort in the development of tools for rule-based query answering over heterogeneous and federated such as InteGraal and TreeForce (see Section 6.1).

7.2 Foundations of rule-based reasoning and query answering

Participants: Jean-François Baget, Pierre Bisquert, Nofar Carmeli, David Carral, Michel Leclère, Marie-Laure Mugnier, Guillaume Pérution-Kihli, Federico Ulliana, Lukas Gerlach (*Technische Universität Dresden*), Sebastian Rudolph (*Technische Universität Dresden*).

Our main contributions for this year revolved around the static analysis problems related to the termination of reasoning strategies. Indeed, it is well understood that as both materialization and virtualization rely on a fixpoint operator, they may not terminate. Thus, it then becomes essential to be able to decide whether a given set of rules enjoys the termination of a given reasoning strategy before this can actually be executed.

Chase termination The most widespread materialization procedure for existential rules is called *the chase*. There are several variants of the chase which differ in terms on their computational costs and termination capabilities. The most powerful variant is the *core chase* which is known to terminate if and only if the knowledge base admits a *finite universal model* (a model that can be embedded into any other model of the knowledge base, hence is sufficient to answer unions of conjunctive queries). Other variants include the *restricted* (aka standard) and *semi-oblivious* (aka skolem) chase. These are less aggressive on the suppression of redundancies produced by materialization and may not terminate even if a finite universal model of the knowledge base exists. A ruleset that enjoys the termination of the core chase over any possible database instance, is also called a *finite expansion set* (FES).

Rewriting termination First-Order (FO) rewritability is a property enjoyed by the sets of existential rules for which every input conjunctive query (CQ) admits a finite equivalent first-order-logic formula, called its *rewriting*. Here, *equivalent* is intended in the sense of query answering, and it means that whenever there is a database that (together with the rules) entails the input query, then the same database (without the rules) directly entails the rewriting. This notion coincides with that of *finite unification set* (FUS), which are the sets of existential rules for which breadth-first query rewriting based on piece-unifiers terminates.

7.2.1 Terminating conditions for disjunctive rules

The disjunctive skolem chase is a sound, complete, and potentially non-terminating procedure for solving Boolean conjunctive query entailment over knowledge bases of disjunctive existential rules. We develop novel acyclicity and cyclicity notions for this procedure; that is, we develop sufficient conditions to determine chase termination and non-termination. Our empirical evaluation shows that our novel notions are significantly more general than existing criteria.

- Published at AAAI Conference on Artificial Intelligence (AAAI-2023) [18] with Lukas Gerlach from TU Dresden.
- Published at Principles of Knowledge Representation and Reasoning Conference (KR-2023) [17] with Lukas Gerlach.

7.2.2 Rewriting with disjunctive existential rules and mappings

We consider the issue of answering unions of conjunctive queries (UCQs) with disjunctive existential rules and mappings. While this issue has already been well studied from a chase perspective, query rewriting within UCQs has hardly been addressed yet. We first propose a sound and complete query rewriting operator, which has the advantage of establishing a tight relationship between a chase step and a rewriting step. The associated breadth-first query rewriting algorithm outputs a minimal UCQ-rewriting when one exists. Second, we show that for any “truly disjunctive” nonrecursive rule, there exists a conjunctive query that has no UCQ-rewriting. It follows that the notion of finite unification sets (FUS), which denotes sets of existential rules such that any UCQ admits a UCQ-rewriting, seems to have little relevance in this

setting. Finally, turning our attention to mappings, we show that the problem of determining whether a UCQ admits a UCQ-rewriting through disjunctive mappings is undecidable. Based on these results, we conclude with a number of open problems.

- Published at Principles of Knowledge Representation and Reasoning Conference (KR-2023) [19] and at Bases de Données Avancées (BDA-2023) [25]

7.2.3 Decidability of bounded treewidth infinite core chase

The core chase, a popular algorithm for answering conjunctive queries (CQs) over existential rules, is guaranteed to terminate and compute a finite universal model whenever one exists, leading to the equivalence of the universal-model-based and the chase-based definitions of finite expansion sets (FES) – a class of rulesets featuring decidable CQ entailment. In case of non-termination, however, it is non-trivial to define a “result” of the core chase, due to its non-monotonicity. This causes complications when dealing with advanced decidability criteria based on the existence of (universal) models of finite treewidth. For these, sufficient chase-based conditions have only been established for weaker, monotonic chase variants. This paper investigates the – prima facie plausible – hypothesis that the existence of a treewidth-bounded universal model and the existence of a treewidth-bounded core-chase sequence coincide – which would conveniently entail decidable CQ entailment whenever the latter holds. Perhaps surprisingly, carefully crafted examples show that both directions of this hypothesized correspondence fail. On a positive note, we are still able to define an aggregation scheme for the infinite core chase that preserves treewidth bounds and produces a finitely universal model, i.e., one that satisfies exactly the entailed CQs. This allows us to prove that the existence of a treewidth-bounded core-chase sequence does warrant decidability of CQ entailment (yet, on other grounds than expected). Hence, for the first time, we are able to define a chase-based notion of bounded treewidth sets of rules that subsumes FES.

