

# Giovanni Simonini

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

CURRICULUM VITAE

UPDATED ON JUNE 2021

## CONTENTS

1	General Information	1
1.1	Education	1
2	Research Activities	2
2.1	Research Appointments	2
2.2	Visiting & Collaborations	3
2.3	Participation to Projects	3
2.4	Professional Services	3
2.5	Awards and Grants	4
2.6	Relevant Talks	4
2.7	Publications	4
3	Teaching	6
3.1	Teaching Appointments	6
3.2	Student Supervisor	7

## 1 GENERAL INFORMATION

*Full Name* Giovanni Simonini  
*Citizenship* Italian  
*email* [simonini@unimore.it](mailto:simonini@unimore.it)  
*Website* [www.giovanisimonini.com](http://www.giovanisimonini.com)

### 1.1 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 RESEARCH ACTIVITIES

### 2.1 Research Appointments

Research Appointments Summary			
Start	End	Institution	Position
01/10/2019	now	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

09/19-now  
RTDb  
UNIMORE

Assistant Professor (RTD b)  
Working on large-scale data curation.

10/18-09/19  
Postdoc at  
MIT CSAIL

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 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 Participation to Projects

### WP Coordinator

2018–2019 Human Brain Project	Deputy coordinator for the Work Package 7.2 “Data federation and data-intensive computing technology” (Human Brain Project - SGA2).
-------------------------------------	---

## 2.4 Professional Services

Editor	·Guest Editor of the special issue of Future Internet (ISSN 1999-5903): “Linked Open Data 2019” <a href="http://tiny.cc/ba796y">http://tiny.cc/ba796y</a>
Organizer	·Co-Chair of SIGMOD 2021 Programming Contest ·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 ·SIGMOD 2021 ·EDBT 2021 ·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 Awards and Grants

- 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/>
  - 2017 · Certificate of merit for national and international research from University of Modena and Reggio Emilia
- Grants*
- 2018 · Grant for the project "Automatic Application-driven Data Cleaning" from the "Fondo di Ateneo per la Ricerca 2018" (FAR JUNIOR) – 35,000 euro. 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

## 2.6 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"

## 2.7 Publications

Citation Metrics (as of June 21)		
	Google Scholar	Scopus
h-index	10	8
Citation Count	326	199

[c] = conference      [j] = journal      [b] = book chapter      [t] = Thesis

- [j7] **Three-dimensional Entity Resolution with JedAI** - Information Systems, Volume 93, November 2020, 101565. G. PAPADAKIS, G. MANDILARAS, L. GAGLIARDELLI, G. SIMONINI, E. THANOS, G. GIANNAKOPOULOS, S. BERGAMASCHI, T. PALPANAS, M. KOUBARAKIS.
- [j6] **BLAST<sub>2</sub>: an Efficient Technique for Loose Schema Information Extraction from heterogeneous Big Data sources** - Journal of Data and Information Quality , 12, 4, Article 18 (November 2020). D. BENEVENTANO, S. BERGAMASCHI, L. GAGLIARDELLI, G. SIMONINI.
- [c16] **RULER: Scaling Up Record-level Matching Rules** - EDBT 2020 (demo). L. GAGLIARDELLI, G. SIMONINI and S. BERGAMASCHI.

- [c15] **JedAI<sup>3</sup>: beyond batch, blocking-based Entity Resolution** - EDBT 2020 (demo). G. PAPADAKIS, L. TSEKOURAS, E. THANOS, N. PITTARAS, G. SIMONINI, D. SKOUTAS, P. ISARIS, G. GIANNAKOPOULOS, T. PALPANAS and M. KOUBARAKIS.
- [c14] **Dagger: A Data (not code) Debugger** - CIDR 2020. - E. REZIG, L. CAO, G. SIMONINI, M. SCHOEMANS, S. MADDEN, M. OUZZANI, N. TANG, M. STONEBRAKER.
- [c13] **Data Civilizer 2.0: A Holistic Framework for Data Preparation and Analytics.** - PVLDB 2019 (DEMO). - E. REZIG, L. CAO, M. STONEBRAKER, G. SIMONINI, W. TAO, S. MADDEN, M. OUZZANI, N. TANG, A. K. ELMAGARMID.
- [j5] **Scaling Entity Resolution: A Loosely Schema-aware Approach.** - Information Systems (2019). - G. SIMONINI, L. GAGLIARDELLI, S. BERGAMASCHI, H. V. JAGADISH.  
<https://doi.org/10.1016/j.is.2019.03.006>
- [c12] **Towards Progressive Search-driven Entity Resolution.** - Italian Symposium on Advanced Database Systems, SEBD 2018. - A. PIETRANGELO, G. SIMONINI, S. BERGAMASCHI, F. NAUMANN, I. KOUMARELAS,
- [j4] **Computing inter-document similarity with Context Semantic Analysis.** - Information Systems (2019). - F. BENEDETTI, D. BENEVENTANO, S. BERGAMASCHI, G. SIMONINI  
<https://doi.org/10.1016/j.is.2018.02.009>.
- [c11] **SparkER: Scaling Entity Resolution in Spark.** - EDBT 2019 (DEMO). - L. GAGLIARDELLI, G. SIMONINI, D. BENEVENTANO, S. BERGAMASCHI.  
[10.5441/002/edbt.2019.66](https://doi.org/10.5441/002/edbt.2019.66)
- [c10] **Schema-agnostic Progressive Entity Resolution.** - IEEE International Conf. on Data Engineering, ICDE 2018. - G. SIMONINI, G. PAPADAKIS, T. PALPANAS, S. BERGAMASCHI. [PDF]
- [c9] **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]
- [c8] **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]
- [b2] **Enhancing Big Data Exploration with Faceted Browsing.** - "Classification, (Big) Data Analysis and Statistical Learning", Springer. - G. SIMONINI, S. ZHU, S. BERGAMASCHI. [PDF]
- [b1] **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]
- [c7] **SOPJ: A Scalable Online Provenance Join for Data Integration.** - IEEE International Conf. on High Performance Computing & Simulation, HPCS 2017. - S. ZHU, G. FIAMENI, G. SIMONINI, S. BERGAMASCHI. [PDF]
- [c6] **BigBench workload executed by using Apache Flink** - Procedia Manufacturing 11 (2017): 695-702. - S. BERGAMASCHI, L. GAGLIARDELLI, G. SIMONINI, S. ZHU [PDF]

