

Giovanni Simonini

Associate Professor at the University of Modena and Reggio Emilia — Italy

CURRICULUM VITAE

UPDATED ON MARCH 2023

1 GENERAL INFORMATION

Full Name Giovanni Simonini
Citizenship Italian
email simonini@unimore.it
Website www.giovannisimonini.com

2 RESEARCH ACTIVITIES

2.1 Research Appointments

Research Appointments Summary			
Start	End	Institution	Position
01/11/2022	now	Università degli Studi di Modena e Reggio Emilia	Associate Professor
01/10/2019	30/10/2022	Università degli Studi di Modena e Reggio Emilia	Assistant Professor
01/10/2018	01/10/2019	Massachusetts Institute of Technology	Postdoctoral Associate CSAIL Database group
01/01/2016	26/09/2018	Università degli Studi di Modena e Reggio Emilia	Postdoc (assegnista di ricerca) DIEF Database group
01/01/2013	31/12/2015	Università degli Studi di Modena e Reggio Emilia	PhD Student DIEF Database group

11/22-now
Assoc. Prof.
UNIMORE
Associate Professor of Big Data Management; main research topics: big data integration, data preparation, and data discovery.

10/19-10/22
RTDb
UNIMORE
Assistant Professor (RTD b)
Main research interests: large-scale data curation; Big Data management and analysis for the environment.

10/18-10/19
Postdoc at
MIT
Postdoctoral Associate at the Massachusetts Institute of Technology, Computer Science and Artificial Intelligence Laboratory. Research topic: Study of new algorithms for automating the data cleaning and data preparation processes within machine learning pipelines.
PI: Prof. Michael STONEBRAKER · MIT CSAIL

01/16-09/18
Postdoc at
UNIMORE
Research topic: Modelling of novel approaches for data integration techniques in the context of Big Data. In particular, working on devising new methods for scaling Entity Resolution (i.e., to identify and fuse different representation of the same real-world entity in databases) on data-intensive parallel computing systems, such as Apache Spark.

PI: Prof. Sonia BERGAMASCHI · UNIMORE DIEF

01/13–12/2015
PhD Student at
UNIMORE

PhD student of the DBGroup of the Università degli Studi di Modena e Reggio Emilia. Research activity: Modelling of new techniques for Big Data Integration. In particular, in the context of large and heterogeneous data: (i) modelling and developing a new technique for blocking (i.e., indexing technique for scaling Entity Resolution); (ii) modelling a novel approach for data discovery/exploration. The resulting thesis won the "2016 PhD Thesis Award" assigned by IEEE Computer Society (Italy section).

Advisor: Prof. Sonia BERGAMASCHI · UNIMORE DIEF

Co-advisor: Prof. H.V. JAGADISH · University of Michigan, Ann Arbor, US

2.2 Visiting & Collaborations

2018
QCRI

Visiting scholar at Qatar Computing Research Institute (QCRI), Doha – Qatar.

10/17
HPI

Visiting researcher at the Hasso Plattner Institute, Potsdam, for collaboration with Prof. Felix Naumann at on a Big Data Exploration & Cleaning project.

03/17– 04/17
Paris Descartes
University

Visiting researcher at the Descartes University, Paris, for collaboration with Prof. Themis Palpanas on a Big Data Cleaning project.

05/15– 11/15
U. of Michigan

Visiting researcher at the University of Michigan, Ann Arbor, for collaboration with Prof. H.V. Jagadish on a Big Data Integration project.

2.3 Grants and Awards

Grants

2023-2025 · National project PRIN 2022 "Discount quality for responsible data science: Human-in-the-Loop for quality data" cod. 202248FWFS – Role: **Local Unit PI**.

2023-2025 · Emilia Romagna project PR-FESR 2022 DATHA PG/2023/311868 – Role: **PI**.

2021 · Grant for "Definition of Machine Learning techniques to incentivize physical activity by means of wearable and smart devices" – private company, Role: **PI**.

2018 · Grant for the project "Automatic Application-driven Data Cleaning" from the "Fondo di Ateneo per la Ricerca 2018" (FAR JUNIOR) – The project has been selected by experts – according to the REPRISE database of MIUR.

2016 · VLDB 2016 Travel Fellowship · (travel and lodging grant, and free VLDB conference registration)

2015 · Grant from CINECA - Project: Entity Resolution for Big Data

2013 · Three year Grant from MIUR for PhD program
2012 · One year Spinner 2013 Scholarship

Awards 2021 · Honorable mention for Artificial intelligence and Big Data from Gruppo 2003
<https://www.scienzainrete.it/premio>

2017 · **IEEE Computer Society Italy Section Chapter PhD Thesis Award**
<http://sites.ieee.org/italy-cs/ieee-computer-society-italy-section-chapter-2016-phd-thesis-award/>

2.4 Professional Services

Editor Associate editor for the ACM Journal of Data and Information Quality (JDIQ), 2024—now

Organizer ·Proceeding chair VLDB 2025.
·Co-Chair of two editions of the ACM SIGMOD Programming Contest (2021 and 2022)
·Program Chair of BDAA 2018 (as part of IEEE HPCS 2018)
·Technical Session Chair “Big Data Integration and IoT for Smart Health Care”, IEEE RTSI 2017