- Published at Principles of Database Systems Conference (PODS-2023) [1] with Sebastian Rudolph.

7.2.4 Effects of rule normalization

Most algorithms developed for existential rules rely upon some normal form that simplifies technical developments. For instance, a common assumption is that rule heads are atomic, i.e., restricted to a single atom. Such assumptions are considered to be made without loss of generality as long as all sets of rules can be normalized while preserving entailment. However, an important question is whether the properties that ensure the decidability of reasoning are preserved as well. We provide a systematic study of the impact of these procedures on the different chase variants with respect to chase (non-)termination and FO-rewritability. This also leads us to study open problems related to chase termination of independent interest.

- Published at Bases de Données Avancées (BDA-2023) [24] with Michaël Thomazo (Valda). Previously published at Principles of Knowledge Representation and Reasoning Conference (KR-2022).

7.3 Algorithms for querying specific data sources

Participants: David Carral, Nofar Carmeli, Marie-Laure Mugnier, Olivier Rodriguez, Florent Tornil, Federico Ulliana, Nikolaos Tziavelis (*Northeastern University, Boston*), Wolfgang Gatterbauer (*Northeastern University, Boston*), Mirek Riedewald (*Northeastern University, Boston*), Benny Kimelfeld (*Israel Institute of Technology*).

7.3.1 Fine grained complexity of query answering

We studied efficient algorithms for Quantile Join Queries, abbreviated as %JQ. A %JQ asks for the answer at a specified relative position (e.g., 50% for the median) under some ordering over the answers to a Join Query (JQ). The goal is to avoid materializing the set of all join answers, and to achieve quasilinear time

in the size of the database, regardless of the total number of answers. We developed a new approach to solving %JQ and showed how this approach allows not just to recover known results, but also generalize them and resolve open cases. The benefit and generality of our approach is shown by using it to establish several new complexity results. First, we prove the tractability of min and max for all acyclic JQs. Second, we extend the previous %JQ dichotomy for sum to all partial sums (over all subsets of the attributes). Third, we handle the intractable cases of sum by devising a *deterministic* approximation scheme that applies to *every acyclic JQ*.

- Published at Principle of Database Systems (PODS-2023) [22] with Nikolaos Tziavelis Wolfgang Gatterbauer, Mirek Riedewald, and Benny Kimelfeld.

Again on the fine-grained complexity of database queries, two additional works have been published this year: a conference paper at Principle of Database Systems (PODS-2023) [28] (also presented at Bases de Données Avancées (BDA-2023) and a journal paper at Transactions on Database Systems (TODS-2023) [29]. These have been done by members of our team before joining Boreal.

7.3.2 Querying JSON document stores

We studied the problem of reasoning on data trees, typically encoded in JSON, which are ubiquitous in data-driven applications. This ubiquity makes urgent the development of novel techniques for querying heterogeneous JSON data in a flexible manner. We designed a rule language for JSON, called constrained tree-rules, whose purpose is to provide a high-level unified view of heterogeneous JSON data and infer implicit information. As reasoning with constrained tree-rules is undecidable, we identify a relevant subset featuring tractable query answering, for which we proposed a novel automata-based query rewriting algorithm. Furthermore, we proposed a novel approach for leveraging NoSQL document stores by means of the instance-aware query-rewriting technique, which uses summaries of data to make the rewriting process both finite and efficient. We presented an extensive experimental analysis on large collections of several million JSON records.

- Published at Proceedings of the VLDB Endowment (PVLDB-2023) [14] and Bases de Données Avancées (BDA-2023) [26].

7.4 Applications of logics and rule-based languages

Participants: Jean-François Baget, Pierre Bisquert, Madalina Croitoru, Michel Leclère, Marie-Laure Mugnier, Guillaume Pérution-Kihli, Mohammed Aziz Sfar Gandoura, Florent Tornil, Federico Ulliana.

7.4.1 Systems for reasoning on heterogeneous and federated data

We demonstrated our approach to integrating and reasoning over heterogeneous and federated data by using our main tool: InteGraal (see Section 6). InteGraal is a highly modular tool constituted by two main components. The first is the data-integration layer which allows the users to build a federated factbase over a collection of sources. The second is the automated reasoning layer, which provides powerful means for the declarative exploitation of data through the expressive formalism of existential rules. The demonstration showcased the use of the tool in use-cases of data exploitation inspired by our collaboration with INRAE, and more precisely in the preparation of data for supporting machine learning and decision support.

- Published at Bases de Données Avancées (BDA-2023) [23].

7.4.2 Knowledge Representation formalisms for testing power plants

In the context of Mohammed Aziz Sfar Gandoura CIFRE PhD funded by EDF, we focussed on the application of a known knowledge representation formalism, namely LTL (Linear Temporal Logic), to do model

checking on logical diagrams (LD). LD are a type of functional specification used for logical controllers in many nuclear power plants. The goal is to check properties on LDs and to generate counter examples serving as validation tests for logical controllers. We proposed a sound and complete LTL encoding framework for LDs allowing the use of model checking (MC) and evaluate different MC techniques on real world LD to efficiently generate counterexamples for verifiable properties.

