

RESEARCH CENTRE

**Inria Center
at Université Grenoble Alpes**

IN PARTNERSHIP WITH:

CNRS, Université de Grenoble Alpes

2022

ACTIVITY REPORT

Project-Team

TYREX

Types and Reasoning for the Web

IN COLLABORATION WITH: Laboratoire d'Informatique de Grenoble
(LIG)

DOMAIN

Perception, Cognition and Interaction

THEME

Data and Knowledge Representation and
Processing

Inria

Contents

Project-Team TYREX	1
1 Team members, visitors, external collaborators	3
2 Overall objectives	3
2.1 Objectives	3
3 Research program	3
3.1 Foundations for Data Manipulation Analysis: Logics and Type Systems	3
3.2 Algebraic Foundations for Optimization of Information Extraction	4
4 Application domains	4
4.1 Querying Large Graphs	4
4.2 Predictive Analytics for Healthcare	4
5 Social and environmental responsibility	4
5.1 Impact of research results	4
6 Highlights of the year	4
6.1 Awards	4
7 New software and platforms	5
7.1 New software	5
7.1.1 MedAnalytics	5
7.1.2 MuIR	5
8 New results	5
8.1 Algebraic Foundations for Distributed Query Evaluation	5
8.2 Query Plan Enumeration	6
8.3 Data Cleaning and Exchange	6
8.4 Neuro-Symbolic Computing	7
9 Bilateral contracts and grants with industry	7
9.1 Bilateral contracts with industry	7
10 Partnerships and cooperations	7
10.1 National initiatives	7
10.1.1 ANR	7
11 Dissemination	9
11.1 Promoting scientific activities	9
11.1.1 Scientific events: organisation	9
11.1.2 Scientific events: selection	9
11.1.3 Journal	9
11.1.4 Research administration	10
11.2 Teaching - Supervision - Juries	10
11.2.1 Teaching	10
11.2.2 Supervision	11
11.2.3 Juries	11
11.3 Popularization	11
11.3.1 Interventions	11
12 Scientific production	11
12.1 Major publications	11
12.2 Publications of the year	11

Project-Team TYREX

Creation of the Project-Team: 2014 July 01

Keywords

Computer sciences and digital sciences

- A2.1.1. – Semantics of programming languages
- A2.1.4. – Functional programming
- A2.1.7. – Distributed programming
- A2.1.10. – Domain-specific languages
- A2.2.1. – Static analysis
- A2.2.4. – Parallel architectures
- A2.2.8. – Code generation
- A2.4. – Formal method for verification, reliability, certification
- A3.1. – Data
 - A3.1.1. – Modeling, representation
 - A3.1.2. – Data management, quering and storage
 - A3.1.3. – Distributed data
 - A3.1.6. – Query optimization
 - A3.1.9. – Database
 - A3.1.10. – Heterogeneous data
 - A3.1.11. – Structured data
- A3.2.1. – Knowledge bases
- A3.2.2. – Knowledge extraction, cleaning
- A3.2.6. – Linked data
- A3.3.3. – Big data analysis
- A3.4. – Machine learning and statistics
 - A3.4.1. – Supervised learning
- A6.3.3. – Data processing
- A7. – Theory of computation
 - A7.1. – Algorithms
 - A7.2. – Logic in Computer Science
- A9.1. – Knowledge
- A9.2. – Machine learning
- A9.7. – AI algorithmics
- A9.8. – Reasoning
- A9.10. – Hybrid approaches for AI

Other research topics and application domains

B2. – Health

B6.1. – Software industry

B6.5. – Information systems

B9.5.1. – Computer science

B9.5.6. – Data science

B9.7.2. – Open data

1 Team members, visitors, external collaborators

Research Scientists

- Pierre Genevès [Team leader, CNRS, Researcher, HDR]
- Nabil Layaida [INRIA, Senior Researcher, HDR]

