

Luca Gagliardelli

CURRICULUM VITAE

LAST UPDATE: SEPTEMBER 2022

PERSONAL INFORMATION

Address Via Pietro Vivarelli 10, 41125 Modena, Italy
Email luca.gagliardelli@unimore.it
Personal page <http://personale.unimore.it/rubrica/dettaglio/luca.gagliardelli>
Phone (+39) 3403809971

Summary				
Start	End	Institute	Position	Research topic
01/01/2022	Present	University of Modena and Reggio Emilia	Fixed-term Researcher (RTDA) DIEF presso DBGroup	Development of architectures and software for Big Data integration and analysis.
16/07/2021	31/12/2021	University of Modena and Reggio Emilia	Postdoc (senior research fellow) DIEF presso DBGroup	Development of architectures and software for Big Data integration and analysis.
01/12/2020	30/06/2021	University of Modena and Reggio Emilia	Postdoc (research fellow) DIEF presso DBGroup	Development of architectures and software for Big Data integration and analysis.
01/12/2019	30/11/2020	University of Modena and Reggio Emilia	Postdoc (research fellow) DSV	LigAdvisor: a webserver for drug design based on Big Data.
01/11/2016	31/10/2019	University of Modena and Reggio Emilia	PhD student DIEF at DBGroup	Big Data Integration
16/05/2015	31/10/2016	University of Modena and Reggio Emilia	Research fellow DIEF at DBGroup	Extension of the MOMIS data integration system.

01/2022-Present
RTDA at
UNIMORE

Fixed-term Researcher (RTD-A) at the DBGroup of the University of Modena and Reggio Emilia.
Research activity: big data integration, data cleaning.

07/2021-12/2021
Research fellow at
UNIMORE

Research fellow at the DBGroup of the University of Modena and Reggio Emilia.
Research activity: modelling of new data integration algorithm in Big Data context. In particular, the research was focused on deploying scalable Entity Resolution techniques on parallel architectures, such as Apache Spark.
Advisor: Prof.ssa Sonia BERGAMASCHI · University of Modena and Reggio Emilia

12/2020-07/2021
Research fellow at
UNIMORE

Research fellow at the DBGroup of the University of Modena and Reggio Emilia.
Research activity: deployment of software architectures for Big Data integration and analysis withing the project: "SUPER: Supercomputing Unified Platform - EmiliaRomagna", funded by the Emilia Romagna region.
Advisor: Prof.ssa Sonia BERGAMASCHI · University of Modena and Reggio Emilia

12/2019-12/2020
Research fellow at
UNIMORE

Research fellow at the DBGroup of the University of Modena and Reggio Emilia.
Research activity: development of a platform that helps researchers to perform drug re-purposing within the project "LigAdvisor: a webserver for drug design based on Big Data" funded by the project FAR 2019.
Advisor: Prof. Domenico BENEVENTANO · University of Modena and Reggio Emilia

11/2016–11/2019
PhD Student at
UNIMORE

PhD Student at the DBGroup of the University of Modena and Reggio Emilia.

Research activity: development of distributed data integration techniques to scale high volume of data. In particular, advanced blocking/indexing techniques and similarity joins on Big Data.

Advisor: Prof.ssa Sonia BERGAMASCHI · Università degli Studi di Modena e Reggio Emilia

Co-Advisor: PhD Giovanni SIMONINI · University of Modena and Reggio Emilia.

05/2015–11/2016
Research fellow at
UNIMORE

Research fellow at the DBGroup dell'University of Modena and Reggio Emilia.

Research activity: semantic data integration for the research project "Open Linked Data dell'Osservatorio Giovani della Regione Emilia Romagna" and extension of the MOMIS data integration system to create interactive dashboards for visualizing the integrated data, funded by the municipality of Modena.