- Published at Formal Methods for Industrial Critical Systems Conference (FMICS 2023) [20]

7.4.3 Food packaging in agronomy

In collaboration with the IATE UMR at INRAE, we proposed an environmental scoring tool for food packaging based on the assessment of three key pillars of packaging sustainability: Materials, Functionality and Post-Usage fate. A participatory process involving relevant food-packaging experts and end users was applied to define the relevant criteria for each pillar. Then, the packaging options for the same food have been ranked according to the Borda voting rule, considering the individual rankings obtained for the various pillars. The proposed methodology was applied to three commercial (milk and sugar) and non-commercial (strawberry) packaging case studies. The obtained ranking has been analyzed with respect to current knowledge in the field.

- Published at Packaging Technology and Science 2023 [12]

7.4.4 Human-robot interactions

In collaboration with the IDH team at LIRMM, we investigated Human-Robot Interactions (HRI) in complex social settings. A robot can affect its social environment beyond the person who is interacting with it. Therefore, we examined the effect of different robot shapes in a multi-person context during dance routines, to understand how the design of the robot enhances the artistic process and through which factors human preferences are being shaped within a novel third party setting human-robot-human interaction (HRHI).

- Published at IEEE International Conference on Robot and Human Interactive Communication (RO-MAN 2023) [16]

8 Bilateral contracts and grants with industry

8.1 Bilateral Grants with Industry

Participants: Madalina Croitoru, Mohammed Aziz Sfar Gandoura.

Madalina Croitoru supervised the CIFRE thesis of M.A. Sfar Gandoura founded by EDF-Research Paris. The thesis started in January 2022. The topic is the automatic generation of testing scenarios for the verification of complex power plant systems.

9 Partnerships and cooperations

9.1 International initiatives

R4Agri

Participants: Pierre Bisquert, David Carral, Akira Charoensit, Marie-Laure Mugnier, Florent Tornil, Federico Ulliana

Title: “R4Agri” - Reasoning on Agricultural Data: Integrating Metrics and Qualitative Perspectives

Partner Institution(s):

- Inria
- DFKI, Germany

Date/Duration: 42 months, started 01/01/2022

Additional info: AI tools supporting competitive and sustainable agriculture need to exploit highly diverse kinds of data and knowledge, from raw data provided by sensors to high level expertise knowledge. Taking numerical agriculture as the targeted application domain, the overall goal of the R4Agri project is to provide a framework for reasoning about knowledge based on heterogeneous data, with a focus on multi-modal and multi-scale sensor data. Main challenges include context-dependent interpretation of sensor data, which involves reasoning about prior knowledge, and query answering techniques that exploit domain knowledge and accommodate the specificities of data sources in a flexible manner. The application potential in this field of world-wide societal and ecological impact will be demonstrated in realistic use cases. project.inria.fr/r4agri/

Federico Ulliana is part of the Inria CDT (Commission Développement Technologique) since May 2022. This committee is in charge of reviewing and following software development projects within the Inria centre at Université Côte d’Azur.

9.2 International research visitors

9.2.1 Visits of international scientists

Karima Belmabrouk

Status: Associate Professor

Institution of origin: Oran University

Country: Algeria

Dates: two weeks, January 2023

Context of the visit: Research stay

Lucía Gómez Álvarez

Status: Postdoctoral Researcher

Institution of origin: TU Dresden

Country: Germany

Dates: two weeks, January 2023

Context of the visit: Research stay

Tim Lyons

Status: Postdoctoral Researcher

Institution of origin: TU Dresden

Country: Germany

Dates: two weeks, February 2023

Context of the visit: Research stay

Piotr Ostropolski-Nalewaja

Status: Postdoctoral Researcher

Institution of origin: TU Dresden

Country: Germany

Dates: two weeks, February 2023

Context of the visit: Research stay

Sebastian Rudolph

Status: Full Professor

Institution of origin: TU Dresden

Country: Germany

Dates: two weeks, February 2023

Context of the visit: Research stay

Jerzey Marcinkowski

Status: Full Professor

Institution of origin: Wroclaw University

Country: Poland

Dates: two weeks, February 2023

Context of the visit: Research stay

9.2.2 Visits to international teams**David Carral**

Visited institution: TU Dresden

Country: Germany

Dates: three weeks, December 2023

Context of the visit: Research collaboration

Mobility program/type of mobility: research stay

9.3 National initiatives

CQFD (ANR PRC, Jan. 2019-Dec. 2024)

Participants: Jean-François Baget, Pierre Bisquert, Nofar Carmeli, David Caral, Michel Leclère, Marie-Laure Mugnier, Guillaume Pérution-Kihli, Olivier Rodriguez, Florent Tornil, Federico Ulliana, Antoine Amarilli (*Télécom ParisTech*), François Goasdoué (*Institut de Recherche en Informatique et Systèmes Aléatoires - IRISA*), Pierre Bourhis (*SPIRALS*), Ioana Manolescu (*CEDAR*), Michaël Thomazo (*VALDA*), Meghyn Biennvenu (*Centre de Recherche en Informatique, Signal et Automatique de Lille - CRISTAL*), Marie-Christine Rousset (*Institut d'Informatique et de Mathématiques Appliquées de Grenoble - IMAG*), Fabrice Jouanot (*Institut d'Informatique et de Mathématiques Appliquées de Grenoble - IMAG*)

