

Luca Gagliardelli

CURRICULUM VITAE

LAST UPDATE: 17/07/2024

PERSONAL INFORMATION

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

RESEARCH EXPERIENCE

Luca Gagliardelli is a Fixed-term Researcher (RTDA - PNRR) at the Database Group (DB-Group) of the Department of Engineering “Enzo Ferrari” of the University of Modena and Reggio Emilia located in Modena, Italy. His research interests lie in the area of big data integration and management. With his research he advanced the field of Entity Resolution, developing scalable meta-blocking techniques that can be used to integrate big data in highly heterogeneous environments, available in an open source framework named [SparkER](#).

After pursuing a Master’s Degree in Computer Engineering in April 2015, he held a research fellow position from May 2015 till October 2016 at the University of Modena and Reggio Emilia. During this period he worked on the project “Open linked data Osservatorio Giovani della Regione Emilia-Romagna” funded by the municipality of Modena and Emilia Romagna region ([LINK](#)), integrating open and proprietary data coming from heterogeneous data sources using and extending the MOMIS data integration system [[c1](#), [j1](#), [c3](#)]. In 2020 he obtained a PhD in Computer Science from the Doctorate school in Information and Communication Technologies of the University of Modena and Reggio Emilia, defending a thesis entitled “Techniques for Big Data Integration in Distributed Computing Environments”. His work was focused on scaling Entity Resolution techniques to integrate Big Data in distributed computing environments [[j3](#), [j5](#), [c7](#), [c8](#), [c9](#), [c10](#)]. During this period, he also worked on side projects, such as the development of the Italian FSHD Registry carried out in collaboration with DataRiver srl and the Department of Biomedical, Metabolic and Neural Sciences of the University of Modena and Reggio Emilia [[c2](#), [j8](#)].

After the PhD, he was a postdoc covering several research fellow positions at the University of Modena and Reggio Emilia, working on multidisciplinary research projects involving the data integration problem:

- “LigAdvisor: a webserver for drug design based on big data” ([LINK](#)), funded by FAR 2019 and carried out with the Department of Life Sciences, involves the design and development of a web platform for drug design in the pharmaceutical field [[j6](#)].
- “Smart monitoring of Local Energy Community” funded by MISE and carried out in collaboration with ENEA, involves the design of a platform for the management of data generated by the Local Energy Communities to reduce the electric consumption [[j9](#)] and promote good use of energy.
- “Digital Experience Platform” funded by the Emilia Romagna Region and carried out in collaboration with Doxee spa, involves the design of a platform for the management of data coming from light/gas billing ([LINK](#)).

- “BD4M: Big Data 4 Manufacturing”, WP leader for the “WP3 - Design and Implementation of a Big Data analysis platform”. Industrial research project carried out in collaboration with DataRiver srl and funded by Bi.Rex, that developed a Big Data platform for Industry 4.0, that can collect and analyze highly heterogeneous data collected from different kinds of industrial machines.

During the postdoc he developed international collaborations with professor Themis Palpanas of the Paris Descartes University and with George Papadakis of the University of Athens, publishing several works on Entity Resolution [j10, j4, j7, c13, c16]. The main work [c13, j10] proposes a novel meta-blocking techniques that was published and presented at VLDB 2022.

From January 2022 till October 2023 he worked as Fixed-term researcher (RTDA, DM 1062) on the project “Big Data e Intelligenza Artificiale per il Virtuosismo Energetico nelle Comunità Energetiche Locali” to develop scalable data integration techniques for the management of energy data coming from local energy communities [c12, j9]. During this period, he received a research grant of 9200 € from the University of Modena and Reggio Emilia (FAR) to carry out research activities on Big Data integration [j10] and was the scientific coordinator of the research/consulting contract “Design and development of microservices for Record Linkage and Entity Matching for the MOMIS data integration platform” between DataRiver srl and the Department of Engineering “Enzo Ferrari” of the University of Modena, worth 5000 €.