Advisor: Prof.ssa Sonia BERGAMASCHI · Università degli Studi di Modena e Reggio Emilia

EDUCATION

03/2020
PhD student at
UNIMORE

UNIVERSITY OF MODENA AND REGGIO EMILIA — ICT International Doctoral School

Thesis: *Techniques for Big Data Integration in Distributed Computing Environments.*

Advisor: Prof.ssa Sonia BERGAMASCHI · University of Modena and Reggio Emilia

Co-Advisor: PhD. Giovanni SIMONINI · University of Modena and Reggio Emilia

04/2015
Master Degree in
Computer
Engineering

UNIVERSITY OF MODENA AND REGGIO EMILIA

Grade: 110/110

10/2011
Bachelor Degree in
Computer
Engineering

UNIVERSITY OF MODENA AND REGGIO EMILIA

Grade: 93/110

TEACHING ACTIVITY

2023
UNIMORE

Assistant professor · 60 hours of frontal lessons for the course "Basi di dati", at the Engineering department "Enzo Ferrari", BSc in Computer Engineering - Mantua.

Assistant professor · 40 hours of frontal lessons for the course "Business intelligence", at the Economics department "Marco Biagi", MSc in International Management.

2022
UNIMORE

Assistant professor · 36 hours of frontal lessons for the course "Basi di dati", at the Engineering department "Enzo Ferrari", BSc in Computer Engineering - Mantua.

2021
UNIMORE

Adjunct professor · 32 hours of frontal lessons for the course "Basi di dati", at the Engineering department "Enzo Ferrari", BSc in Computer Engineering - Mantua.

Lecturer · Lecturer for the course "Data Analytics" for the students of the International Doctorate school in Information and Communication Technologies of the University of Modena and Reggio Emilia – [\[LINK\]](#).

2020
UNIMORE

Adjunct professor · 32 hours of frontal lessons for the course "Basi di dati", at the Engineering department "Enzo Ferrari", BSc in Computer Engineering - Mantua.

2019
UNIMORE

Adjunct professor · 27 hours of frontal lessons for the course "Sistemi informativi e basi di dati L-Z", at the "DISMI" department, BSc in Management Engineering - Reggio Emilia.

Adjunct professor · 4 hours of frontal lessons for the course "Corso di perfezionamento Metodologie, Tecniche e Tool per Analisi di Big Data".

PARTICIPATION IN RESEARCH PROJECTS

10/2020–01/2022

Digital Experience Platform

The project is developed in collaboration with Doxee S.p.a. and is funded by the Emilia Romagna region. The goal of the project is to develop a platform to manage a high volume of data coming from several types of bills (e.g. light, water, phone) and perform data analysis to get useful insights to improve the user experience.

10/2020–04/2021

Smart monitoring of Local Energy Community

The project is funded by MISE and supervised by ENEA. The goal of the project is to design a platform to collect and analyze the data acquired by the sensors that monitor the energy network to allow a conscious use of energy by the users of the so-called Energy Communities.

COLLABORATIONS

01/03/2021–
01/08/2021
Paris Descartes
University

Internship carried out under the supervision of the professor Themis Palpanas at the Descartes University, Paris. During this period I worked on a **Big Data Cleaning** project that has as outcome an article that was published at the international conference VLDB 2022.

Note: the internship was carried out in remote work due to the Covid-19 restrictions.

STUDENT SUPERVISOR

MSc in Computer
Engineering

(Co-advisor) Marco Esposito. "Gestione e analisi di Big Data conformi al GDPR: il progetto del database Digital eXperience Platform"
2022 UNIMORE.

BSc in Computer
Engineering

(Co-advisor) Davide Abba. "Full-stack project "Gastronomia La Cilba" "
2021 UNIMORE.

(Co-advisor) Mario Cristiano. "Relational Database Management System: Analisi dell'ambiente MariaDB"
2021 UNIMORE.

(Co-advisor) Davide Corradi. "Full-Text Search e applicazioni: Confronto tra Microsoft SQL Server ed Elasticsearch"
2021 UNIMORE.