Reviewer ·IEEE Transactions on Knowledge and Data Engineering
·Neurocomputing
·Data & Knowledge Engineering
·ACM Journal of Data and Information Quality

PC Member ·ICDE 2022, 2023(Best reviewer)
·SIGMOD 2021, 2023, 2025
·VLDB 2023, 2025
·EDBT 2022, 2024
·DOLAP 2024
·DOLAP 2024
·TheWebConf (Industry track) 2022, 2024
·International Conference on Database Systems for Advanced Applications 2019 (DEMO PC)
·International Conference on Model and Data Engineering
·BDAA (as part of HPCS) 2017 & 2015 & 2014
·ICDE 2018 (external reviewer)
·WI 2017 (external reviewer)
·VLDB 2014 (external reviewer)

*Research
Evaluation
Committee* ·Evaluation of the Research and Professional Activities of the Institutes of the Czech Academy of
Sciences for 2010-2014

2.5 Relevant Talks

01/2019 · MIT and QCRI meeting – Doha, Qatar: “The Data Civilizer Project”
04/2018 · ICDE – France: “Schema-agnostic Progressive Entity Resolution”
03/2017 · Paris Descartes University – France: “Entity Resolution for Big Data”
09/2016 · VLDB – India: “BLAST: a loosely schema-aware meta-blocking approach for entity resolution”

2.6 Education

03/2016
PhD at
UNIMORE

UNIVERSITÀ DEGLI STUDI DI MODENA E REGGIO EMILIA — ICT International Doctoral School
Grade: Ottimo (highest grade)
Thesis: *Loosely Schema-aware Techniques for Big Data Integration*
Advisor: Prof. Sonia BERGAMASCHI · Università degli Studi di Modena e Reggio Emilia
Co-advisor: Prof. H.V. JAGADISH · University of Michigan, Ann Arbor, US

02/2012
MSc in Computer
Engineering

UNIVERSITÀ DEGLI STUDI DI MODENA E REGGIO EMILIA
Grade: 110/110 cum laude

2.7 Publications

Metrics (Mar 2023)		
	Google Scholar	Scopus
#citazioni	691	491
h-index	15	14

- [1] L. Gagliardelli, G. Papadakis, G. **Simonini**, S. Bergamaschi, and T. Palpanas, “GSM: A generalized approach to supervised meta-blocking for scalable entity resolution,” *Inf. Syst.*, vol. 120, p. 102307, 2024. DOI: [10.1016/J.IS.2023.102307](https://doi.org/10.1016/J.IS.2023.102307). [Online]. Available: <https://doi.org/10.1016/j.is.2023.102307>.
- [2] A. D. Angelis, M. Mazzei, F. Piai, P. Merialdo, G. **Simonini**, L. Zecchini, S. Bergamaschi, D. Firmani, X. Chu, P. Li, and R. Wu, “Experiences and lessons learned from the SIGMOD entity resolution programming contests,” *SIGMOD Rec.*, vol. 52, no. 2, pp. 43–47, 2023. DOI: [10.1145/3615952.3615965](https://doi.org/10.1145/3615952.3615965). [Online]. Available: <https://doi.org/10.1145/3615952.3615965>.
- [3] A. Aslam, G. **Simonini**, L. Gagliardelli, A. Mozzillo, and S. Bergamaschi, “HKS: efficient data partitioning for stateful streaming,” in *Big Data Analytics and Knowledge Discovery - 25th International Conference, DaWaK 2023, Penang, Malaysia, August 28-30, 2023, Proceedings*, R. Wrembel, J. Gamper, G. Kotsis, A. M. Tjoa, and I. Khalil, Eds., ser. Lecture Notes in Computer Science, vol. 14148, Springer, 2023, pp. 386–391. DOI: [10.1007/978-3-031-39831-5_35](https://doi.org/10.1007/978-3-031-39831-5_35). [Online]. Available: https://doi.org/10.1007/978-3-031-39831-5_35.
- [4] D. Firmani, J. G. Mathew, D. Santoro, G. **Simonini**, and L. Zecchini, “Bridging the gap between buyers and sellers in data marketplaces with personalized datasets,” in *Proceedings of the 31st Symposium of Advanced Database Systems, Galzingano Terme, Italy, July 2nd to 5th, 2023*, D. Calvanese, C. Diamantini, G. Faggioli, N. Ferro, S. Marchesin, G. Silvello, and L. Tanca, Eds., ser. CEUR Workshop Proceedings, vol. 3478, CEUR-WS.org, 2023, pp. 525–534. [Online]. Available: <https://ceur-ws.org/Vol-3478/paper01.pdf>.