Since November 2023 he has been a Fixed-term researcher (RTDA, PNRR) working within the Work Package 5 “Digital Maktaba” of the PNRR project “ITSERR - Italian Strengthening of the ESFRI RI RESILIENCE”, which has the goal of building a library that integrates documents written in non-latin languages (e.g. Arabic) exploiting, also, the use of Large Language Models (LLMs). Since one of the problems in using LLMs is related to the usage costs, in current and future work he plans to develop Entity Resolution techniques that leverage LLMs in a pay-as-you-go fashion to obtain the best results within a fixed budget.

Summary			
Start	End	Institute	Position
01/11/2023	Present	University of Modena and Reggio Emilia	Fixed-term Researcher (RTDA - PNRR) DIEF at DBGroup
01/01/2022	31/10/2023	University of Modena and Reggio Emilia	Fixed-term Researcher (RTDA - DM 1062) DIEF at DBGroup
16/07/2021	31/12/2021	University of Modena and Reggio Emilia	Postdoc (senior research fellow) DIEF at DBGroup
01/12/2020	30/06/2021	University of Modena and Reggio Emilia	Postdoc (research fellow) DIEF at DBGroup
01/12/2019	30/11/2020	University of Modena and Reggio Emilia	Postdoc (research fellow) DSV
01/11/2016	31/10/2019	University of Modena and Reggio Emilia	PhD student DIEF at DBGroup
16/05/2015	31/10/2016	University of Modena and Reggio Emilia	Research fellow DIEF at DBGroup

- 11/2023-Present
RTDA at
UNIMORE
- Fixed-term Researcher (RTDA - PNRR) at the DBGroup of the University of Modena and Reggio Emilia.**
Research activity: the research activity is carried out within the Work Package 5 “Digital Maktaba” of the PNRR project “ITSERR - Italian Strengthening of the ESFRI RI RESILIENCE” (Avviso MUR 3264/2022) funded by EU – NextGenerationEU - Grant No IR0000014. The goal is to develop techniques for acquiring, indexing, categorizing, and integrating documents written in non-latin languages (mainly in Arabic), to the end of building a digital library that integrates data coming from different single libraries. In the big data integration field, this project faces non-trivial problems such as multilanguage data integration (the same document could be translated into multiple languages), authorship identification in Arabic documents, topic identification to categorize documents, and more. To solve these problems, we are working on techniques to exploit the power of recently available LLM engines such as ChatGPT 4.
- 01/2022-10/2023
RTDA at
UNIMORE
- Fixed-term Researcher (RTDA - DM 1092) at the DBGroup of the University of Modena and Reggio Emilia.**
Research activity: development of scalable techniques and algorithms for the management and analysis of big data, with a focus on data produced by the Local Energy Communities. The activity was conducted within the project “Big Data e Intelligenza Artificiale per il Virtuosismo Energetico nelle Comunità Energetiche Locali”, funded by the program “Fondo Sociale Europeo REACT-EU - PON “Ricerca e Innovazione” 2014 – 2020 – Azione IV.6 Contratti di ricerca su tematiche Green (D.M. 1062 del 10/08/ 2021) CUP: E95F21002330001”.
- 07/2021-12/2021
Research fellow at
UNIMORE
- Research fellow at the DBGroup of the University of Modena and Reggio Emilia.**
Research activity: modeling of new data integration algorithms in Big Data context. In particular, the research 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 within the project: “SUPER: Supercomputing Unified Platform - Emilia Romagna”, 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 web server 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 volumes of data. In particular, advanced blocking/indexing techniques and similarity joins on Big Data.
Advisor: Prof.ssa Sonia BERGAMASCHI · University of Modena and 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 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

