



**ITA
DATA
2025**

The 4th Italian Conference on Big Data and Data Science

September 9-11, 2025
Turin, Italy

Sponsored by



Supported by



ITADATA2025

Welcome Message from General and Program Chairs

Welcome to the fourth edition of The Italian Conference on Big Data and Data Science (ITADATA)! Welcome to Turin!

The Italian Conference on Big Data and Data Science (ITADATA) is the annual event of the CINI National Laboratory on Data Science aimed at putting together Italian researchers and professionals from academia, industry and government working in the area of big data and data science, as well as on other related fields such as AI, security, privacy and HPC. The main scope of ITADATA is to discuss and shape the future of Big Data and Data Science, considering the multidisciplinary, complex, heterogeneous, and data-centric environment in which modern distributed systems are operating. Modern business processes increasingly require to timely manage and analyze large amounts of data, and to guarantee low-latency access to the data themselves. In this context, data science plays the fundamental role of developing and defining all the technologies needed to support construction of value on data in multidisciplinary areas at increasing complexity. ITADATA covers both theoretical and practical research pertaining to data, from data governance to data processing and analysis. It also covers research in related domains where data and data science technologies are key, such as cloud, artificial intelligence, edge and IoT, high-performance computing, blockchain, security and privacy, assurance and certification.

ITADATA2025 is the fourth edition of the Conference, hosted by Dipartimento di Informatica, Università degli Studi di Torino. This year, in addition to the main research track and the presentation track, ITADATA also includes a discussion paper track, with a summary of already-published papers or papers presenting ongoing work. Two selected distinguished keynote speeches, by Prof. Angela Bonifati, Dr. Horst Simon, and Prof. Emilio Porcu, will be delivered at the conference. Five tutorials on specific methodologies, solutions, and prototypes, three thematic workshops on vertical aspects related to big data and data science, and one panel on data-centric AI complete the program.

The fruitful collaboration between the CINI National Laboratory on Data Science and CINI National Lab on "HPC: Key Technologies and Tools" continues also this year with BigHPC, the third edition of the track dedicated to the enabling technologies for data science, including parallel, distributed, network-based, and quantum computing.

ITADATA2025 confirmed the growing trend that started in 2022. The conference attracted submissions from national and international researchers of different disciplines of interest. More specifically, the conference received a total of 50 submissions (25% increase from 2024). Following a rigorous double-blind review process, each submission was reviewed by at least three experts in the relevant areas for each paper, based on their significance, novelty, technical quality, presentation, and practical impact. After an intense post-review discussion by the program committee, the conference accepted 22 regular paper, 11 discussion papers, and 10 presentations.

ITADATA2025 was made possible by the effort of many volunteers who dedicated time and energy for the success of the conference. We would like to thank them all, as well as the program committee members for their efforts in reading, reviewing, and selecting the papers. Our appreciation extends to all the external reviewers for assisting the program committee. The quality of submissions was high and the selection of the papers to be presented was difficult. This process has resulted in a high-quality final program, which will certainly spark interesting discussions during the conference.

We would like to thank all authors who submitted their work to the conference and all of you who have come to ITADATA2025, hoping that you find it both stimulating and enjoyable.

Finally, we would like to thank all sponsors that contributed to the conference: Anti-Financial Crime Digital Hub, ICSC, Intesa Sanpaolo Innovation Center, Engineering, CeRICT, Namirial, Dipartimento di Informatica of the Università degli Studi di Torino, as well as our supporters INAF, USC-C, and PNRR Project FAIR.

General chairs

Claudio A. Ardagna

Università degli Studi di Milano

Mirko Polato

Università degli Studi di Torino

Luigi Romano

Università degli Studi di Napoli "Parthenope"

Program chairs

Nicola Bena

Università degli Studi di Milano

Michelangelo Ceci

Università degli Studi di Bari "Aldo Moro"

Roberto Esposito

Università degli Studi di Torino

Riccardo Torlone

Università degli Studi Roma Tre

Organizing Committee

Steering Committee

Claudio A. Ardagna, Università degli Studi di Milano
Angela Bonifati, Université Claude Bernard Lyon 1
Angelo Ciaramella, Università degli Studi di Napoli "Parthenope"
Davide Dalle Carbonare, Engineering
Ernesto Damiani, Khalifa University
Donato Malerba, Università degli Studi di Bari "Aldo Moro"