(Co-advisor) Francesca Prontera. "Scalabilità di Tecniche di Analisi di Dati con la libreria Modin"
2021 UNIMORE.

(Co-advisor) Luisa Bottiglieri. "TimescaleDB: analisi di time-series di consumi energetici"
2021 UNIMORE.

(Co-advisor) Giovanni Bonisoli. "Sviluppo di un'applicazione web per l'utilizzo del framework SparkER"
2017 UNIMORE.

AWARDS AND GRANTS

- Awards* 2017 · Young Professional event @ RTSI2017 · (second prize).
- Grants* 2019 · Christine Collet EDBT/ICDT Student Participation Award · (Free registration and launches for the EDBT international conference)
- 2016 · three years PhD scholarship funded by the Emilia Romagna region.

TALKS

- September 2022 · VLDB, Sidney, Australia (remote due to Covid-19): “Generalized Supervised Meta-blocking”
- March 2022 · DataPlat, Edinburgh, UK (remote due to Covid-19): “DXP: Billing Data Preparation for Big Data Analytics”
- June 2020 · SEBD, Villasimius, Italy (remote due to Covid-19): “Scaling Up Record-level Matching Rules”
- March 2020 · EDBT, Copenhagen, Denmark (remote due to Covid-19): “RulER: Scaling Up Record-level Matching Rules”
- March 2019 · EDBT, Lisboa, Portugal: “Sparker: Scaling entity resolution in spark”
- July 2018 · TE2018, Modena, Italy: “BigDedup: a big data integration toolkit for duplicate detection in industrial scenarios”

PROFESSIONAL SERVICES

- PC member* · 24th International Conference on Information Integration and Web Intelligence (iiWAS 2022).
- 24th International Conference on Big Data Analytics and Knowledge Discovery (DaWaK 2022).
- 23rd International Conference on Big Data Analytics and Knowledge Discovery (DaWaK 2021).
- Revisor* · IEEE Transactions on Knowledge and Data Engineering (TKDE).
- Knowledge-Based Systems (KNOSIS).
- ACM SIGMOD Reproducibility 2023.
- External revisor* · ICDE 2022.
- SIGMOD 2022.
- SIGMOD 2021.
- VLDB 2020.
- EDBT 2020.
- ICDE 2018.

WORK EXPERIENCE

- 2015–Now
Collaborator at DataRiver Consultant for data integration projects.
<http://datariver.it/chi-siamo/team/>
- 2/2019– 3/2019
Full stack developer Development of an interactive map for the municipality of Modena, title of the project “Mappa multimediale di Modena, le architetture del Novecento”. [[Sito web](#)]

COMPUTER SKILLS

<i>Programming Languages</i>	SQL; PYTHON; SCALA; JAVA; JAVASCRIPT; SHELL
<i>Big Data Frameworks</i>	SPARK (SPARK, SPARKSQL, MLlib); DELTALAKE
<i>Other Frameworks</i>	DOCKER; IPYTHON; ANGULAR; SCIKIT-LEARN; SCRAPY; MONGODB; SQL SERVER; MYSQL; POSTGRESQL; L ^A T _E X;