QUALIFICATIONS

17/12/2023 –
17/12/2034

NATIONAL SCIENTIFIC QUALIFICATION AS ASSOCIATE PROFESSOR IN THE ITALIAN HIGHER EDUCATION SYSTEM

Disciplinary field: 09/H1 – Information processing systems

Released by: MUR – Italian Ministry for Universities and Research

2015

STATE EXAM FOR SECTION A AND COMPUTER SCIENCE SECTOR

Released by: *University of Modena and Reggio Emilia*

TEACHING ACTIVITY

He started teaching during the PhD period as a subject expert for the course "Basi di dati" (Database Foundation), at the Engineering department "Enzo Ferrari", BSc in Computer Engineering, doing exercises on the Entity Relationship model and SQL. Subsequently, he covered the same role for the course Big Data Management and Governance, at the Engineering department "Enzo Ferrari", MSc in Computer Engineering, doing theoretical and practical lessons on advanced SQL (stored procedures, triggers, transactions) and Apache Spark. In the academic year 2018/2019 he was an adjunct professor for the course "Sistemi informativi e basi di dati L-Z", teaching the SQL language through practical lessons to undergraduate students in Management Engineering. In the same year, he also held a seminary on Apache Spark (theory and practical use) within the course "Corso di perfezionamento Metodologie, Tecniche e Tool per Analisi di Big Data" [LINK]. In 2020 he was a lecturer for the course "Data Analytics" for the PhD students of the ICT Doctorate School of the University of Modena and

Reggio Emilia to present the enabling Big Data technologies such as MapReduce paradigm, HDFS, and the practical use of Apache Spark.

Since the academic year 2022/2023 he has been the teacher holder for the courses "Basi di dati" (Database Foundation), at the Engineering Department "Enzo Ferrari", BSc in Computer Engineering - Mantua, and "Business Intelligence" at the Economics department "Marco Biagi", MSc in International Management - Modena. The first course is taught in Italian and provides the knowledge to design, implement, and use databases through theoretical and practical lessons. It starts from the designing of a database from requisites written in natural language with the use of the Entity-Relationship model, then the logical design with the relational model, and in the end the implementation and querying through the SQL language. The latter is taught in English, since most of the students are foreign/Erasmus, and has a more practical focus on the methodologies and technologies used in a typical Business Intelligence pipeline. The students learn how to manage, clean, and integrate data, build effective dashboards, and do some basic data mining operations.

During his teaching activities, he mentored several BSc students, and an MSc student as outlined subsequently.

Academic year
2023/2024
UNIMORE

Assistant professor · 72 hours of frontal lessons for the course "Basi di dati" (Database Foundation), at the Engineering department "Enzo Ferrari", BSc in Computer Engineering - Mantua.
Course Language: Italian. **Enrolled students:** 40.

Assistant professor · 38 hours of frontal lessons for the course "Business intelligence", at the Economics department "Marco Biagi", MSc in International Management.
Course Language: English. **Enrolled students:** 80.

Academic year
2022/2023
UNIMORE

Assistant professor · 60 hours of frontal lessons for the course "Basi di dati" (Database Foundation), at the Engineering department "Enzo Ferrari", BSc in Computer Engineering - Mantua.
Course Language: Italian. **Enrolled students:** 40.

Assistant professor · 38 hours of frontal lessons for the course "Business intelligence", at the Economics department "Marco Biagi", MSc in International Management.
Course Language: English. **Enrolled students:** 70.

Academic year
2021/2022
UNIMORE

Assistant professor · 36 hours of frontal lessons for the course "Basi di dati" (Database Foundation), at the Engineering department "Enzo Ferrari", BSc in Computer Engineering - Mantua.
Course Language: Italian. **Enrolled students:** 38.

Academic year
2020/2021
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.
Course Language: Italian. **Enrolled students:** 50.

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 – [\[LINK\]](#).

Course Language: English. **Enrolled PhD students:** 10.

Academic year
2019/2020
UNIMORE

Adjunct professor · 32 hours of frontal lessons for the course "Basi di dati" (Database Foundation), at the Engineering department "Enzo Ferrari", BSc in Computer Engineering - Mantua.
Course Language: Italian. **Enrolled students:** 61.

Academic year
2018/2019
UNIMORE

Adjunct professor · 27 hours of frontal lessons for the course "Sistemi informativi e basi di dati L-Z" (Database Foundation), at the "DISMI" department, BSc in Management Engineering - Reggio Emilia.
Course Language: Italian. **Enrolled students:** 80.

Adjunct professor · 4 hours of frontal lessons on Big Data analysis with Apache Spark for the course "Corso di perfezionamento Metodologie, Tecniche e Tool per Analisi di Big Data".
Course Language: Italian. **Enrolled students:** 10.

STUDENT SUPERVISOR

PhD Students

(Co-tutor) Giovanni Sullutrone. "Exploring the Potential of Large Language Models for Multilingual Historical Document Analysis and Semantic Cataloguing in Digital Libraries"
XXXIX Cycle [\[LINK\]](#).