- [5] L. Gagliardelli, D. Beneventano, M. Esposito, L. Zecchini, G. **Simonini**, S. Bergamaschi, F. Miselli, and G. C. Miano, "DXP: billing data preparation for big data analytics," *CoRR*, vol. abs/2312.12902, 2023. DOI: [10.48550/ARXIV.2312.12902](https://doi.org/10.48550/ARXIV.2312.12902). arXiv: [2312.12902](https://arxiv.org/abs/2312.12902). [Online]. Available: <https://doi.org/10.48550/arXiv.2312.12902>.
- [6] L. Gagliardelli, G. Papadakis, G. **Simonini**, S. Bergamaschi, and T. Palpanas, "A general approach to supervised meta-blocking," in *Proceedings of the 31st Symposium of Advanced Database Systems, Galzingano Terme, Italy, July 2nd to 5th, 2023*, D. Calvanese, C. Diamantini, G. Faggioli, N. Ferro, S. Marchesin, G. Silvello, and L. Tanca, Eds., ser. CEUR Workshop Proceedings, vol. 3478, CEUR-WS.org, 2023, pp. 141–150. [Online]. Available: <https://ceur-ws.org/Vol-3478/paper84.pdf>.
- [7] L. Gagliardelli, L. Zecchini, L. Ferretti, D. Beneventano, G. **Simonini**, S. Bergamaschi, M. Orsini, L. Magnotta, E. Mescoli, A. Livaldi, N. Gessa, P. D. Sabbata, G. D'Agosta, F. Paolucci, and F. Moretti, "A big data platform exploiting auditable tokenization to promote good practices inside local energy communities," *Future Gener. Comput. Syst.*, vol. 141, pp. 595–610, 2023. DOI: [10.1016/J.FUTURE.2022.12.007](https://doi.org/10.1016/J.FUTURE.2022.12.007). [Online]. Available: <https://doi.org/10.1016/j.future.2022.12.007>.
- [8] A. Mozzillo, L. Zecchini, L. Gagliardelli, A. Aslam, S. Bergamaschi, and G. Simonini, "Evaluation of dataframe libraries for data preparation on a single machine," *CoRR*, vol. abs/2312.11122, 2023. DOI: [10.48550/ARXIV.2312.11122](https://doi.org/10.48550/ARXIV.2312.11122). arXiv: [2312.11122](https://arxiv.org/abs/2312.11122). [Online]. Available: <https://doi.org/10.48550/arXiv.2312.11122>.
- [9] G. **Simonini**, L. Zecchini, F. Naumann, and S. Bergamaschi, "Entity resolution on-demand for querying dirty datasets," in *Proceedings of the 31st Symposium of Advanced Database Systems, Galzingano Terme, Italy, July 2nd to 5th, 2023*, D. Calvanese, C. Diamantini, G. Faggioli, N. Ferro, S. Marchesin, G. Silvello, and L. Tanca, Eds., ser. CEUR Workshop Proceedings, vol. 3478, CEUR-WS.org, 2023, pp. 410–419. [Online]. Available: <https://ceur-ws.org/Vol-3478/paper70.pdf>.
- [10] L. Zecchini, G. **Simonini**, S. Bergamaschi, and F. Naumann, "Brewer: Entity resolution on-demand," *Proc. VLDB Endow.*, vol. 16, no. 12, pp. 4026–4029, 2023. DOI: [10.14778/3611540.3611612](https://doi.org/10.14778/3611540.3611612). [Online]. Available: <https://www.vldb.org/pvldb/vol16/p4026-zecchini.pdf>.
- [11] L. Gagliardelli, G. Papadakis, G. **Simonini**, S. Bergamaschi, and T. Palpanas, "Generalized supervised meta-blocking," *Proc. VLDB Endow.*, vol. 15, no. 9, pp. 1902–1910, 2022. DOI: [10.14778/3538598.3538611](https://doi.org/10.14778/3538598.3538611). [Online]. Available: <https://www.vldb.org/pvldb/vol15/p1902-gagliardelli.pdf>.