Faculty Members

- Ugo Comignani [GRENOBLE INP, Associate Professor]
- Nils Gesbert [GRENOBLE INP, Associate Professor, working 50% until 2022-08-31 then 80%]

PhD Students

- Amela Fejza [UGA, ATER]
- Luisa Werner [UGA]

Technical Staff

- Laurent Carcone [ERCIM]
- Sarah Chlyah [Inria, Engineer]

Administrative Assistant

- Helen Pouchot-Rouge-Blanc [INRIA]

2 Overall objectives

2.1 Objectives

We work on the foundations of the next generation of data analytics and data-centric programming systems. These systems extend ideas from programming languages, artificial intelligence, data management systems, and theory. Data-intensive applications are increasingly more demanding in sophisticated algorithms to represent, store, query, process, analyse and interpret data. We build and study data-centric programming methods and systems at the core of artificial intelligence applications. Challenges include the robust and efficient processing of large amounts of structured, heterogeneous, and distributed data.

On the data-intensive application side, our current focus is on building efficient and scalable analytics systems. Our technical contributions particularly focus on the optimization, compilation, and synthesis of information extraction and analytics code, in particular with large amounts of data.

On the theoretical side, we develop the foundations of data-centric systems and analytics engines with a particular focus on the analysis and typing of data manipulations. We focus in particular on the foundations of programming with distributed data collections. We also study the algebraic and logical foundations of query languages, for their analysis and their evaluation.

3 Research program

3.1 Foundations for Data Manipulation Analysis: Logics and Type Systems

We develop methods for the static analysis of queries and programs that manipulate structured data (such as trees or graphs). One originality of our research is that we develop type-systems based on decision procedures for expressive logics. One major scientific difficulty here consists in dealing with problems

of high computational complexity (sometimes even close to the frontier of decidability), and therefore in finding useful trade-offs between programming expressivity, complexity, succinctness, algorithmic techniques and effective implementations.

3.2 Algebraic Foundations for Optimization of Information Extraction

We explore and develop intermediate languages based on algebraic foundations for the representation, characterization, transformations and compilation of queries. In particular, we investigate two lines of algebraic foundations. First, we study extensions of the relational algebra for optimizing expressive recursive queries. Second, we also explore monad comprehensions and in particular monoid calculi for the generation of efficient and scalable code on big data frameworks. When transforming and optimizing algebraic terms, we rely on cost-based searches of equivalent terms. We thus develop cost models whose purpose is to estimate the time, space and network costs of query evaluation. One difficulty is to estimate these costs in architectures where data and computations are distributed, and where the modeling of data transfers is critical.

4 Application domains

4.1 Querying Large Graphs

Increasingly large amounts of graph-structured data become available. We develop methods that apply to the efficient evaluation of graph queries over large — and potentially distributed — graphs. In particular, we consider the SPARQL query language, which is the standard language for querying graphs structured in the Resource Description Format (RDF). We also consider other increasingly popular graph query languages such as Cypher queries for extracting information from property graphs. We compile graph queries into lower-level distributed primitives found in big data frameworks such as Apache Spark, etc. Applications of graph querying are ubiquitous: large knowledge bases, social networks, road networks, trust networks and fraud detection for cryptocurrencies, publications graphs, web graphs, recommenders, etc.

4.2 Predictive Analytics for Healthcare

One major expectation of data science in healthcare is the ability to leverage on digitized health information and computer systems to better apprehend and improve care. The availability of large amounts of clinical data and in particular electronic health records opens the way to the development of quantitative models for patients that can be used to predict health status, as well as to help prevent disease and adverse effects.

In collaboration with the Grenoble University Hospital (CHUGA), we explore solutions to the problem of predicting important clinical outcomes such as patient mortality, based on clinical data. This raises many challenges including dealing with a very high number of potential predictor variables and resource-consuming data preparation stages.

5 Social and environmental responsibility

5.1 Impact of research results

Our work on query optimization helps in reducing resource consumption in information extraction.