(Co-tutor) Luca Sala. "Data Management, analytics and intelligent AI-based knowledge extraction for multilingual and multi-alphabetic heritages"
XXXVIII Cycle [\[LINK\]](#).

(Co-tutor) Sania Aftar. "Data Management, analytics and intelligent AI-based knowledge extraction for multilingual and multi-alphabetic heritages"
XXXVIII Cycle [\[LINK\]](#).

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

(Advisor) Manuel Olmi. "Full-stack project "Gestione Esami""
2023 UNIMORE.

(Co-advisor) Michele Rinaldi. "Entity Resolution progressiva con graph embedding"
2022 UNIMORE.

(Co-advisor) Arturo Bianchi. "Documentazione e aggiunte alla libreria per l'Entity Resolution SparkER"
2022 UNIMORE.

(Co-advisor) Davide Bucciarelli. "SIARD Format: analisi di un nuovo modo di preservare database"
2021 UNIMORE.

(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.

PARTICIPATION IN RESEARCH PROJECTS

4/2023–11/2023

Design and development of microservices for Record Linkage and Entity Matching for the MOMIS data integration platform

Scientific coordinator of the research/consulting contract worth 5000 € between DataRiver srl and the Department of Engineering “Enzo Ferrari” of the University of Modena and Reggio Emilia for the design and development of the microservices for Record Linkage and Entity Matching for the MOMIS data integration platform.

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.

7/2020–01/2022

BD4M: Big Data 4 Manufacturing

Work Package leader for the “WP3 - Design and Implementation of a Big Data analysis platform”. The industrial research project was carried out in collaboration with DataRiver srl and funded by Bi.Rex “Area 2 – Big Data per il Manufacturing” di cui in attuazione dell’art. 1, comma 115, Legge 11 dicembre 2016 n. 232 e in conformità e nel rispetto dell’art. 4, D.Lgs. 19 agosto 2016 n. 175, del D.M. Sviluppo Economico 12 settembre 2017 n. 214, del Decreto Direttoriale del Ministero dello Sviluppo Economico 29 gennaio 2018.

The goal of the project is to develop a Big Data platform for Industry 4.0, that can collect and analyze highly heterogeneous data collected from different kinds of industrial machines.

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 [c13].

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

AWARDS AND GRANTS

Awards

2017 · Young Professional event @ RTSI2017 · (second prize).

Grants

2022 · Research grant of 9200 € from the Engineering Department of the University of Modena and Reggio Emilia (FAR) for my research project about data cleaning and integration.

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

- June 2023 · SEBD, Galzignano Terme, Italy: “How To Generalize Supervised Meta-blocking”
- September 2022 · VLDB, Sidney, Australia (remote due to Covid-19): “Generalized Supervised Meta-blocking” - [\[LINK\]](#)
- March 2022 · DataPlat, Edinburgh, UK (remote due to Covid-19): “ECDP: A Big Data Platform for the Smart Monitoring of Local Energy Communities” - [\[LINK\]](#)
- 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*
 - 25th International Conference on Big Data Analytics and Knowledge Discovery (DaWaK 2023).
 - 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*
 - Elsevier Information Systems.
 - Data Mining & Knowledge Discovery.
 - Data & Knowledge Engineering.
 - IEEE Transactions on Knowledge and Data Engineering (TKDE).
 - Knowledge-Based Systems (KNOSIS).
 - ACM SIGMOD Reproducibility 2023.
- External revisor*
 - SIGMOD 2024.
 - 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

- Programming*
Languages
 - SQL; PYTHON; SCALA; JAVA; JAVASCRIPT; SHELL

Big Data Frameworks

SPARK (SPARK, SPARKSQL, MLlib); DELTALAKE

*Other Frameworks*DOCKER; IPYTHON; ANGULAR; SCIKIT-LEARN; SCRAPY; MONGODB; SQL SERVER; MYSQL; POSTGRESQL; L^AT_EX; TABLEAU;**OTHER INFORMATION***Languages*ITALIAN · Mother tongue
ENGLISH · Professional knowledge**PUBLICATIONS**