- [12] L. Gagliardelli, G. Papadakis, G. **Simonini**, S. Bergamaschi, and T. Palpanas, "Generalized supervised meta-blocking (technical report)," *CoRR*, vol. abs/2204.08801, 2022. DOI: [10.48550/ARXIV.2204.08801](https://doi.org/10.48550/ARXIV.2204.08801). arXiv: [2204.08801](https://arxiv.org/abs/2204.08801). [Online]. Available: <https://doi.org/10.48550/arXiv.2204.08801>.
- [13] L. Gagliardelli, L. Zecchini, D. Beneventano, G. **Simonini**, S. Bergamaschi, M. Orsini, L. Magnotta, E. Mescoli, A. Livaldi, N. Gessa, P. D. Sabbata, G. D'Agosta, and F. Paolucci, "ECDP: A big data platform for the smart monitoring of local energy communities," in *Proceedings of the Workshops of the EDBT/ICDT 2022 Joint Conference, Edinburgh, UK, March 29, 2022*, M. Ramanath and T. Palpanas, Eds., ser. CEUR Workshop Proceedings, vol. 3135, CEUR-WS.org, 2022. [Online]. Available: https://ceur-ws.org/Vol-3135/dataplat%5C_short4.pdf.
- [14] G. **Simonini**, L. Gagliardelli, M. Rinaldi, L. Zecchini, G. D. Sabbata, A. Aslam, D. Beneventano, and S. Bergamaschi, "Progressive entity resolution with node embeddings," in *Proceedings of the 30th Italian Symposium on Advanced Database Systems, SEBD 2022, Tirrenia (PI), Italy, June 19-22, 2022*, G. Amato, V. Bartalesi, D. Bianchini, C. Gennaro, and R. Torlone, Eds., ser. CEUR Workshop Proceedings, vol. 3194, CEUR-WS.org, 2022, pp. 52–60. [Online]. Available: <https://ceur-ws.org/Vol-3194/paper6.pdf>.
- [15] G. **Simonini**, L. Zecchini, S. Bergamaschi, and F. Naumann, "Entity resolution on-demand," *Proc. VLDB Endow.*, vol. 15, no. 7, pp. 1506–1518, 2022. DOI: [10.14778/3523210.3523226](https://doi.org/10.14778/3523210.3523226). [Online]. Available: <https://www.vldb.org/pvldb/vol15/p1506-simonini.pdf>.
- [16] G. M. Mandilaras, G. Papadakis, L. Gagliardelli, G. **Simonini**, E. Thanos, G. Giannakopoulos, S. Bergamaschi, T. Palpanas, M. Koubarakis, A. Lara-Clares, and A. Fariña, "Reproducible experiments on three-dimensional entity resolution with jedai," *Inf. Syst.*, vol. 102, p. 101 830, 2021. DOI: [10.1016/J.IS.2021.101830](https://doi.org/10.1016/J.IS.2021.101830). [Online]. Available: <https://doi.org/10.1016/j.is.2021.101830>.
- [17] G. **Simonini**, H. Sacconi, L. Gagliardelli, L. Zecchini, D. Beneventano, and S. Bergamaschi, "The case for multi-task active learning entity resolution (discussion paper)," in *Proceedings of the 29th Italian Symposium on Advanced Database Systems, SEBD 2021, Pizzo Calabro (VV), Italy, September 5-9, 2021*, S. Greco, M. Lenzerini, E. Masciari, and A. Tagarelli, Eds., ser. CEUR Workshop Proceedings, vol. 2994, CEUR-WS.org, 2021, pp. 363–370. [Online]. Available: <https://ceur-ws.org/Vol-2994/paper40.pdf>.
- [18] D. Beneventano, S. Bergamaschi, L. Gagliardelli, and G. Simonini, "BLAST2: An efficient technique for loose schema information extraction from heterogeneous big data sources," *ACM J. Data Inf. Qual.*, vol. 12, no. 4, 18:1–18:22, 2020. DOI: [10.1145/3394957](https://doi.org/10.1145/3394957). [Online]. Available: <https://doi.org/10.1145/3394957>.

- [19] L. Gagliardelli, G. **Simonini**, and S. Bergamaschi, “Ruler: Scaling up record-level matching rules,” in *Proceedings of the 23rd International Conference on Extending Database Technology, EDBT 2020, Copenhagen, Denmark, March 30 - April 02, 2020*, A. Bonifati, Y. Zhou, M. A. V. Salles, A. Böhm, D. Olteanu, G. H. L. Fletcher, A. Khan, and B. Yang, Eds., OpenProceedings.org, 2020, pp. 611–614. DOI: [10.5441/002/EDBT.2020.76](https://doi.org/10.5441/002/EDBT.2020.76). [Online]. Available: <https://doi.org/10.5441/002/edbt.2020.76>.
- [20] L. Gagliardelli, G. **Simonini**, and S. Bergamaschi, “Scaling up record-level matching rules,” in *Proceedings of the 28th Italian Symposium on Advanced Database Systems, Villasimius, Sud Sardegna, Italy (virtual due to Covid-19 pandemic), June 21-24, 2020*, M. Agosti, M. Atzori, P. Ciaccia, and L. Tanca, Eds., ser. CEUR Workshop Proceedings, vol. 2646, CEUR-WS.org, 2020, pp. 12–23. [Online]. Available: <https://ceur-ws.org/Vol-2646/27-paper.pdf>.
- [21] G. Papadakis, G. M. Mandilaras, L. Gagliardelli, G. **Simonini**, E. Thanos, G. Giannakopoulos, S. Bergamaschi, T. Palpanas, and M. Koubarakis, “Three-dimensional entity resolution with jedai,” *Inf. Syst.*, vol. 93, p. 101 565, 2020. DOI: [10.1016/J.IS.2020.101565](https://doi.org/10.1016/J.IS.2020.101565). [Online]. Available: <https://doi.org/10.1016/j.is.2020.101565>.
- [22] G. Papadakis, L. Tsekouras, E. Thanos, N. Pittaras, G. **Simonini**, D. Skoutas, P. Isaris, G. Giannakopoulos, T. Palpanas, and M. Koubarakis, “Jedai³ : Beyond batch, blocking-based entity resolution,” in *Proceedings of the 23rd International Conference on Extending Database Technology, EDBT 2020, Copenhagen, Denmark, March 30 - April 02, 2020*, A. Bonifati, Y. Zhou, M. A. V. Salles, A. Böhm, D. Olteanu, G. H. L. Fletcher, A. Khan, and B. Yang, Eds., OpenProceedings.org, 2020, pp. 603–606. DOI: [10.5441/002/EDBT.2020.74](https://doi.org/10.5441/002/EDBT.2020.74). [Online]. Available: <https://doi.org/10.5441/002/edbt.2020.74>.
- [23] E. K. Rezig, L. Cao, G. **Simonini**, M. Schoemans, S. Madden, N. Tang, M. Ouzzani, and M. Stonebraker, “Dagger: A data (not code) debugger,” in *10th Conference on Innovative Data Systems Research, CIDR 2020, Amsterdam, The Netherlands, January 12-15, 2020, Online Proceedings*, www.cidrdb.org, 2020. [Online]. Available: <http://cidrdb.org/cidr2020/papers/p35-rezig-cidr20.pdf>.
- [24] L. Zecchini, G. **Simonini**, and S. Bergamaschi, “Entity resolution on camera records without machine learning,” in *Proceedings of the 2nd International Workshop on Challenges and Experiences from Data Integration to Knowledge Graphs co-located with 46th International Conference on Very Large Data Bases, DI2KG@VLDB 2020, Tokyo, Japan, August 31, 2020*, F. Piai, D. Firmani, V. Crescenzi, A. D. Angelis, X. L. Dong, M. Mazzei, P. Merialdo, and D. Srivastava, Eds., ser. CEUR Workshop Proceedings, vol. 2726, CEUR-WS.org, 2020. [Online]. Available: <https://ceur-ws.org/Vol-2726/paper3.pdf>.