CQFD (Complex ontological Queries over Federated heterogeneous Data), coordinated by Federico Ulliana (BOREAL), involves participants from Inria Saclay (CEDAR team), Inria Paris (VALDA team), Inria Nord Europe (SPIRALS team), IRISA, LIG, LTCI, and LaBRI. The aim of this project is to tackle two crucial challenges in OMQA (Ontology Mediated Query Answering), namely, heterogeneity, that is, the possibility to deal with multiple types of data-sources and database management systems, and federation, that is, the possibility of cross-querying a collection of heterogeneous datasources. By featuring 8 different partners in France, this project aims at consolidating a national community of researchers around the OMQA issue.

www.lirmm.fr/cqfd/

Convergence institute #DigitAg (2017-2026)

Participants: Jean-François Baget, Patrice Buche, Madalina Croitoru, Marie-Laure Mugnier, Federico Ulliana.

Located in Montpellier, #DigitAg (for Digital Agriculture) gathers 17 founding members: research institutes, including Inria, the University of Montpellier and higher-education institutes in agronomy, transfer structures and companies. Its objective is to support the development of digital agriculture. BOREAL is involved in this project on the issues of designing data and knowledge management systems adapted to agricultural information systems, and of developing methods for integrating different types of information and knowledge (generated from data, experts, models). A PhD thesis (Elie Najm, 2019-2022) investigated knowledge representation and reasoning for the design of new agroecological systems, in collaboration with the research laboratory ABSys - Biodiversified Agrosystems (formerly UMR SYSTEM).

www.hdigitag.fr/en/

9.4 Collaborations with Industry

We conducted a 6 month collaboration with data scientists from the CEMIS team at Naval Group. This collaboration targeted a central topic for the team, namely the use of reasoning on knowledge for integrating and exploiting heterogeneous databases. The collaboration outcomes include (1) a proof of concept for the case studied which is based on InteGraal (see Section 6) and (2) a collaborative 3-years project proposal.

10 Dissemination

Participants: Jean-François Baget, Pierre Bisquert, Nofar Carmeli, David Carral, Madalina Croitoru, Michel Leclère, Marie-Laure Mugnier, Florent Tornil, Federico Ulliana.

10.1 Promoting scientific activities

10.1.1 Scientific events: organisation

At the LIRMM level, Madalina Croitoru is co-chair of the transversal axis and workshop on Artificial Intelligence for human-robot interaction.

10.1.2 Scientific events: selection

Chair of conference program committees Teams members have acted as area chairs in the following conferences:

- KR 2023 (20th International Conference on Principles of Knowledge Representation and Reasoning, Area chair (Marie-Laure Mugnier)
- ICCS 2023 (28th International Conferences on Conceptual Structures): Steering Committee - Madalina Croitoru

Member of the conference program committees

International

- AAI 2023 (37th AAI Conference on Artificial Intelligence): PC member - Federico Ulliana
- AAMAS 2023 (22st International Conference on Autonomous Agents and Multiagent Systems): Senior PC member - Madalina Croitoru
- KR 2023 (20th International Conference on Principles of Knowledge Representation and Reasoning): PC member - David Carral and Michel Leclère
- DL 2023 (36th International Workshop on Description Logics): PC member - David Carral
- PODS 2023 (Symposium on Principles of Database Systems): PC member - Nofar Carmeli
- PODS 2024 (Symposium on Principles of Database Systems): PC member - Nofar Carmeli

National

- IC 2023 (Ingénierie des Connaissances 2023) - PC member - Michel Leclère
- BDA 2023 (Gestion de Données – Principes, Technologies et Applications) - PC member - Nofar Carmeli

10.1.3 Invited talks

- Nofar Carmeli - Invited talk *Accessing Answers to Conjunctive Queries with Ideal Time Guarantees* - DL 2023 (36th International Workshop On Description Logics), September 2nd
- Nofar Carmeli - Invited talk *Query answering: tractability beyond acyclicity* - Graph and Databases workshop, Lyon, March 3rd
- Nofar Carmeli - Invited participation and talks in two workshops of the Logic and Algorithms in Database Theory and AI program at the Simons Institute, Berkeley, California
 - *Accessing Answers to Unions of Conjunctive Queries with Ideal Time Guarantees* at the Fine-Grained Complexity, Logic, and Query Evaluation workshop, September 25th
 - *Direct Access for Conjunctive Queries with Aggregation* at the Logic and Algebra for Query Evaluation workshop, November 14th

10.1.4 Scientific expertise

Award committees

- Nofar Carmeli was a member of the ICDT (International Conference on Database Theory) 2024 test-of-time award committee.
- Marie-Laure Mugnier was a member of the jury for the SIF / Gilles Kahn prize, which rewards the best PhD thesis in computer science at the national level.

Recruitment committees

- Federico Ulliana was a member of the recruitment committee for a Assistant Professor position, (Polytech, U. Montpellier).
- Marie-Laure Mugnier was a member of recruitment committees for a Professor position (IUT, U. Montpellier) and an Assistant Professor position (Science Faculty, U. Bordeaux).