General Chairs

Claudio A. Ardagna, Università degli Studi di Milano
Mirko Polato, Università degli Studi di Torino
Luigi Romano, Università degli Studi di Napoli "Parthenope"

Program Chairs

Nicola Bena, Università degli Studi di Milano
Michelangelo Ceci, Università degli Studi di Bari
Roberto Esposito, Università degli Studi di Torino
Riccardo Torlone, Università degli Studi Roma Tre

Program Chairs – Presentation Track

Devis Bianchini, Università degli Studi di Brescia
Emanuel Di Nardo, Università degli Studi di Napoli "Parthenope"
Salvatore Distefano, Università degli Studi di Messina

Panel Chairs

Antonio Maratea, Università degli Studi di Napoli "Parthenope"
Ruggero Pensa, Università degli Studi di Torino
Giancarlo Ruffo, Università degli Studi del Piemonte Orientale
"Amedeo Avogadro"

Workshop Co-Chairs

Filippo Berto, Università degli Studi di Milano
Mattia Cerrato, University of Mainz
Beniamino Di Martino, Università degli Studi della Campania
"Luigi Vanvitelli"

Demo/Tutorial Chairs

Chiara Boldrini, CNR - IIT
Stefano Cirillo, Università degli Studi di Salerno
Attilio Fiandrotti, Università degli Studi di Torino

Local Arrangement Co-Chairs

Carlo Alberto Barbano, Università degli Studi di Torino
Samuele Fonio, Università degli Studi di Torino
Eniko Puhl, CINI
Lorenzo Sciandra, Università degli Studi di Torino

Website Co-Chairs

Antonio Pellicani, Università degli Studi di Bari "Aldo Moro"
Antongiacomo Polimeno, Università degli Studi di Milano
Vincenzo Turturro, Università degli Studi di Bari "Aldo Moro"

Publication Co-Chair

Alex Della Bruna, CINI

Publicity Chair

Antongiacomo Polimeno, Università degli Studi di Milano

Publicity Co-Chairs

Claudio Mattutino, Università degli Studi di Torino
Alex Mircoli, Università Politecnica delle Marche

Hackaton Co-Chairs

Luca Amoroso, Università degli Studi di Napoli "Parthenope"
Filippo Berto, Università degli Studi di Milano
Antonio Maratea, Università degli Studi di Napoli "Parthenope"
Fabrizio Montecchiani, Università degli Studi di Perugia

Program Committee

Marco Anisetti, Università degli Studi di Milano
Maurizio Atzori, Università degli Studi di Cagliari
Domenico Beneventano, Università degli Studi di Modena e Reggio Emilia
Nadia Bennani, INSA Lyon
Sonia Bergamaschi, Università degli Studi di Modena e Reggio Emilia
Alessio Cascione, Università di Pisa
Angelo Ciaramella, Università degli Studi di Napoli "Parthenope"
Stefano Cirillo, Università degli Studi di Salerno
Lorenzo Colombi, Università degli Studi di Ferrara
Salvatore Contino, Università degli Studi di Palermo
Giordano D'Aloisio, Università degli Studi dell'Aquila
Mattia D'Emidio, Università degli Studi dell'Aquila
Simon Dahdal, Università degli Studi di Ferrara
Edoardo Di Caro, Università degli Studi di Ferrara
Claudia Diamantini, Università Politecnica delle Marche
Sandro Luigi Fiore, Università degli Studi di Trento
Massimiliano Garda, Università degli Studi di Brescia
Chirine Ghedira-Guegan, Université Jean Moulin Lyon 3
Francesco Gullo, Unicredit
Mirko Lai, Università degli Studi del Piemonte Orientale "Amedeo Avogadro"
Domenico Lembo, Sapienza Università di Roma
Andrea Loddo, Università degli Studi di Cagliari
Antonella Longo, Università degli Studi del Salento
Samira Maghool, Università degli Studi di Milano
Donato Malerba, Università degli Studi di Bari "Aldo Moro"