- [25] F. Benedetti, D. Beneventano, S. Bergamaschi, and G. Simonini, "Computing inter-document similarity with context semantic analysis," *Inf. Syst.*, vol. 80, pp. 136–147, 2019. DOI: [10.1016/J.IS.2018.02.009](https://doi.org/10.1016/J.IS.2018.02.009). [Online]. Available: <https://doi.org/10.1016/j.is.2018.02.009>.
- [26] D. Beneventano, S. Bergamaschi, L. Gagliardelli, and G. Simonini, "Entity resolution and data fusion: An integrated approach," in *Proceedings of the 27th Italian Symposium on Advanced Database Systems, Castiglione della Pescaia (Grosseto), Italy, June 16-19, 2019*, M. Mecella, G. Amato, and C. Gennaro, Eds., ser. CEUR Workshop Proceedings, vol. 2400, CEUR-WS.org, 2019. [Online]. Available: <https://ceur-ws.org/Vol-2400/paper-17.pdf>.
- [27] L. Gagliardelli, G. **Simonini**, D. Beneventano, and S. Bergamaschi, "Sparker: Scaling entity resolution in spark," in *Advances in Database Technology - 22nd International Conference on Extending Database Technology, EDBT 2019, Lisbon, Portugal, March 26-29, 2019*, M. Herschel, H. Galhardas, B. Reinwald, I. Fundulaki, C. Binnig, and Z. Kaoudi, Eds., OpenProceedings.org, 2019, pp. 602–605. DOI: [10.5441/002/EDBT.2019.66](https://doi.org/10.5441/002/EDBT.2019.66). [Online]. Available: <https://doi.org/10.5441/002/edbt.2019.66>.
- [28] E. K. Rezig, L. Cao, M. Stonebraker, G. **Simonini**, W. Tao, S. Madden, M. Ouzzani, N. Tang, and A. K. Elmagarmid, "Data civilizer 2.0: A holistic framework for data preparation and analytics," *Proc. VLDB Endow.*, vol. 12, no. 12, pp. 1954–1957, 2019. DOI: [10.14778/3352063.3352108](https://doi.org/10.14778/3352063.3352108). [Online]. Available: <http://www.vldb.org/pvldb/vol12/p1954-rezig.pdf>.
- [29] G. **Simonini**, L. Gagliardelli, S. Bergamaschi, and H. V. Jagadish, "Scaling entity resolution: A loosely schema-aware approach," *Inf. Syst.*, vol. 83, pp. 145–165, 2019. DOI: [10.1016/J.IS.2019.03.006](https://doi.org/10.1016/J.IS.2019.03.006). [Online]. Available: <https://doi.org/10.1016/j.is.2019.03.006>.
- [30] G. **Simonini**, G. Papadakis, T. Palpanas, and S. Bergamaschi, "Schema-agnostic progressive entity resolution," *IEEE Trans. Knowl. Data Eng.*, vol. 31, no. 6, pp. 1208–1221, 2019. DOI: [10.1109/TKDE.2018.2852763](https://doi.org/10.1109/TKDE.2018.2852763). [Online]. Available: <https://doi.org/10.1109/TKDE.2018.2852763>.
- [31] G. **Simonini**, G. Papadakis, T. Palpanas, and S. Bergamaschi, "Schema-agnostic progressive entity resolution (extended version)," *CoRR*, vol. abs/1905.06385, 2019. arXiv: [1905.06385](https://arxiv.org/abs/1905.06385). [Online]. Available: <http://arxiv.org/abs/1905.06385>.
- [32] S. Bergamaschi, D. Beneventano, F. Mandreoli, R. Martoglia, F. Guerra, M. Orsini, L. Po, M. Vincini, G. **Simonini**, S. Zhu, L. Gagliardelli, and L. Magnotta, "From data integration to big data integration," in *A Comprehensive Guide Through the Italian Database Research Over the Last 25 Years*, ser. Studies in Big Data, S. Flesca, S. Greco, E. Masciari, and D. Saccà, Eds., vol. 31, Springer International