PhD / HDR juries

- Nofar Carmeli was a member of the PhD committee for Caroline Brosse, U. Clermont Auvergne
- Marie-Laure Mugnier was a member of the PhD committee for Hui Yang, U. Paris-Saclay (May 2023)
- Marie-Laure Mugnier was a member of the PhD committee for Julie Cailler, U. Montpellier (December 2023)

10.1.5 Research administration

- Madalina Croitoru has been a deputy member of the CNU section 27 (Computer Science) since September 2019.
- Madalina Croitoru was a deputy director of the Computer Science Department at the Faculty of Science, University of Montpellier from September 2021 to January 2023.
- Marie-Laure Mugnier has been president of the “Section 27 Committee” (Computer Science) of the University of Montpellier since July 2021.
- Marie-Laure Mugnier has been a member of the Council (the Human Resources Commission) of the Scientific Department MIPS (Mathematics Informatics Physics and Systems) of the University of Montpellier since 2016.
- Jean-François Baget represents the Boreal team at LIRMM’s Computer Science Department meetings (Comité des Projets INFO) and at LIRMM’s Laboratory teams meetings (CIEL) since 2021.
- Guillaume Pérution-Kihli has been a PhD student member of LIRMM’s CL (Conseil de Laboratoire) and has been representing students at the doctoral school I2S.
- Guillaume Pérution-Kihli has been elected president of the PhD students council.
- Federico Ulliana is part of the Inria CDT (Commission Développement Technologique) since May 2022. This committee is in charge of reviewing and following software development projects within the Inria centre at Université Côte d’Azur.

10.2 Teaching - Supervision - Juries

10.2.1 Teaching

Madalina Croitoru, Michel Leclère, and Marie-Laure Mugnier, do an average of 200 teaching hours per year at the Computer Science department of the Science Faculty. They are in charge of courses in Logics (Licence), Artificial Intelligence (Master), Knowledge Representation and Reasoning (Master), Theory of Data and Knowledge Bases (Master), Social and Semantic Web (Master) and Multi-Agent Systems (Master). Concerning full-time researchers in 2023, Jean-François Baget, David Carral, and Federico Ulliana (on secondment from Montpellier University) taught in the Computer Science Master; topics include Theory of Data and Knowledge Bases, Datawarehouses, Big-Data and NoSQL systems.

Guillaume Pérution-Kihli has been coaching students from University of Montpellier for SWERC since 2020.

10.2.2 Supervision

PhD

- PhD. Guillaume Pérution-Kihli, “Data Management in the Existential Rule Framework: Translation of Queries and Constraints”. Supervisors: Michel Leclère and Marie-Laure Mugnier. Started in September 2020 and defended in December 2023 [27].
- PhD. Olivier Rodriguez, “Querying key-value store under semantic constraints”. Supervisors: Federico Ulliana and Marie-Laure Mugnier. Started February 2019 and suspended in December 2023.

The following PhD theses are in progress:

- Akira Charoensit "Explainable Artificial Intelligence for Rule-Based Query Languages". Supervisors, David Carral, Pierre Bisquert, Federico Ulliana. Started in May 2023.
- David Carral is co-supervising with Michael Thomazo (Inria VALDA) the PhD of Lucas Larroque, started in October 2023. Lucas Larroque is a member of the Inria VALDA team.
- Mohammed Aziz Sfar Gandoura, “Génération de scénarios de tests pour les systèmes de contrôle-commande logique : une application pour les centrales nucléaires palier N4”. Supervisors: Madalina Croitoru and Dina Irofiti (EDF). Started in January 2022. As already mentioned, this PhD will continue in the context of the IDH team at LIRMM.

Engineers Many team members (Federico Ulliana, Michel Leclère, Pierre Bisquert, Marie-Laure Mugnier, and Jean-François Baget) follow on a regular basis the development of the InteGraal software and jointly supervise Florent Tornil’s work as an engineer.

Interns This year, the team has welcomed two interns.

- Paul Fontaine (L3 U. Montpellier, 2 months) worked on an algorithm for forgetting predicates for a fragment of existential rules. Supervisors: Pierre Bisquert and Michel Leclère.
- Sarah Michel (L2 U. Montpellier, 1 month) worked on an interactive shell for InteGraal. Supervisors: Pierre Bisquert and Michel Leclère.

10.3 Popularization

Nofar Carmeli was a judge and problem setter in the Southwestern Europe Regional Contest (SWERC) of the International Collegiate Programming Contest (ICPC) 2023-2024.

