Zoé Chevallier

Education

2026 – **PostDoctoral Researcher**
Supervisor: Matteo Lissandrini
University of Verona, Italy

2021 – 2025 **PhD Candidate**
Title: Data Discovery and Extraction in the Web of Data
Supervisor: Zoubida Kedad, Béatrice Finance et Frédéric Chaillan
DAVID Laboratory, University of Versailles Paris Saclay, France
Grand Paris Sud, Lieusaint, France

2016 – 2019 **Computer Engineering Diploma**
ISTY, *Institut des Sciences et Techniques des Yvelines*, Velizy, France
2018 **Exchange semester at the University of Quebec at Chicoutimi** (4 months)

2014 – 2016 **CPGE – Classe Préparatoire aux Grandes Écoles**
Intensive preparatory program in Mathematics, Physics and Engineering Science
Lycée Marceau, Chartres, France

2014 **Scientific Baccalaureate, major in Biology, with elective in Computer Science**
French national secondary school diploma
Lycée Marceau, Chartres, France

Description of Research Activities

During my PhD

Title *Data Discovery and Extraction in the Web of Data*
Supervisor Zoubida Kedad, Béatrice Finance, Frédéric Chaillan

My PhD research, in collaboration with the Grand Paris Sud local authority, aimed to support public decision-making by extracting relevant information from the Web of Data to specific application needs. The work focused on the identification and extraction of data to populate a target schema describing these needs. During this PhD, we had three contributions: (i) identification of contributive data sources based on semi-supervised learning, (ii) a scalable approach to identify contributive data sources, and (iii) the generation of extraction queries to populate the target schema.
We first focused on identifying contributive sources to populate a target schema, in a context of irregular data. We suppose that the alignment between the target schema and the data source is not always provided. But in this context, data sources may still contain candidate instances

that are relevant to populate the target schema. We propose an approach for extracting candidate instance patterns, i.e. property sets characterizing the candidate instances, from a data source [5][2]. This approach, based on a semi-supervised learning algorithm, iteratively computes the similarity between source entities and target schema classes, taking into account the candidate instances already identified for these classes.
We also tackled the scalability issues for the identification of contributive data sources. Our approach based on a semi-supervised learning algorithm is not well suited to massive data sources due to its time complexity. We introduce an indexing-based approach that significantly reduces the execution time while producing results of comparable quality to those obtained with our semi-supervised learning based approach. This approach indexes the data source, as well as fictive entities representing classes of the target schema to identify sets of similar entities that are then used to identify the candidate instances.
Finally, we addressed the problem of query generation in order to populate a target schema with valid data with respect to a set of constraints defined on this schema using the SHACL language. To this end, we propose an approach for the automatic generation of extraction queries, expressed in SPARQL, that minimizes the number of generated queries [3][1]. A lattice structure is used to group the pattern of candidate instances that can be extracted in a same query. These queries include all the required filters to extract only the valid entities with respect to the constraints defined on the target schema.

During my Internship

Title *Towards a Digital Twin for Smart Buildings*

Supervisor Béatrice Finance

I completed my final-year internship as part of the development of a digital twin for a smart building at CESI Engineering School. The goal of the project was to enable simulations of the building's evolution. We proposed a data architecture designed to describe a smart building, integrating both static data such as the building's physical description and dynamic data coming from sensors [4]. These data are initially described using domain-specific tools (Building Information Modeling) and are then transformed into RDF using the ifcOWL ontology vocabulary. The architecture also includes data streams from the building's sensors, described using the SOSA/SSN ontologies.

Teaching

2025 – 2026 **Data Structures and Algorithms**

Level: Second year of bachelor's degree in computer science, 2x36h tutorials

Activity: As part of this course, I supervised tutorial sessions. The course covers the concepts of complexity, as well as different data structures (arrays, stacks/queues, linked lists, and different trees), and the principles of algorithms (recursion, sorting algorithms, tree processing algorithms).

2024-2025 **Introduction to programming**

Level: First year of Bachelor's degree in Computer Science, 19h Practical Sessions

Activity: supervision of practical sessions, project mentoring

Python Language

2024-2025 **Programming, Software engineering**

Level: First year of Master's degree in Computer Science, 2x24h Practical Sessions

Activity: supervision of practical sessions, project mentoring

Java Language, Object-Oriented Programming

2024-2025 **Programming bridging course**

Level: First year of Master's degree in Computer Science, 21h Practical Sessions

Activity: supervision of practical sessions, project mentoring

Java Language, Object-Oriented Programming

2024-2025 **Introduction to Databases**

Level: Second year of Bachelor's degree (two groups not specialized in Computer Science), 2x36h tutorials