- Publishing, 2018, pp. 43–59. DOI: [10.1007/978-3-319-61893-7_3](https://doi.org/10.1007/978-3-319-61893-7_3). [Online]. Available: https://doi.org/10.1007/978-3-319-61893-7_3.
- [33] L. Gagliardelli, S. Zhu, G. **Simonini**, and S. Bergamaschi, “Bigdedup: A big data integration toolkit for duplicate detection in industrial scenarios,” in *Transdisciplinary Engineering Methods for Social Innovation of Industry 4.0 - Proceedings of the 25th ISPE Inc. International Conference on Transdisciplinary Engineering, TE 2018, Modena, Italy, July 3-6, 2018*, M. Peruzzini, M. Pellicciari, C. Bil, J. Stjepandic, and N. Wognum, Eds., ser. Advances in Transdisciplinary Engineering, vol. 7, IOS Press, 2018, pp. 1015–1023. DOI: [10.3233/978-1-61499-898-3-1015](https://doi.org/10.3233/978-1-61499-898-3-1015). [Online]. Available: <https://doi.org/10.3233/978-1-61499-898-3-1015>.
 - [34] L. Magnotta, L. Gagliardelli, G. **Simonini**, M. Orsini, and S. Bergamaschi, “MOMIS dashboard: A powerful data analytics tool for industry 4.0,” in *Transdisciplinary Engineering Methods for Social Innovation of Industry 4.0 - Proceedings of the 25th ISPE Inc. International Conference on Transdisciplinary Engineering, TE 2018, Modena, Italy, July 3-6, 2018*, M. Peruzzini, M. Pellicciari, C. Bil, J. Stjepandic, and N. Wognum, Eds., ser. Advances in Transdisciplinary Engineering, vol. 7, IOS Press, 2018, pp. 1074–1081. DOI: [10.3233/978-1-61499-898-3-1074](https://doi.org/10.3233/978-1-61499-898-3-1074). [Online]. Available: <https://doi.org/10.3233/978-1-61499-898-3-1074>.
 - [35] A. Pietrangelo, G. **Simonini**, S. Bergamaschi, F. Naumann, and I. K. Koumarelas, “Towards progressive search-driven entity resolution,” in *Proceedings of the 26th Italian Symposium on Advanced Database Systems, Castellana Grotte, Italy, June 24-27, 2018*, S. Bergamaschi, T. D. Noia, and A. Maurino, Eds., ser. CEUR Workshop Proceedings, vol. 2161, CEUR-WS.org, 2018. [Online]. Available: <https://ceur-ws.org/Vol-2161/paper16.pdf>.
 - [36] G. **Simonini**, L. Gagliardelli, S. Zhu, and S. Bergamaschi, “Enhancing loosely schema-aware entity resolution with user interaction,” in *2018 International Conference on High Performance Computing & Simulation, HPCS 2018, Orleans, France, July 16-20, 2018*, IEEE, 2018, pp. 860–864. DOI: [10.1109/HPCS.2018.00138](https://doi.org/10.1109/HPCS.2018.00138). [Online]. Available: <https://doi.org/10.1109/HPCS.2018.00138>.
 - [37] G. **Simonini**, G. Papadakis, T. Palpanas, and S. Bergamaschi, “Schema-agnostic progressive entity resolution,” in *34th IEEE International Conference on Data Engineering, ICDE 2018, Paris, France, April 16-19, 2018*, IEEE Computer Society, 2018, pp. 53–64. DOI: [10.1109/ICDE.2018.00015](https://doi.org/10.1109/ICDE.2018.00015). [Online]. Available: <https://doi.org/10.1109/ICDE.2018.00015>.
 - [38] S. Zhu, L. Gagliardelli, G. **Simonini**, and D. Beneventano, “How improve set similarity join based on prefix approach in distributed environment,” in *2018 International Conference on High Performance Computing & Simulation, HPCS 2018, Orleans, France, July 16-20, 2018*, IEEE, 2018, pp. 844–851. DOI: [10.1109/HPCS.2018.00136](https://doi.org/10.1109/HPCS.2018.00136). [Online]. Available: <https://doi.org/10.1109/HPCS.2018.00136>.