Jean-François Baget made a demonstration of the Integraal Software at the “Journées Portes Ouvertes” of LIRMM (october 2023). *Les machines à voter rêvent-elles de moutons macronistes.*

11 Scientific production

11.1 Major publications

- [1] J.-F. Baget, M.-L. Mugnier and S. Rudolph. ‘Bounded Treewidth and the Infinite Core Chase: Complications and Workarounds toward Decidable Querying’. In: *Proceedings of the 42nd ACM SIGMOD-SIGACT-SIGAI Symposium on Principles of Database Systems*. SIGMOD/PODS 2023 - International Conference on Management of Data. Seattle, WA, United States: ACM, 18th June 2023, pp. 291–302. DOI: [10.1145/3584372.3588659](https://doi.org/10.1145/3584372.3588659). URL: <https://hal-lirmm.ccsd.cnrs.fr/lirmm-04320944>.
- [2] C. Bourgaux, D. Carral, M. Krötzsch, S. Rudolph and M. Thomazo. ‘Capturing Homomorphism-Closed Decidable Queries with Existential Rules (Extended Abstract)’. In: *Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence*. IJCAI-ECAI 2022 - 31st International Joint Conference on Artificial Intelligence - 25th European Conference on Artificial Intelligence. Vienna, Austria, 23rd July 2022, pp. 5269–5273. DOI: [10.24963/ijcai.2022/733](https://doi.org/10.24963/ijcai.2022/733). URL: <https://hal.archives-ouvertes.fr/hal-03801049>.
- [3] N. Carmeli and L. Segoufin. ‘Conjunctive Queries With Self-Joins, Towards a Fine-Grained Complexity Analysis’. In: PODS’23. Seattle, United States, 18th June 2023. URL: <https://inria.hal.science/hal-04136055>.
- [4] N. Carmeli, N. Tziavelis, W. Gatterbauer, B. Kimelfeld and M. Riedewald. ‘Tractable Orders for Direct Access to Ranked Answers of Conjunctive Queries’. In: *ACM Transactions on Database Systems* 48.1 (31st Mar. 2023), pp. 1–45. DOI: [10.1145/3578517](https://doi.org/10.1145/3578517). URL: <https://inria.hal.science/hal-04278100>.
- [5] D. Carral, L. Larroque, M.-L. Mugnier and M. Thomazo. ‘Normalisations of Existential Rules: Not so Innocuous!’ In: KR 2022 - 19th International Conference on Principles of Knowledge Representation and Reasoning. Haifa, Israel, 2022, pp. 102–111. URL: <https://hal-lirmm.ccsd.cnrs.fr/lirmm-03762686>.
- [6] L. Gerlach and D. Carral. ‘Do Repeat Yourself: Understanding Sufficient Conditions for Restricted Chase Non-Termination’. In: KR 2023 - 20th International Conference on Principles of Knowledge Representation and Reasoning. Proceedings of the 20th International Conference on Principles of Knowledge Representation and Reasoning. Rhodes, Greece: International Joint Conferences on Artificial Intelligence Organization, 2nd Sept. 2023, pp. 301–310. DOI: [10.24963/kr.2023/30](https://doi.org/10.24963/kr.2023/30). URL: <https://hal.science/hal-04212137>.
- [7] L. Gerlach and D. Carral. ‘General Acyclicity and Cyclicity Notions for the Disjunctive Skolem Chase’. In: *AAAI-23 Technical Tracks 5*. AAAI 2023 - 37th Conference on Artificial Intelligence. Vol. 37. Technical Track on Knowledge Representation and Reasoning 5. Washington, United States, 27th June 2023, pp. 6372–6379. DOI: [10.1609/aaai.v37i5.25784](https://doi.org/10.1609/aaai.v37i5.25784). URL: <https://hal.science/hal-04212112>.
- [8] M. Leclère, M.-L. Mugnier and G. Pérution-Kihli. ‘Query Rewriting with Disjunctive Existential Rules and Mappings’. In: *Proceedings of the 20th International Conference on Principles of Knowledge Representation and Reasoning*. KR 2023 - 20th International Conference on Principles of Knowledge Representation and Reasoning. Rhodes, Greece: International Joint Conferences on Artificial Intelligence Organization, 2023, pp. 429–439. DOI: [10.24963/kr.2023/42](https://doi.org/10.24963/kr.2023/42). URL: <https://hal-lirmm.ccsd.cnrs.fr/lirmm-04272014>.
- [9] P. Ostropolski-Nalewaja, J. Marcinkowski, D. Carral and S. Rudolph. ‘A Journey to the Frontiers of Query Rewritability’. In: PODS 2022 - 41st ACM SIGMOD-SIGACT-SIGAI Symposium on Principles of Database Systems. Philadelphia, United States, 12th June 2022, pp. 359–367. DOI: [10.1145/3517804.3524163](https://doi.org/10.1145/3517804.3524163). URL: <https://hal.science/hal-03599601>.
- [10] O. Rodriguez, F. Ulliana and M.-L. Mugnier. ‘Scalable Reasoning on Document Stores via Instance-Aware Query Rewriting’. In: *Proceedings of the VLDB Endowment (PVLDB)* 16.11 (24th Aug. 2023), pp. 2699–2713. DOI: [10.14778/3611479.3611481](https://doi.org/10.14778/3611479.3611481). URL: <https://hal-lirmm.ccsd.cnrs.fr/lirmm-04305787>.

- [11] N. Tziavelis, N. Carmeli, W. Gatterbauer, B. Kimelfeld and M. Riedewald. ‘Efficient Computation of Quantiles over Joins’. In: SIGMOD/PODS 2023 - International Conference on Management of Data. Seattle, WA, United States: ACM, 18th June 2023, pp. 303–315. DOI: [10.1145/3584372.3588670](https://doi.org/10.1145/3584372.3588670). URL: <https://hal-lirmm.ccsd.cnrs.fr/lirmm-04277973>.