OTHER INFORMATION

<i>Languages</i>	ITALIAN · Mother tongue
	ENGLISH · Professional knowledge

PUBLICATIONS

Metrics (June 2022)		
	Google Scholar	Scopus
h-index	7	7
Citations	153	107

Publications list organized by type and sorted by publication date

[c] = conference [j] = journal [bc] = book chapter [t] = thesis [sw] = software

<i>Journal</i>	<p>[j8] The Italian National Registry for FSHD: an enhanced data integration and an analytics framework towards Smart Health Care and Precision Medicine for a rare disease - Orphanet Journal of Rare Diseases, 2021 - C. BETTIO, V. SALSÌ, M. ORSINI, E. CALANCI, L. MAGNOTTA, L. GAGLIARDELLI, J. KINOSHITA, S. BERGAMASCHI, R. TUPLER. [PDF]</p> <p>[j7] Reproducible experiments on Three-Dimensional Entity Resolution with JedAI - Information Systems, 2021 - G. PAPADAKIS, G. MANDILARAS, L. GAGLIARDELLI, G. SIMONINI, E. THANOS, G. GIANNAKOPOULOS, S. BERGAMASCHI, T. PALPANAS, M. KOUBARAKIS. [PDF]</p> <p>[j6] LigAdvisor: a versatile and user-friendly web-platform for drug design - Nucleic Acids Research (NAR), 2021 - L. PINZI, A. TINIVELLA, L. GAGLIARDELLI, D. BENEVENTANO, G. RASTELLI. [PDF]</p> <p>[j5] BLAST2 An Efficient Technique for Loose Schema Information Extraction from Heterogeneous Big Data Sources - Journal of Data and Information Quality (JDIQ), 2020 - D. BENEVENTANO, S. BERGAMASCHI, L. GAGLIARDELLI, G. SIMONINI. [PDF]</p> <p>[j4] Three-dimensional Entity Resolution with JedAI - Information Systems, 2020 - G. PAPADAKIS, G. MANDILARAS, L. GAGLIARDELLI, G. SIMONINI, E. THANOS, G. GIANNAKOPOULOS, S. BERGAMASCHI, T. PALPANAS, M. KOUBARAKIS. [PDF]</p> <p>[j3] Scaling entity resolution: A loosely schema-aware approach - Information Systems, 2019 - G. SIMONINI, L. GAGLIARDELLI, S. BERGAMASCHI, HV Jagadish. [PDF] [CODE]</p> <p>[j2] BigBench workload executed by using Apache Flink - Procedia Manufacturing 11 (2017): 695-702. - S. BERGAMASCHI, L. GAGLIARDELLI, G. SIMONINI, S. ZHU. [PDF]</p> <p>[j1] Driving Innovation in Youth Policies With OpenData - IC₃K 2015. Communications in Computer and Information Science, 2016 - D. BENEVENTANO, S. BERGAMASCHI, L. GAGLIARDELLI, L. PO. [PDF]</p>
----------------	--