- [39] S. Zhu, G. Fiameni, G. **Simonini**, and S. Bergamaschi, "SOPJ: A scalable online provenance join for data integration," in *2017 International Conference on High Performance Computing & Simulation, HPCS 2017, Genoa, Italy, July 17-21, 2017*, IEEE, 2017, pp. 79–85. DOI: [10.1109/HPCS.2017.23](https://doi.org/10.1109/HPCS.2017.23). [Online]. Available: <https://doi.org/10.1109/HPCS.2017.23>.
- [40] S. Bergamaschi, D. Ferrari, F. Guerra, G. **Simonini**, and Y. Velegrakis, "Providing insight into data source topics," *J. Data Semant.*, vol. 5, no. 4, pp. 211–228, 2016. DOI: [10.1007/S13740-016-0063-6](https://doi.org/10.1007/S13740-016-0063-6). [Online]. Available: <https://doi.org/10.1007/s13740-016-0063-6>.
- [41] G. **Simonini** and S. Bergamaschi, "Enhancing entity resolution efficiency with loosely schema-aware techniques," in *24th Italian Symposium on Advanced Database Systems, SEBD 2016, Ugento, Lecce, Italy, June 19-22, 2016, Ugento, Lecce, Italia, June 19-22, 2016*, M. A. Bochicchio and G. Mecca, Eds., Matematicamente.it, 2016, pp. 270–277.
- [42] G. **Simonini**, S. Bergamaschi, and H. V. Jagadish, "BLAST: a loosely schema-aware meta-blocking approach for entity resolution," *Proc. VLDB Endow.*, vol. 9, no. 12, pp. 1173–1184, 2016. DOI: [10.14778/2994509.2994533](https://doi.org/10.14778/2994509.2994533). [Online]. Available: <http://www.vldb.org/pvldb/vol9/p1173-simonini.pdf>.
- [43] F. Guerra, G. **Simonini**, and M. Vincini, "Supporting image search with tag clouds: A preliminary approach," *Adv. Multim.*, vol. 2015, 439020:1–439020:10, 2015. DOI: [10.1155/2015/439020](https://doi.org/10.1155/2015/439020). [Online]. Available: <https://doi.org/10.1155/2015/439020>.
- [44] G. **Simonini** and S. Zhu, "Big data exploration with faceted browsing," in *2015 International Conference on High Performance Computing & Simulation, HPCS 2015, Amsterdam, Netherlands, July 20-24, 2015*, IEEE, 2015, pp. 541–544. DOI: [10.1109/HPCSim.2015.7237087](https://doi.org/10.1109/HPCSim.2015.7237087). [Online]. Available: <https://doi.org/10.1109/HPCSim.2015.7237087>.
- [45] S. Bergamaschi, D. Ferrari, F. Guerra, and G. **Simonini**, "Discovering the topics of a data source: A statistical approach," in *Proceedings of the Workshop on Surfacing the Deep and the Social Web co-located with the 13th International Semantic Web Conference (ISWC 2014), Riva del Garda, Trentino, Italy, October 19, 2014*, P. R. da Cunha, N. T. Nguyen, O. Boucelma, B. Cautis, and Y. Velegrakis, Eds., ser. CEUR Workshop Proceedings, vol. 1310, CEUR-WS.org, 2014. [Online]. Available: <https://ceur-ws.org/Vol-1310/paper3.pdf>.
- [46] M. Interlandi, G. **Simonini**, and S. Bergamaschi, "Towards declarative imperative data-parallel systems," in *22nd Italian Symposium on Advanced Database Systems, SEBD 2014, Sorrento Coast, Italy, June 16-18, 2014*, S. Greco and A. Picariello, Eds., 2014, pp. 97–104.

- [47] G. **Simonini** and F. Guerra, "Using big data to support automatic word sense disambiguation," in *International Conference on High Performance Computing & Simulation, HPCS 2014, Bologna, Italy, 21-25 July, 2014*, IEEE, 2014, pp. 311–314. doi: [10.1109/HPCSIM.2014.6903701](https://doi.org/10.1109/HPCSIM.2014.6903701). [Online]. Available: <https://doi.org/10.1109/HPCSim.2014.6903701>.
- [48] S. Bergamaschi, F. Guerra, and G. Simonini, "Keyword search over relational databases: Issues, approaches and open challenges," in *Bridging Between Information Retrieval and Databases - PROMISE Winter School 2013, Bressanone, Italy, February 4-8, 2013. Revised Tutorial Lectures*, N. Ferro, Ed., ser. Lecture Notes in Computer Science, vol. 8173, Springer, 2013, pp. 54–73. doi: [10.1007/978-3-642-54798-0_3](https://doi.org/10.1007/978-3-642-54798-0_3). [Online]. Available: https://doi.org/10.1007/978-3-642-54798-0_3.

3 TEACHING

3.1 Teaching Appointments

Teaching Appointments Summary			
Academic Year	Position	Institution	Course
2023-now	Assistant Professor	Università degli Studi di Modena e Reggio Emilia	Big Data Management C.d.L. Ingegneria Informatica
2023-now	Associate Professor	Università degli Studi di Modena e Reggio Emilia	Quantitative Methods and Computer Science C.d.L. Data Analysis for Economics and Management
2022-2023	Assistant Professor	Università degli Studi di Modena e Reggio Emilia	Quantitative Methods and Computer Science C.d.L. Data Analysis for Economics and Management
2021-2023	Assistant Professor	Università degli Studi di Modena e Reggio Emilia	Big Data Management and Governance C.d.L. Ingegneria Informatica
2020-2023	Assistant Professor	Università degli Studi di Modena e Reggio Emilia	HRIS and Data Science C.d.L. Economia (Relazioni del Lavoro)
2020-2023	Assistant Professor	Università degli Studi di Modena e Reggio Emilia	Basi di Dati C.d.L. Ingegneria Informatica (Mantova)
2017/2018	Adjunct Professor	Università degli Studi di Modena e Reggio Emilia	Basi di Dati C.d.L. Ingegneria Informatica
2016/2017	Adjunct Professor	Università degli Studi di Modena e Reggio Emilia	Sistemi Informativi e Basi di Dati C.d.L. Ingegneria Gestionale
2017	Short Course Professor	Università degli Studi di Modena e Reggio Emilia	Big data analytics and visualization Doctorate School in ICT
2017	Short Course Professor	Università degli Studi di Modena e Reggio Emilia	Big data analytics and visualization Ordine degli Ingegneri di Modena
2013-2016	Teaching Assistant	Università degli Studi di Modena e Reggio Emilia	Basi di Dati C.d.L. Ingegneria Informatica
2013-2016	Teaching Assistant	Università degli Studi di Modena e Reggio Emilia	Tecnologia delle Basi di Dati C.d.L. Ingegneria Informatica (magistrale)
2014-2018	Short Course Instructor	CINECA	Tools and techniques for massive data analysis (multiple editions for PhD students and researchers)