6 Highlights of the year

6.1 Awards

Luisa Werner received the best PhD presentation award of the MIAI days for her works entitled “Neural symbolic integration of knowledge extraction and reasoning on graph data”.

7 New software and platforms

7.1 New software

7.1.1 MedAnalytics

Keywords: Big data, Predictive analytics, Distributed systems

Functional Description: We implemented a method for the automatic detection of at-risk profiles based on a fine-grained analysis of prescription data at the time of admission. The system relies on an optimized distributed architecture adapted for processing very large volumes of medical records and clinical data. We conducted practical experiments with real data of millions of patients and hundreds of hospitals. We demonstrated how the various perspectives of big data improve the detection of at-risk patients, making it possible to construct predictive models that benefit from volume and variety. Parts of this prototype implementation are described in the publications DSAA'18, Big Data'18, CHIL'21, UAI'21.

Publications: [hal-01517087](#), [hal-01877742](#), [hal-03124966](#), [hal-03125018](#), [hal-03160473](#), [hal-03066941](#), [hal-03266004](#)

Contact: Pierre Genevès

Partner: CHU Grenoble

7.1.2 MuIR

Name: Mu Intermediate Representation

Keywords: Optimizing compiler, Querying

Functional Description: This is a prototype of an intermediate language representation, i.e. an implementation of algebraic terms, rewrite rules, query plans, cost model, query optimizer, and query evaluators. This includes a distributed evaluator of algebraic terms using Apache Spark. Concepts of this implementation have been described in the SIGMOD'20 and CIKM'20 publications, among others, and the distributed evaluator and query optimizers are described in 2021 preprints.

Publications: [hal-01673025](#), [hal-03295445](#), [hal-03004218](#), [hal-03517826](#)

Contact: Pierre Genevès

8 New results

8.1 Algebraic Foundations for Distributed Query Evaluation

Participants: Pierre Genevès, Nabil Layaida, Nils Gesbert, Sarah Chlyah.

Distributed Evaluation of Graph Queries using Recursive Relational Algebra. We have investigated the distributed evaluation of μ -RA queries. We present a system called Dist- μ -RA for the distributed evaluation of recursive graph queries. Dist- μ -RA builds on the recursive relational algebra and extends it with evaluation plans suited for the distributed setting. The goal is to offer expressivity for high-level queries while providing efficiency at scale and reducing communication costs. Experimental results on both real and synthetic graphs show the effectiveness of the proposed approach compared to existing systems [4].

An Algebra with a Fixpoint Operator for Distributed Data Collections. Big data programming frameworks are becoming increasingly important for the development of applications, for which performance and scalability are critical. In those complex frameworks, optimizing code by hand is hard and time-consuming, making automated optimization particularly necessary. In order to automate optimization, a prerequisite is to find suitable abstractions to represent programs; for instance, algebras based on monads or monoids to represent distributed data collections. Currently, however, such algebras do not represent recursive programs in a way which allows analyzing or rewriting them. We extend a monoid algebra with a fixpoint operator for representing recursion as a first class citizen and show how it allows new optimizations. The fixpoint operator is suitable for modeling recursive computations with distributed data collections. We show that under reasonable conditions this fixpoint can be evaluated by parallel loops with one final merge rather than by a global loop requiring network overhead after each iteration. We also propose several rewrite rules, showing when and how filters can be pushed through recursive terms, and how to filter inside a fixpoint before a join. Experiments with the Spark platform illustrate performance gains brought by these systematic optimizations [5, 3].

8.2 Query Plan Enumeration

Participants: Amela Fejza, Pierre Genevès, Nabil Layaïda.