Conference

- [c14] **Progressive Entity Resolution with Node Embeddings**, SEBD 2022. - G. SIMONINI, L. GAGLIARDELLI, M. RINALDI, L. ZECCHINI, G. DE SABBATA, A. ASLAM, D. BENEVENTANO, S. BERGAMASCHI. [PDF]
- [c13] **Generalized Supervised Meta-blocking** - 48th International Conference on Very Large Databases, VLDB 2022. - L. GAGLIARDELLI, G. PAPADAKIS, G. SIMONINI, S. BERGAMASCHI, T. PALPANAS. [PDF]
- [c12] **ECDP: A Big Data Platform for the Smart Monitoring of Local Energy Communities** - 1st International Workshop on Data Platform Design, Management, and Optimization, DataPlat 2022. - L. GAGLIARDELLI, L. ZECCHINI, D. BENEVENTANO, G. SIMONINI, S. BERGAMASCHI, M. ORSINI, L. MAGNOTTA, E. MESCOI, A. LIVALDI, N. GESSA, P. DE SABBATA, G. D'AGOSTA, F. PAOLUCCI, F. MORETTI. [PDF]
- [c11] **The Case for Multi-task Active Learning Entity Resolution** - 29th Italian Symposium on Advanced Database Systems, SEBD 2021. - G. SIMONINI, H. SACCANI, L. GAGLIARDELLI, L. ZECCHINI, D. BENEVENTANO, S. BERGAMASCHI. [PDF]
- [c10] **Scaling Up Record-level Matching Rules** - 28th Italian Symposium on Advanced Database Systems, SEBD 2020. - L. GAGLIARDELLI, G. SIMONINI, S. BERGAMASCHI. [PDF]
- [c9] **RuLER: Scaling Up Record-level Matching Rules** - 23rd International Conference on Extending Database Technology, EDBT 2020. - L. GAGLIARDELLI, G. SIMONINI, S. BERGAMASCHI. [PDF]
- [c8] **Entity resolution and data fusion: An integrated approach** - 27th Italian Symposium on Advanced Database Systems, SEBD 2019. - D. BENEVENTANO, S. BERGAMASCHI, L. GAGLIARDELLI, G. SIMONINI. [PDF]
- [c7] **Sparker: Scaling entity resolution in Spark** - 22nd International Conference on Extending Database Technology, EDBT 2019. - L. GAGLIARDELLI, G. SIMONINI, D. BENEVENTANO, S. BERGAMASCHI. [PDF] [CODE]
- [c6] **How improve Set Similarity Join based on prefix approach in distributed environment.** - IEEE International Conf. on High Performance Computing & Simulation, HPCS 2018. - S. ZHU, L. GAGLIARDELLI, D. BENEVENTANO, G. SIMONINI. [PDF]
- [c5] **Enhancing Loosely Schema-aware Entity Resolution with User Interaction.** - IEEE International Conf. on High Performance Computing & Simulation, HPCS 2018. - G. SIMONINI, L. GAGLIARDELLI, S. ZHU, S. BERGAMASCHI. [PDF]
- [c4] **Bigdedup: a big data integration toolkit for duplicate detection in industrial scenarios** - IEEE 25th International Conference on Transdisciplinary Engineering, TE 2018. - L. GAGLIARDELLI, S. ZHU, G. SIMONINI, S. BERGAMASCHI. [PDF]
- [c3] **Momis dashboard: A powerful data analytics tool for industry 4.0** - IEEE 25th International Conference on Transdisciplinary Engineering, TE 2018. - L. MAGNOTTA, L. GAGLIARDELLI, G. SIMONINI, M. ORSINI, S. BERGAMASCHI. [PDF]
- [c2] **The Italian FSHD Registry: an enhanced data integration and analytics framework for smart health care** - IEEE 3rd International Forum on Research and Technologies for Society and Industry, RTSI 2017. - M. ORSINI, E. CALANCHI, L. MAGNOTTA, L. GAGLIARDELLI, M. GOVI, F. MELE, R. TUPLER. [PDF]
- [c1] **Open Data for Improving Youth Policies** - International Joint Conference on Knowledge Discovery, Knowledge Engineering, IC3K 2015 - D. BENEVENTANO, S. BERGAMASCHI, L. GAGLIARDELLI, L. Po. [PDF]

Book chapter

[bc1] **From Data Integration to Big Data Integration.** - “A Comprehensive Guide Through the Italian Database Research 2018”, Springer. - S. BERGAMASCHI, D. BENEVENTANO, F. MANDREOLI, R. MARTOGLIA, F. GUERRA, M. ORSINI, L. PO, M. VINCINI, G. SIMONINI, S. ZHU, L. GAGLIARDELLI, L. MAGNOTTA. [[PDF](#)]

Thesis

[t] **Techniques for Big Data Integration in Distributed Computing Environments** - PhD Thesis (2020) - L. Gagliardelli. [[PDF](#)]

Software

[sw] **JedAI** - è un framework completo che contiene diversi tool per eseguire integrazione dati. Personalmente mi sono occupato di sviluppare la parte distribuita basata su Apache Spark. [[LINK](#)]

[sw] **RuIER** - è un tool che utilizza una nuova tecnica per trovare record simili applicando regole complesse di join di similarità su uno o più attributi realizzato utilizzando il framework Apache Spark. [[LINK](#)]

[sw] **SparkER** - è un tool che implementa tecniche avanzate di blocking per eseguire Entity Resolution realizzato utilizzando il framework Apache Spark. [[LINK](#)]