2020
UNIMORE

Adjunct Professor (professore a contratto) · 27 hours front teaching for the course of Foundation of Databases (Basi di Dati), DIF (Mantova) — corso di laurea triennale in Ingegneria Informatica dell'Università degli Studi di Modena e Reggio Emilia.

2017/2018
UNIMORE

Adjunct Professor (professore a contratto) · 27 hours front teaching for the course of Foundation of Databases (Basi di Dati), DIF — corso di laurea triennale in Ingegneria Informatica dell'Università degli Studi di Modena e Reggio Emilia.

2016/2017
UNIMORE

Adjunct Professor (professore a contratto) · 54 hours front teaching for the course of Foundation of Databases and Information Systems (Sistemi informativi e basi di dati), DISMI — corso di laurea triennale in Ingegneria Gestionale dell'Università degli Studi di Modena e Reggio Emilia.

2017 UNIMORE	Short Course Instructor · 16 hours front teaching for the Doctorate School in ICT at the University of Modena and Reggio Emilia, course name: “Big data analytics and visualization”.
2017 UNIMORE	Short Course Instructor · 16 hours front teaching for the “ordine degli ingegneri di Modena”, course name: “Big data analytics and visualization”.
2013-2016 UNIMORE	Teaching Assistant · Advanced Database Technologies (Tecnologia delle Basi di Dati) - Teaching Assistant — corso di laurea magistrale in Ingegneria Informatica dell’ Università degli Studi di Modena e Reggio Emilia.
2013-2016 UNIMORE	Teaching Assistant · Introduction to Database (Basi di Dati e Lab.) — corso di laurea triennale in Ingegneria Informatica dell’ Università degli Studi di Modena e Reggio Emilia.
2014-present CINECA	Tools and techniques for massive data analysis at CINECA I held three days courses (×4 editions) on tools and techniques for Big Data (focused on the Apache Spark ecosystem), promoted and hosted by CINECA. Link: http://www.hpc.cineca.it/content/tools-and-techniques-massive-data-analysis

3.2 Student Supervisor

Academic Board:	·Member of the Academic Board for the Ph.D. in ICT of the Department of Engineering “Enzo Ferrari”, Università degli Studi di Modena e Reggio Emilia – since February 2019.
PhD students:	·(Co-advisor) Luca Zecchini. “TBD” – (XXXIV Cycle) PhD Università degli Studi di Modena e Reggio Emilia. Runner-up of SIGMOD 2020 programming contest ·(Co-advisor) Giuseppe Fiameni. “A distributed HPC infrastructure to process very large scientific data sets” – (XXXI Cycle) PhD obtained in 2019 at Università degli Studi di Modena e Reggio Emilia. Now at NVIDIA ·(Co-advisor) Luca Gagliardelli. “Big Data Integration” – (XXXII Cycle)PhD expected for 2020 at Università degli Studi di Modena e Reggio Emilia. Now at UniMoRe
MSc students:	·Gabriele Gorreri. “Human Resource Information Systems at Coopservice” – Graduated in 2022 at Università degli Studi di Modena e Reggio Emilia. (Advisor) ·Fabio Miselli. “Studio e progettazione di una tecnica di Multi-Source Entity Resolution” – Graduated in 2018 at Università degli Studi di Modena e Reggio Emilia. (Co-Advisor) ·Alberto Pietrangelo. “Progressive Search-driven Entity Resolution” – Graduated in 2018 at Università degli Studi di Modena e Reggio Emilia. (Co-Advisor) ·Giacomo Amici. “A Query-Driven Approach to Entity Resolution based on Data Ordering” – Graduated in 2018 at Università degli Studi di Modena e Reggio Emilia. (Co-Advisor)

·Andrea Spina. “Algoritmi di Machine Learning Distribuiti per Big Data: Confronto dei Sistemi Apache Flink e Apache Spark” – Graduated in 2016 at Università degli Studi di Modena e Reggio Emilia. (Co-Advisor)

·Guido Mazza. “Big Data streaming processing engines under the umbrella of the Apache Foundation: benchmark and industrial applications” – Graduated in 2016 at Università degli Studi di Modena e Reggio Emilia. (Co-Advisor)