- [t] **Loosely Schema-aware Techniques for Big Data Integration** - PhD Thesis (2016) - G. Simonini [PDF]
- [j3] **BLAST: a loosely schema-aware meta-blocking approach for entity resolution** - Proceedings of the VLDB Endowment (PVLDB) 9.12 (2016): 1173-1184. - G. SIMONINI, S. BERGAMASCHI, H. V. JAGADISH [PDF] [CODE]
- [j2] **Providing Insight into Data Source Topics** - Journal on Data Semantics (2016): 1-18. - S. BERGAMASCHI, F. GUERRA, D. FERRARI, G. SIMONINI, Y. VELEGRAKIS [PDF]
- [c5] **Big data exploration with faceted browsing** - IEEE International Conf. on High Performance Computing & Simulation, HPCS 2015 - G. SIMONINI, Z. SONG [PDF]
- [c4] **Discovering the topics of a data source: a statistical approach** - SWSD Workshop @ISWC 2014 - G. SIMONINI, F. GUERRA, S. BERGAMASCHI [PDF]
- [c3] **Towards Declarative Imperative Data-parallel Systems** - Italian Symposium on Advanced Database Systems, SEBD 2014 - M. INTERLANDI, G. SIMONINI, S. BERGAMASCHI [PDF]
- [c2] **Using big data to support automatic Word Sense Disambiguation** - IEEE International Conf. on High Performance Computing & Simulation, HPCS 2014 - G. SIMONINI, F. GUERRA [PDF]
- [j1] **Supporting Image Search with Tag Clouds: a Preliminary Approach** - Advances in Multimedia, 2015 - G. SIMONINI, F. GUERRA, M. VINCINI [PDF]
- [c1] **Keyword Search over Relational Databases: Issues, Approaches and Open Challenges** - Bridging Between Information Retrieval and Databases. Springer, Berlin, Heidelberg, 2014. 54-73. - S. BERGAMASCHI, F. GUERRA, G. SIMONINI [PDF]

### 3 TEACHING

#### 3.1 Teaching Appointments

2020 <i>UNIMORE</i>	Adjunct Professor (professore a contratto) · 27 hours front teaching for the course of Foundation of Databases (Basi di Dati), DIEF (Mantova) — corso di laurea triennale in Ingegneria Informatica dell' Università degli Studi di Modena e Reggio Emilia.
2017/2018 <i>UNIMORE</i>	Adjunct Professor (professore a contratto) · 27 hours front teaching for the course of Foundation of Databases (Basi di Dati), DIEF — corso di laurea triennale in Ingegneria Informatica dell' Università degli Studi di Modena e Reggio Emilia.
2016/2017 <i>UNIMORE</i>	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 <i>UNIMORE</i>	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 <i>UNIMORE</i>	Short Course Instructor · 16 hours front teaching for the “ordine degli ingegneri di Modena”, course name: “Big data analytics and visualization”.

Teaching Appointments Summary			
Academic Year	Position	Institution	Course
2020-now	Assistant Professor	Università degli Studi di Modena e Reggio Emilia	Human Resource Information Systems and Data Science C.d.L. Economia (Relazioni del Lavoro)
2020-now	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)

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:*

- 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.
- Alberto Pietrangelo. "Progressive Search-driven Entity Resolution" – Graduated in 2018 at Università degli Studi di Modena e Reggio Emilia.
- 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.
- 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.
- 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.