11.2 Publications of the year

International journals

- [12] J. Frojan, P. Bisquert, P. Buche, N. Gontard, L. Boone, T. Nhu Thuy, A. Vermeulen, P. Ragaert, J. Dewulf and V. Guillard. ‘Scoring methodology for comparing the environmental performance of food packaging’. In: *Packaging Technology and Science* 36.6 (27th Feb. 2023), pp. 439–463. DOI: [10.1002/pts.2720](https://doi.org/10.1002/pts.2720). URL: <https://hal.inrae.fr/hal-04043460>.
- [13] E. Najm, M.-L. Mugnier, C. Gary, J.-F. Baget, R. Métral and L. Garcia. ‘Integrating data and knowledge to support the selection of service plant species in agroecology’. In: *Computers and Electronics in Agriculture* 217 (Feb. 2024), p. 108594. DOI: [10.1016/j.compag.2023.108594](https://doi.org/10.1016/j.compag.2023.108594). URL: <https://hal-lirmm.ccsd.cnrs.fr/lirmm-04393663>.
- [14] O. Rodriguez, F. Ulliana and M.-L. Mugnier. ‘Scalable Reasoning on Document Stores via Instance-Aware Query Rewriting’. In: *Proceedings of the VLDB Endowment (PVLDB)* 16.11 (24th Aug. 2023), pp. 2699–2713. DOI: [10.14778/3611479.3611481](https://doi.org/10.14778/3611479.3611481). URL: <https://hal-lirmm.ccsd.cnrs.fr/lirmm-04305787>.

International peer-reviewed conferences

- [15] J.-F. Baget, M.-L. Mugnier and S. Rudolph. ‘Bounded Treewidth and the Infinite Core Chase: Complications and Workarounds toward Decidable Querying’. In: *Proceedings of the 42nd ACM SIGMOD-SIGACT-SIGAI Symposium on Principles of Database Systems*. SIGMOD/PODS 2023 - International Conference on Management of Data. Seattle, WA, United States: ACM, 18th June 2023, pp. 291–302. DOI: [10.1145/3584372.3588659](https://doi.org/10.1145/3584372.3588659). URL: <https://hal-lirmm.ccsd.cnrs.fr/lirmm-04320944>.
- [16] S.-S. Cîrcu, B. Yun, A. Kheddar, C.-y. Chen and M. Croitoru. ‘Dance, Dance, Dance With My Hands: Third-Party Human Robot-Human Interactions’. In: RO-MAN 2023 - 32nd IEEE International Conference on Robot and Human Interactive Communication. 2023 32nd IEEE International Conference on Robot and Human Interactive Communication (RO-MAN). Busan, South Korea: IEEE, 28th Aug. 2023, pp. 1997–2002. DOI: [10.1109/RO-MAN57019.2023.10309353](https://doi.org/10.1109/RO-MAN57019.2023.10309353). URL: <https://hal.science/hal-04317927>.
- [17] L. Gerlach and D. Carral. ‘Do Repeat Yourself: Understanding Sufficient Conditions for Restricted Chase Non-Termination’. In: KR 2023 - 20th International Conference on Principles of Knowledge Representation and Reasoning. Proceedings of the 20th International Conference on Principles of Knowledge Representation and Reasoning. Rhodes, Greece: International Joint Conferences on Artificial Intelligence Organization, 2nd Sept. 2023, pp. 301–310. DOI: [10.24963/kr.2023/30](https://doi.org/10.24963/kr.2023/30). URL: <https://hal.science/hal-04212137>.
- [18] L. Gerlach and D. Carral. ‘General Acyclicity and Cyclicity Notions for the Disjunctive Skolem Chase’. In: *AAAI-23 Technical Tracks* 5. AAAI 2023 - 37th Conference on Artificial Intelligence. Vol. 37. Technical Track on Knowledge Representation and Reasoning 5. Washington, United States, 27th June 2023, pp. 6372–6379. DOI: [10.1609/aaai.v37i5.25784](https://doi.org/10.1609/aaai.v37i5.25784). URL: <https://hal.science/hal-04212112>.
- [19] M. Leclère, M.-L. Mugnier and G. Pérution-Kihli. ‘Query Rewriting with Disjunctive Existential Rules and Mappings’. In: *Proceedings of the 20th International Conference on Principles of Knowledge Representation and Reasoning*. KR 2023 - 20th International Conference on Principles of Knowledge Representation and Reasoning. Rhodes, Greece: International Joint Conferences on Artificial Intelligence Organization, 2023, pp. 429–439. DOI: [10.24963/kr.2023/42](https://doi.org/10.24963/kr.2023/42). URL: <https://hal-lirmm.ccsd.cnrs.fr/lirmm-04272014>.