Efficient Enumeration of Recursive Plans in Transformation-based Query Optimizers. Query optimizers built on the transformation-based Volcano/Cascades framework are used in many database systems. Transformations proposed earlier on the logical query dag (LQDAG) data structure, which is key in such a framework, focus only on recursion-free queries. We propose the recursive logical query dag (RLQDAG) which extends the LQDAG with the ability to capture and transform recursive queries, leveraging recent developments in recursive relational algebra. Specifically, this extension includes: (i) the ability of capturing and transforming sets of recursive relational terms thanks to (ii) annotated equivalence nodes used for guiding transformations that are more complex in the presence of recursion; and (iii) RLQDAG rewrite rules that transform sets of subterms in a grouped manner, instead of transforming individual terms in a sequential manner; and that (iv) incrementally update the necessary annotations. Core concepts of the RLQDAG are formalized using a syntax and formal semantics with a particular focus on subterm sharing and recursion. The result is a clean generalization of the LQDAG transformation-based approach, enabling more efficient explorations of plan spaces for recursive queries. An implementation of the proposed approach shows significant performance gains compared to the state-of-the-art [6].

Exploring Property Graphs with Recursive Path Patterns. We demonstrate a system for recursive query answering over property graphs. The novelty of the system resides in its ability to optimize and efficiently answer recursive path patterns in queries for property graphs. The system is based on a complete implementation of the μ -recursive relational algebra [1]. It also includes parsers and compilers adapted for property graphs so that one can formulate, optimize and answer queries that navigate recursively along paths in property graphs. We demonstrate the system on three real datasets, including the exploration of chains of drug interactions [7].

8.3 Data Cleaning and Exchange

Participants: Ugo Comignani.

Provenance-aware Discovery of Functional Dependencies on Integrated Views. The automatic discovery of functional dependencies (FDs) has been widely studied as one of the hardest problems in data

profiling. Existing approaches have focused on making the FD computation efficient while inspecting single relations at a time. In this paper, for the first time we address the problem of inferring FDs for multiple relations as they occur in integrated views by solely using the functional dependencies of the base relations of the view itself. To this purpose, we leverage logical inference and selective mining and show that we can discover most of the exact FDs from the base relations and avoid the full computation of the FDs for the integrated view itself, while at the same time preserving the lineage of FDs of base relations. We propose algorithms to speedup the inferred FD discovery process and mine FDs on-the-fly only from necessary data partitions. We present InFine(INferred FunctIoNal dEpendency), an end-to-end solution to discover inferred FDs on integrated views by leveraging provenance information of base relations. Our experiments on a range of real-world and synthetic datasets demonstrate the benefits of our method over existing FD discovery methods that need to rerun the discovery process on the view from scratch and cannot exploit lineage information on the FDs. We show that InFine outperforms traditional methods necessitating the full integrated view computation by one to two order of magnitude in terms of runtime. It is also the most memory efficient method while preserving FD provenance information using mainly inference from base table with negligible execution time.

These results were presented at the ICDE 2022 conference [2].

8.4 Neuro-Symbolic Computing

Participants: Luisa Werner, Nabil Layaïda, Pierre Genevès.

On the Replicability of Knowledge Enhanced Neural Networks in a Graph Neural Network Framework.

In order to extend Knowledge Enhanced Neural Networks, we investigate the replicability of the approach and present a re-implementation of Knowledge Enhanced Neural Networks based on a Graph Neural Network framework (PyTorch Geometric). Knowledge Enhanced Neural Networks integrate prior knowledge in the form of logical formulas into an Artificial Neural Network by adding additional Knowledge Enhancement layers. The obtained results show that the model outperforms pure neural models as well as Neural-Symbolic models. Our long term goal is to be able to address more complex and large-scale knowledge graphs and to benefit from the wide range of functionalities available in PyTorch Geometric. To ensure that our implementation produces the same results, we replicate the original transductive experiments and explain the various challenges and the steps that we went through to reach that goal [8].

9 Bilateral contracts and grants with industry

9.1 Bilateral contracts with industry

Participants: Pierre Genevès, Nabil Layaïda.

Collaboration with the French Opensee fintech startup located in Paris about query optimization for multidimensional data.