Metrics (July 2024)		
	Google Scholar	Scopus
h-index	10	8
Citations	318	217

Publications list organized by type and sorted by publication date

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

Journal

- [j11] **Stream-aware indexing for distributed inequality join processing** - Information Systems, 2024. - A. ASLAM, G. SIMONINI, L. GAGLIARDELLI, L. ZECCHINI, S. BERGAMASCHI [PDF]
- [j10] **GSM: A generalized approach to Supervised Meta-blocking for scalable entity resolution** - Information Systems, 2024. - L. GAGLIARDELLI, G. PAPADAKIS, G. SIMONINI, S. BERGAMASCHI, T. PALPANAS. [PDF]
- [j9] **A Big Data Platform Exploiting Auditable Tokenization to Promote Good Practices inside Local Energy Communities** - Future Generation Computer Systems, 2023. - L. GAGLIARDELLI, L. ZECCHINI, FERRETTI, D. BENEVENTANO, G. SIMONINI, S. BERGAMASCHI, M. ORSINI, L. MAGNOTTA, E. MESCOLO, A. LIVALDI, N. GESSA, P. DE SABBATA, G. D'AGOSTA, F. PAOLUCCI, F. MORETTI. [PDF]
- [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. SALSU, M. ORSINI, E. CALANCI, L. MAGNOTTA, L. GAGLIARDELLI, J. KINOSHITA, S. BERGAMASCHI, R. TUPLER. [PDF]
- [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]
- [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]
- [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]
- [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]

- [j3] **Scaling entity resolution: A loosely schema-aware approach** - Information Systems, 2019 - G. SIMONINI, L. GAGLIARDELLI, S. BERGAMASCHI, HV Jagadish. [PDF] [CODE]
- [j2] **BigBench workload executed by using Apache Flink** - Procedia Manufacturing 11 (2017): 695-702. - S. BERGAMASCHI, L. GAGLIARDELLI, G. SIMONINI, S. ZHU. [PDF]
- [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]

Conference

- [c19] **A Novel Methodology for Topic Identification in Hadith**, IRCDL 2024. - S. AFTAR, L. GAGLIARDELLI, A. EL GANADI, F. RUOZZI, S. BERGAMASCHI.
- [c18] **A Big Data Platform for the Management of Local Energy Communities Data**, AIPREF @ IEEE BigData 2023. - S. BERGAMASCHI, L. GAGLIARDELLI.
- [c17] **HKS: Efficient Data Partitioning for Stateful Streaming**, DAWAK 2023. - A. ASLAM, G. SIMONINI, L. GAGLIARDELLI, A. MOZZILLO, S. BERGAMASCHI.
- [c16] **A general approach to Supervised Meta-Blocking**, SEBD 2023. - L. GAGLIARDELLI, G. PAPADAKIS, G. SIMONINI, S. BERGAMASCHI, T. PALPANAS.
- [c15] **Big Data Integration for Data-Centric AI**, ItaData 2022. - S. BERGAMASCHI, D. BENEVENTANO, G. SIMONINI, L. GAGLIARDELLI, A. ASLAM, G. DE SABBATA, L. ZECCHINI.
- [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. MESCOLO, 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](#)]

Book

- [b1] **Progetto di Basi di Dati Relazionali**, Bonomo Editore. D. BENEVENTANO, S. BERGAMASCHI, L. GAGLIARDELLI, F. GUERRA, M. VINCINI. ISBN 9788869723032 [[LINK](#)]

Thesis

- [t1] **Techniques for Big Data Integration in Distributed Computing Environments** - PhD Thesis (2020) - L. Gagliardelli. [[PDF](#)]

Software

- [sw1] **JedAI** - is a complete framework that contains several tools for performing data integration. Personally I contributed by developing the distributed part based on Apache Spark. [[LINK](#)]
- [sw2] **RuIER** - is a tool that uses a new technique to find similar records by applying complex similarity join rules on one or more attributes, developed by using the Apache Spark framework. [[LINK](#)]
- [sw3] **SparkER** - is a tool that implements advanced blocking techniques to perform Entity Resolution, developed by using the Apache Spark framework. [[LINK](#)]