- [20] A. Sfar, D. Carral, D. Irofti and M. Croitoru. ‘Testing Logical Diagrams in Power Plants: A Tale of LTL Model Checking’. In: FMICS 2023 - Formal Methods for Industrial Critical Systems. Vol. 14290. Lecture Notes in Computer Science. Antwerp, Netherlands: Springer Nature Switzerland, 17th Sept. 2023, pp. 189–204. DOI: [10.1007/978-3-031-43681-9_11](https://doi.org/10.1007/978-3-031-43681-9_11). URL: <https://hal-lirmm.ccsd.cnrs.fr/lirmm-04214807>.
- [21] S. Shekhar, A. Favier, R. Alami and M. Croitoru. ‘A Knowledge Rich Task Planning Framework for Human-Robot Collaboration’. In: AI 2023 - 43rd SGAI International Conference on Artificial Intelligence. Vol. 14381. Lecture Notes in Computer Science (LNAI). Cambridge, United Kingdom, Dec. 2023, pp. 259–265. DOI: [10.1007/978-3-031-47994-6_25](https://doi.org/10.1007/978-3-031-47994-6_25). URL: <https://laas.hal.science/hal-04385641>.
- [22] N. Tziavelis, N. Carmeli, W. Gatterbauer, B. Kimelfeld and M. Riedewald. ‘Efficient Computation of Quantiles over Joins’. In: SIGMOD/PODS 2023 - International Conference on Management of Data. Seattle, WA, United States: ACM, 18th June 2023, pp. 303–315. DOI: [10.1145/3584372.3588670](https://doi.org/10.1145/3584372.3588670). URL: <https://hal-lirmm.ccsd.cnrs.fr/lirmm-04277973>.

National peer-reviewed Conferences

- [23] J.-E. Baget, P. Bisquert, M. Leclère, M.-L. Mugnier, G. Pérution-Kihli, F. Tornil and F. Ulliana. ‘InteGraal: a Tool for Data-Integration and Reasoning on Heterogeneous and Federated Sources’. In: *Proceedings of the conference BDA’2023*. BDA 2023 - 39e Conférence sur la Gestion de Données – Principes, Technologies et Applications. Montpellier, France, 23rd Oct. 2023. URL: <https://hal-lirmm.ccsd.cnrs.fr/lirmm-04304601>.
- [24] D. Carral, L. Larroque, M. Thomazo and M.-L. Mugnier. ‘Normalisations of Existential Rules: Not so Innocuous!’. In: BDA 2023 - 39e Conférence sur la Gestion de Données – Principes, Technologies et Applications. Montpellier, France, 23rd Oct. 2023. URL: <https://hal-lirmm.ccsd.cnrs.fr/lirmm-04315377>.
- [25] M. Leclère, M.-L. Mugnier and G. Pérution-Kihli. ‘Query Rewriting with Disjunctive Existential Rules and Mappings’. In: BDA 2023 - 39e Conférence sur la Gestion de Données - Principes, Technologies et Applications. Montpellier, France, 23rd Oct. 2023. DOI: [10.48550/arXiv.2306.05973](https://doi.org/10.48550/arXiv.2306.05973). URL: <https://hal-lirmm.ccsd.cnrs.fr/lirmm-04272033>.
- [26] *Best Paper*
O. Rodriguez, F. Ulliana and M.-L. Mugnier. ‘Scalable Reasoning on Document Stores via Instance-Aware Query Rewriting’. In: BDA 2023 - 39e Conférence sur la Gestion de Données – Principes, Technologies et Applications. Montpellier, France, 23rd Oct. 2023. URL: <https://hal-lirmm.ccsd.cnrs.fr/lirmm-04305822>.

Doctoral dissertations and habilitation theses

- [27] G. Pérution-Kihli. ‘Data Management in the Existential Rule Framework: Translation of Queries and Constraints’. Université de Montpellier, 18th Dec. 2023. URL: <https://theses.hal.science/tel-04393631>.

11.3 Cited publications

- [28] N. Carmeli and L. Segoufin. ‘Conjunctive Queries With Self-Joins, Towards a Fine-Grained Complexity Analysis’. In: *PODS’23*. Seattle, United States, June 2023. URL: <https://inria.hal.science/hal-04136055>.
- [29] N. Carmeli, N. Tziavelis, W. Gatterbauer, B. Kimelfeld and M. Riedewald. ‘Tractable Orders for Direct Access to Ranked Answers of Conjunctive Queries’. In: *ACM Transactions on Database Systems* 48.1 (Mar. 2023), pp. 1–45. DOI: [10.1145/3578517](https://doi.org/10.1145/3578517). URL: <https://inria.hal.science/hal-04278100>.

- [30] N. Mitton, L. Brossard, T. Bouadi, F. Garcia, R. Gautron, N. Hilgert, D. Ienco, C. Largouët, E. Lutton, V. Masson, R. Martin-Clouaire, M.-L. Mugnier, P. Neveu, P. Preux, H. Raynal, C. Roussey, A. Termier and V. Bellon Maurel. 'Foundations and state of the art'. In: *Agriculture and Digital Technology: Getting the most out of digital technology to contribute to the transition to sustainable agriculture and food systems*. White book Inria 6. Acknowledgements (contribution, proofreading, editing) – Isabelle Piot-Lepetit. INRIA, 2022, pp. 30–75. URL: <https://hal.inrae.fr/hal-03609470>.