Activity: supervision of tutorial sessions

Modeling, creation, and querying of a relational database, entity-relationship modeling, relational algebra, and SQL language

2022-2023 **Introduction to Databases**

Level: Second year of Bachelor's degree (specialized in Mathematics and Informatics applied on Human and Social Science), 30h tutorials

Activity: supervision of tutorial sessions, preparation of exam questions

Modeling, creation, and querying of a relational database, entity-relationship modeling, relational algebra, and SQL language

2017 - 2021 **Private math tutoring**

Level: From 9th grade to 12th grade

Activity: Designing customized lessons based on students' needs, helping them understand exercises, and providing support for exam preparation

Service

2022 – 2024 **PhD students' representative**, DAVID laboratory

Activity: Participated in general assemblies and communicated information to the lab PhD students, served as liaison between the laboratory's head and PhD students, and organized a student seminar within the laboratory.

2024 **Junior Committee Member**, Junior Conference on Data Science and Engineering (JDSE)

Activity: Conference organization, drafted and distributed the call for papers, managed event communications, invited speakers, and acted as a reviewer for scientific papers

Publications

International Conferences

[1] Z. Chevallier, Z. Kedad, B. Finance, and F. Chaillan. 2024. Generating SPARQL Queries for Data Discovery. In Advances in Databases and Information Systems (ADBIS 2024), J. Tekli, J. Gamper, R. Chbeir, and Y. Manolopoulos (Eds.). *Lecture Notes in Computer Science*, Vol. 14918. Springer, Cham. https://doi.org/10.1007/978-3-031-70626-4_5

[2] Z. Chevallier, Z. Kedad, B. Finance, and F. Chaillan. 2024. Identifying Relevant Data in RDF Sources. In *Proceedings of the 18th International Conference on Research Challenges in Information Science (RCIS 2024), Part II*, J. Araújo, J. L. de la Vara, M. Y. Santos, and S. Assar (Eds.). *Lecture Notes in Business Information Processing*, Vol. 514. Springer, 92–99. https://doi.org/10.1007/978-3-031-59468-7_11

[3] Z. Chevallier, Z. Kedad, B. Finance, and F. Chaillan. 2024. Data Search and Discovery in RDF Sources. In *The Semantic Web: ESWC 2024 Satellite Events, Part I*, A. Meroño-Peñuela, Ó. Corcho, P. Groth, E. Simperl, V. Tamma, A. G. Nuzzolese, M. Poveda-Villalón, M. Sabou, V. Presutti, I. Celino, A. Revenko, J. Raad, B. Sartini, and P. Lisena (Eds.). *Lecture Notes in Computer Science*, Vol. 15344. Springer, 188–192. https://doi.org/10.1007/978-3-031-78952-6_24

International Workshop

[4] Z. Chevallier, B. Finance, and B. Cohen Boulakia. 2020. A Reference Architecture for Smart Building Digital Twin. In *Proceedings of the International Workshop on Semantic Digital Twins (SeDiT@ESWC 2020)*, R. García-Castro, J. Davies, G. Antoniou, and C. Fortuna (Eds.). *CEUR Workshop Proceedings*, Vol. 2615. CEUR-WS.org. https://ceur-ws.org/Vol-2615/paper2.pdf

National Conference

[5] Z. Chevallier, Z. Kedad, B. Finance, and F. Chaillan. 2023. Identification de données pertinentes dans des sources RDF. In *Actes de la Conférence sur la Gestion de Données – Principes, Technologies et Applications (BDA 2023)*.

Posters

2023 **Research Day of Paris Saclay University**. Data integration for decision support in Smart Cities

2023 **GDR MaDICS - CNRS**. Data integration for decision support in Smart Cities

2022 **Research Day of Paris Saclay University**. Data integration for decision support in Smart Cities

Talks

2024 **Invited Speaker.** Identifying Relevant Data in RDF Sources.
K-GALS Workshop@ADBIS

2023 **Paper presentation, BDA national conference** (39th Conference on Data Management: Principles, Technologies, and Applications)

2022 **PhD Day**, DAVID Laboratory

2022 **TIGA project seminar** (Territory of Innovation with Great Ambition)
TIGA project a call for expressions of interest launched as part of the third Investment for the Future Program (PIA 3), aimed at transforming and supporting innovative and ambitious territorial projects.

Referees

Zoubida Kedad (*PhD Supervisor*)
Title : Professor, University of Versailles Paris Saclay
Mail : zoubida.kedad@uvsq.fr

Béatrice Finance (*PhD Co-Supervisor*)
Title : Assistant Professor (HDR), University of Versailles Paris Saclay
Mail : beatrice.finance@uvsq.fr

Nicolas Travers
Title : Professor, ESILV Engineering School
Mail : nicolas.travers@devinci.fr