Antonio Maratea, Università degli Studi di Napoli "Parthenope"
Davide Martinenghi, Politecnico di Milano
Fabrizio Montecchiani, Università degli Studi di Perugia
Vincenzo Moscato, Università degli Studi di Napoli "Federico II"
Serena Nicolazzo, Università degli Studi del Piemonte Orientale "Amedeo Avogadro"
Antonino Nocera, Università degli Studi di Pavia
Vincenzo Pasquadibisceglie, Università degli Studi di Bari "Aldo Moro"
Antonio Pellicani, Università degli Studi di Bari "Aldo Moro"
Luca Piovesan, Università degli Studi del Piemonte Orientale "Amedeo Avogadro"
Roberto Pirrone, Università degli Studi di Palermo
Tommaso Piselli, Università degli Studi di Perugia
Domenico Potena, Università Politecnica delle Marche
Giancarlo Ruffo, Università degli Studi del Piemonte Orientale "Amedeo Avogadro"
Salvatore Ruggieri, Università degli Studi di Pisa
Anna Sapienza, Università degli Studi del Piemonte Orientale "Amedeo Avogadro"
Irene Siragusa, Università degli Studi di Palermo
Giandomenico Solimando, Università degli Studi di Salerno
Michael Soprano, Università degli Studi di Udine
Giovanni Stilo, Università degli Studi dell'Aquila
Alessandro Sperduti, Università degli Studi di Padova
Domenico Talia, Università della Calabria
Orazio Tomarchio, Università degli Studi di Catania
Genoveva Vargas-Solar, CNRS
Michela Vespa, Università degli Studi di Ferrara

Logistics & Social Events

ITADATA2025 is held in two locations at the Università degli Studi di Torino: *Cavallerizza Reale* (rooms: *Aula magna* and *Aula multifunzionale*) and *Rettorato* (room: *Sala blu*). The **registration desk** will be located at Cavallerizza Reale, while **coffee breaks** and **lunches** will be held at **Rettorato**.



Social Event

The ITADATA 2025 Social Event will take place on **9th September starting at 18:00**.

Guided Tour

The social event will start at **18:00 with a one-hour guided tour of the city of Turin**. The tour will be organized in groups of up to 25 people and will depart from **Piazza Castello in front of the Teatro Regio**. There will be four guides: three Italian-speaking and one English-speaking.

Aperitivo

After the guided tour (**around 19:00**), participants are invited to join us at **GreenWood (Via Sant'Ottavio, 18/3, 10124 Torino)**, a stylish lounge bar offering stunning views of the Mole Antonelliana. Food will be served buffet-style. Each participant will receive one complimentary drink, redeemable with a ticket provided at the start of the aperitif, and there will also be free Prosecco available while supplies last!



Social Dinner

The Social dinner will take place on **10th September starting at 20:00** at **Circolo Eridano (Corso Moncalieri 88, Torino)**, a rowing club and very nice restaurant along the Po river. The ITADATA organization will provide a **shuttle service from and to Piazza Castello**. The shuttle has a capacity of ~50 passengers and will operate two departures, scheduled for approximately:

- Piazza Castello → Circolo Eridano: 19:30 and 19:45.
- Circolo Eridano → Piazza Castello: 23:00 and 23:15.

Program Overview

Legend					
Plenary	ITADATA	BigHPC	Workshop	Demo/Tutorial	ITADATAHack

Tuesday September 9th

2025-09-09	Aula Magna (Cavallerizza)	Aula Multifunzionale (Cavallerizza)	Sala Blu (Rettorato)
9:00-10:30	ITADATAHack		Workshop “Scientific HPC in the pre-Exascale era (2nd edition)”
10:30-11:00	Coffee Break		
11:00-11:15	Institutional Greetings		Workshop “Scientific HPC in the pre-Exascale era (2nd edition)”
11:15-12:15	Data Science for Healthcare I	Tutorial “Tutorial Mining and learning on temporal networks”	
12:15-13:00	Large Language Models		
13:00-14:00	Lunch		
14:00-15:00	Advances in Data Science and AI	Demo “Open Data and Air Quality Data for Personalized Route Planning in Smart Cities”	Workshop “Scientific HPC in the pre-Exascale era (2nd edition)”
15:00-16:00	Data Science and AI for Good I		
16:00-16:15	Coffee Break		
16:15-16:45	Award		Workshop “Scientific HPC in the pre-Exascale era (2nd edition)”
16:45-17:00	ITADATAHack		
18:00-20:30	Social event		