10 Partnerships and cooperations

10.1 National initiatives

10.1.1 ANR

CLEAR

Participants: Pierre Genevès, Nabil Layaïda, Nils Gesbert, Sarah Chlyah, Amela Fejza.

- Title: Compilation of intermediate Languages into Efficient big dAta Runtimes
- Call: Appel à projets générique 2016 défi 'Société de l'information et de la communication' – JCJC
- Duration: January 2017 – Mars 2022
- Coordinator: Pierre Genevès
- See also: tyrex.inria.fr/clear
- Abstract: This project addresses one fundamental challenge of our time: the construction of effective programming models and compilation techniques for the correct and efficient exploitation of big and linked data. We study high-level specifications of pipelines of data transformations and extraction for producing valuable knowledge from rich and heterogeneous data. We investigate how to synthesize code which is correct and optimized for execution on distributed infrastructures.

Newcare

Participants: Pierre Genevès, Nabil Layaïda, Luisa Werner.

- Title: Network for hEalth Workers : Covid And oRganization of Emergency teams – NEWCARE
- Duration: January 2021 – Mars 2024
- Coordinator: Marie-Estelle BINET (Laboratoire d'Economie Appliquée de Grenoble)
- Abstract: This research project has several objectives. The first is to create an original database to describe the characteristics and interactions between caregivers working in healthcare teams in the emergency department. These data will be extracted (or desilated) from the PREDIMED clinical data warehouse (CDW), which gathers health and administrative data from patients and healthcare professionals working at Grenoble University Hospital. Then, the analysis of social networks will allow us to identify the modes of collaboration in place between caregivers and their ability to adapt to their environment. Impact evaluation methods will allow us to estimate the impact of the organizational changes caused by the covid-19 health crisis on the quality of work and the well-being of healthcare professionals.

QualiHealth

Participants: Ugo Comignani.

- Title: Enhancing the Quality of Health Data
- Call: Appel à projets Projets de Recherche Collaborative – Entreprise (PRCE)
- Duration: 2018-2022
- Coordinator: Angela Bonifati
- Others partners: LIMOS, Université Clermont Auvergne. LIS, Université d'Aix-Marseille. HEGP, INSERM, Paris. Inst. Cochin, INSERM, Paris. Gnubila, Argonay. The University of British Columbia, Vancouver (Canada)

- **Abstract:** This research project is geared towards a system capable of capturing and formalizing the knowledge of data quality from domain experts, enriching the available data with this knowledge and thus exploiting this knowledge in the subsequent quality-aware medical research studies. We expect a quality-certified collection of medical and biological datasets, on which quality-certified analytical queries can be formulated. We envision the conception and implementation of a quality-aware query engine with query enrichment and answering capabilities.

To reach this ambitious objective, the following concrete scientific goals must be fulfilled: (1) An innovative research approach, that starts from concrete datasets and expert practices and knowledge to reach formal models and theoretical solutions, will be employed to elicit innovative quality dimensions and to identify, formalize, verify and finally construct quality indicators able to capture the variety and complexity of medical data; those indicators have to be composed, normalized and aggregated when queries involve data with different granularities and of different quality dimensions (e.g., mixing incomplete and inaccurate data); and (2) In turn, those complex aggregated indicators have to be used to provide new quality-driven query answering, refinement, enrichment and data analytics techniques. A key novelty of this project is the handling of data which are not rectified on the original database but sanitized in a query-driven fashion: queries will be modified, rewritten and extended to integrate quality parameters in a flexible and automatic way.

Participation to MIAI Chairs

Participants: Pierre Genevès, Nabil Layaïda, Amela Fejza, Luisa Werner.

P. Genevès is member of the board of the DeepCare MIAI Chair. A. Fejza participates to the DeepCare MIAI Chair. N. Layaïda, L. Werner and P. Genevès also participate to the Knowledge communication and evolution MIAI Chair.

11 Dissemination

Participants: Pierre Genevès, Nabil Layaïda, Nils Gesbert, Sarah Chlyah, Ugo Comignani, Amela Fejza, Luisa Werner.