Wednesday September 10th

2025-09-10	Aula Magna (Cavallerizza)	Aula Multifunzionale (Cavallerizza)	Sala Blu (Rettorato)
9:00-10:00	Keynote Speech - Angela Bonifati "Property Graphs to the Rescue for Declarative Causal Analysis"		
10:00-11:00	Panel "Data-centric AI"		
11:00-11:30	Coffee Break		
11:30-12:15	Security, Privacy, and Beyond	BigHPC	Workshop "Secure Data Architectures and Post-Quantum Strategies: Technologies for Resilient ICT Systems"
12:15-13:00	Data Science: Technical Advances and Societal Perspectives		
13:00-14:00	Lunch		
14:00-16:00	Tutorial "Implementing Trusted European Data Spaces: From Strategy to Practice"	BigHPC	Workshop "Secure Data Architectures and Post-Quantum Strategies: Technologies for Resilient ICT Systems"
16:00-16:30	Coffee Break		
16:30-17:15	Storage and Querying	BigHPC	
17:15-18:00	Data Science and AI for Good II		
20:00-23:00	Social dinner		

Thursday September 11th

2025-09-11	Aula Magna (Cavallerizza)	Aula Multifunzionale (Cavallerizza)	Sala Blu (Rettorato)
9:30-10:30	Keynote Speech - Horst Simon and Emilio Porcu “ADIA Lab and Three Types of Climate Data Networks”		
10:30-11:00	Coffee Break		
11:00-12:00	Tutorial AFCH	Tutorial “Intelligent Statistics: Applied Methods in Healthcare, Conflict Monitoring, and High-Frequency Trading”	Workshop “Risks and Unintended Harms of Generative AI Systems”
12:00-13:00		Data Science and AI for Good III	
13:00-14:00	Lunch		
14:00-14:45	Fairness and Applications of Data Science		Workshop “Risks and Unintended Harms of Generative AI Systems”
14:45-15:30	Data Science for Healthcare II		

Program Details

Data Science for Healthcare I

September 9, 11:15–12:15

Session chair: Michelangelo Ceci, *Università degli Studi di Bari “Aldo Moro”*

Room: Aula Magna (Cavallerizza)

Diabetes Management: Non-Invasive Glucose Monitoring via ECG and Machine Learning

Matteo Ruggiero, Alessia Auriemma Citarella, Fabiola De Marco, Luigi Di Biasi, Genoveffa Tortora

Improving Preliminary Diagnosis in Italian Through a Retrieval-Augmented Medical Chatbot

Mariano Barone, Gian Marco Orlando, Marco Perillo, Giuseppe Riccio, Antonio Romano, Diego Russo, Ferdinando Tammaro, Valerio La Gatta, Marco Postiglione, Vincenzo Moscato

DART: A Structured Dataset of Regulatory Drug Documents in Italian for Clinical NLP

Mariano Barone, Antonio Laudante, Giuseppe Riccio, Antonio Romano, Marco Postiglione, Vincenzo Moscato

Homomorphic Encryption for EDSS Classification: Safeguarding Patient Privacy in MRI-Based Assessment of Multiple Sclerosis

Stefano Cirillo, Vincenzo Deufemia, Luigi Di Biasi, Giuseppe Polese, Giandomenico Solimando, Genoveffa Tortora

Large Language Models

September 9, 12:15–13:00

Session chair: Chair: Nicola Bena, *Università degli Studi di Milano*

Room: Aula Magna (Cavallerizza)

Investigating Edge Fine-Tuning of Large Language Models in a Federated Environment

Lorenzo Colombi, Michela Vespa, Francesco Resca, Sara Cavicchi, Edoardo di Caro, Elena Bellodi, Mauro Tortonesi, Cesare Stefanelli

Question Answering through Retrieval-Augmented Large Language Models: an Experimental Evaluation

Claudia Diamantini, Alex Mircoli, Alessia Pagnotta, Domenico Potena, Maria Cristina Recchioni, Emanuele Storti

PARSAL: Pipeline for Automatic Retrieval and Structuring of Academic Literature

Salvatore Contino, Irene Siragusa, Giovanni Sciortino, Roberto Pirrone

Advances in Data Science and AI

September 9, 14:00–15:00

Session chair: Claudia Diamantini, *Università Politecnica delle Marche*

Room: Aula Magna (Cavallerizza)

Space-efficient private estimation of quantiles

Massimo Cafaro, Angelo Coluccia, Italo Epicoco, Marco Pulimeno

HybridKAN: Leveraging Multi-Sized Sub-MLPs for Enhanced Performance

Fabio D'Andreamatteo, Andrea D'Angelo, Giovanni Stilo

BoXoR-C: Coherence and Directionality of Counterfactual Explanations

Antonio Maratea, Domenico Lugubre

Non-Euclidean Geometries for Prototype-Based Image Classification: a Reality Check