11.1 Promoting scientific activities

11.1.1 Scientific events: organisation

- P. Genevès organized the MFML Axis Workshop 2022, a yearly workshop on Formal Methods Model and Languages organized of the LIG laboratory.

11.1.2 Scientific events: selection

Member of the conference program committees

- P. Genevès is program committee member for the PLDI 2023 and SIGMOD 2023 conferences.
- U. Comignani is a PhD Thesis awards committee member of the BDA 2022 conference

11.1.3 Journal

Reviewer - reviewing activities

- U. Comignani is reviewer for the following journals: IEEE TKDE, the VLDB journal, the Journal of Data and Information Quality

11.1.4 Research administration

- N. Layaïda is president of the Hiring committee CRCN-ISFP of the Centre Inria de l'Université de Rennes 2023. The Inria Rennes hiring commission is in charge of hiring permanent researchers CRCN-ISFP of the Research Center.
- N. Layaïda was a member of Professor hiring committee at Université de Lyon 1. 2022. Hiring committee member for a Professor position for the LIRIS Laboratory in the Database Group.
- N. Layaïda is a member of the scientific committee of the LabEx PERSYVAL-lab (Pervasive Systems and Algorithms).
- N. Layaïda is a member of the Scientific Board of Digital League, the digital cluster of Auvergne-Rhône-Alpes.
- P. Genevès is member of the board of the DeepCare MIAI Chair
- P. Genevès is co-responsible of the Doctoral School MSTII that organizes the PhD program in Computer Science at the University Grenoble Alpes.
- P. Genevès is member of the board of the LIG laboratory, responsible for the "Formal Methods, Models and Languages" research axis regrouping 4 research teams working in this field.

11.2 Teaching - Supervision - Juries

11.2.1 Teaching

- Master: P. Genevès is co-responsible and teacher of the M2-level course 'Fundamentals of Data Processing and Distributed Knowledge' of the MOSIG program at UGA (36h)
- Master: P. Genevès is co-responsible and teacher of the M2-level course 'Accès à l'information: du web des données au web sémantique' in the ENSIMAG ISI 3A program at Grenoble-INP (30h)
- Master : N. Gesbert, academic tutorship of an apprentice, 6 h eq TD, M1, Grenoble INP
- Master : N. Gesbert, 'Construction d'applications Web', 27 h eq TD, M1, Grenoble INP
- Master : N. Gesbert, 'Principes des systèmes de gestion des bases de données', 58 h eq TD, M1, Grenoble INP
- Master : N. Gesbert, 'Introduction to lambda-calculus', 4 h eq TD, M2, UGA-Grenoble INP (MOSIG)
- Licence : N. Gesbert, 'Logique pour l'informatique', 45 h eq TD, L3, Grenoble INP
- N. Gesbert is in charge of the L3-level course 'logique pour l'informatique' and of the M1-level course 'Principes des systèmes de gestion de bases de données (SEOC)'.
- Master : U. Comignani is co-responsible of the "BigData" master, co-accredited between Grenoble Ecole de Management and Grenoble INP
- Master : U. Comignani is in charge of the 'Projets fil rouges', 10 h eq TD, MS BigData, Grenoble INP
- Master : U. Comignani, 'Principes des systèmes de gestion de bases de données', 99.5 h eq TD, M1, Grenoble INP
- Master : U. Comignani is in charge of the 'Projet BD', 64 h eq TD, M1, Grenoble INP
- Master : U. Comignani, 'Stockage et traitement de données à grande échelle', 28.5 h eq TD, M2, Grenoble INP
- Master : U. Comignani, academic tutorship of an apprentice, 10 h eq TD, M1, Grenoble INP

11.2.2 Supervision

- PhD in progress: Luisa Werner, Neural Symbolic Integration, PhD started in October 2020, co-supervised by Nabil Layaïda and Pierre Genevès.
- PhD thesis of Amela Fejza, On the Optimization of Recursive Plan Enumeration with an Application to Property Graph Queries, supervised by Pierre Genevès, defended on January 11, 2023.
- PhD thesis of Sarah Chlyah, On Algebraic Foundations for the Optimization of Iterative Programming with Distributed Data Collections, co-supervised by Pierre Genevès and Nabil Layaïda, defended on May 5, 2022.

11.2.3 Juries

- N. Layaïda has been reviewer (Rapporteur) for the thesis of Issam GHABRI. Efficacité énergétique des phases de conception et d'exploitation des entrepôts de données. PhD in computer science. Ecole Nationale Supérieure de Mécanique et d'Aérotechnique. 29 December 2022.
- N. Layaïda has been reviewer (Rapporteur) for the thesis of Théo Ducros, Reasoning in Description Logics Augmented with Refreshing Variables. PhD in computer science from Université Clermont Auvergne. 27 September 2022.

11.3 Popularization

11.3.1 Interventions

- Luisa Werner gave a presentation on “Neural symbolic integration of knowledge extraction and reasoning on graph data” at the MIAI days, for which she received the best PhD presentation award.
- Sarah Chlyah gave a presentation on “Neurosymbolic integration” for the Atelier d'Innovation at Centre Inria de l'Université de Grenoble Alpes.
- Sarah Chlyah gave a presentation on “Distributed evaluation of recursive algebra” for the service d'expérimentation et développement of the Centre Inria de l'Université de Grenoble Alpes.

12 Scientific production

12.1 Major publications

- [1] L. Jachiet, P. Genevès, N. Gesbert and N. Layaïda. ‘On the Optimization of Recursive Relational Queries: Application to Graph Queries’. In: *SIGMOD 2020 - ACM International Conference on Management of Data*. Portland, United States, June 2020, pp. 1–23. DOI: [10.1145/3318464.3380567](https://doi.org/10.1145/3318464.3380567). URL: <https://hal.inria.fr/hal-01673025>.

12.2 Publications of the year

International peer-reviewed conferences

- [2] U. Comignani, L. Berti-Équille, N. Novelli and A. Bonifati. ‘Provenance-aware Discovery of Functional Dependencies on Integrated Views’. In: *ICDE 2022 - 38th IEEE International Conference on Data Engineering*. Kuala Lumpur, Malaysia: IEEE, 9th May 2022, pp. 1–15. URL: <https://hal.archives-ouvertes.fr/hal-03474955>.

Doctoral dissertations and habilitation theses

- [3] S. Chlyah. ‘On Algebraic Foundations for the Optimization of Iterative Programming with Distributed Data Collections’. Université Grenoble Alpes [2020-....], 23rd May 2022. URL: <https://theses.hal.science/tel-03783672>.

Reports & preprints

- [4] S. Chlyah, P. Genevès and N. Layaïda. *Distributed Evaluation of Graph Queries using Recursive Relational Algebra*. 7th Oct. 2022. URL: <https://hal.inria.fr/hal-03295445>.
- [5] S. Chlyah, N. Gesbert, P. Genevès and N. Layaïda. *On the Optimization of Iterative Programming with Distributed Data Collections*. 24th May 2022. URL: <https://hal.inria.fr/hal-02066649>.
- [6] A. Fejza, P. Genevès and N. Layaïda. *Efficient Enumeration of Recursive Plans in Transformation-based Query Optimizers*. 24th Oct. 2022. URL: <https://hal.inria.fr/hal-03692274>.
- [7] A. Fejza, P. Genevès and N. Layaïda. *Exploring Labeled Graphs with Recursive Path Patterns*. 24th Oct. 2022. URL: <https://hal.inria.fr/hal-03517826>.
- [8] L. Werner, N. Layaïda and P. Genevès. *On the Replicability of Knowledge Enhanced Neural Networks in a Graph Neural Network Framework*. 9th June 2022. URL: <https://hal.inria.fr/hal-03692362>.