Samuele Fonio, Silvia Grosso, Sara Bouchenak

Data Science and AI for Good I

September 9, 15:00–16:00

Session chair: Lorenzo Colombi, *Università degli Studi di Ferrara*

Room: Aula Magna (Cavallerizza)

Representation of the temporal ego-networks through graph evolution rules: a tool for Web3 applications

Alessia Galdeman, Matteo Zignani, Christian Quadri, Sabrina Gaito

Augmented Narratives: A Data-Driven Platform for Interactive Storytelling based on Artificial Intelligence and Knowledge Representation techniques

Luigi Colucci Cante, Mariangela Graziano, Beniamino di Martino

Real-Time Detection of Social Media Disinformation Using DISARM: Twitter Case Study

Daniele Granata, Roberto Nardone

Security, Privacy, and Beyond

September 10, 11:30–12:15

Session chair: Chiara Boldrini, *CNR, IIT*

Room: Aula Magna (Cavallerizza)

Bridging LLMs and Symbolic Reasoning in Educational QA Systems: Insights from the XAI Challenge at IJCNN 2025

Long S. T. Nguyen, Khang H. N. Vo, Thu H. A. Nguyen, Tuan C. Bui, Duc Q. Nguyen, Thanh-Tung Tran, Anh D. Nguyen, Minh L. Nguyen, Fabien Baldacci, Thang H. Bui, Emanuel Di Nardo, Angelo Ciaramella, Son H. Le, Ihsan Ullah, Lorenzo Di Rocco, Tho T. Quan

Network Traffic-Based Malware Detection for Android Devices

Dincy R. Arikkat, Hasna Jabeen N. K., Mirko Lai, Serena Nicolazzo, Antonino Nocera, Anafa P., Vinod P., Rafidha Rehiman K.A.

A Privacy-Preserving Schema Matching Technique based on k-Anonymity

Domenico Beneventano, Sonia Bergamaschi, Carmelo La Corte

Data Science: Technical Advances and Societal Perspectives

September 10, 12:15–13:00

Session chair: Matteo Zignani, *Università degli Studi di Milano, France*

Room: Aula Magna (Cavallerizza)

Dividi et Impera: Enhancing Synthetic Data Fidelity through Data Partitioning

Yordanos Nebiyu Yifru, Salvatore Distefano

Exploring the Dataset Landscape for Automated Propaganda Detection: A Data-Centric Insight

Marco Usai, Davide Antonio Mura, Andrea Loddo, Manuela Sanguinetti, Luca Zedda, Cecilia Di Ruberto, Maurizio Atzori

The BotBid Project: Data Education at Work

Giorgio Delzanno, Giovanna Guerrini, Enrica Roccotiello, Angela Sugliano, Giovanni Zanone

S-PIC4CHU: Semantics-Enriched Techniques for Data Preparation in Data Science

Gianvincenzo Alfano, Ilaria Bartolini, Diego Calvanese, Paolo Ciaccia, Sergio Greco, Davide Lanti, Pasquale Leonardo Lazzaro, Emilia Lenzi, Davide Martinenghi, Cristian Molinaro, Marco Patella, Letizia Tanca, Riccardo Torlone, Irina Trubitsyna

Storage and Querying

September 10, 16:30–17:15

Session chair: Antonio Maratea, *Università degli Studi di Napoli "Parthenope"*

Room: Aula Magna (Cavallerizza)

From Data Sources to User Screens: a Multimodal hEalth Data lakehouse for ITALy (MEDITA)

Maria Paola Priola

Logic-based Semantics and Query Entailment for Knowledge Graphs

Roberto Maria Delfino, Maurizio Lenzerini, Antonella Poggi

Simplicial Information System

Emanuela Merelli (Università degli Studi di Camerino), Matteo Rucco (Biocentis)

A new framework for Integrated distributed data management, processing, and advanced analytics supporting Digital Twin systems: a Proof of Concept

Nicola Masi, Chiara Maria Capizzi, Francesco Patanè, Giuseppe Veneziano, Davide Profeta, Sergio Comella, Giada Sajeva, Lorenzo Gori, Susanna Bonura

Controlled Query Evaluation under Epistemic Protection Policies

Gianluca Cima, Domenico Lembo, Lorenzo Marconi, Riccardo Rosati, Domenico Fabio Savo

Data Science and AI for Good II

September 10, 17:15–18:00

Session chair: Roberto Esposito, *Università degli Studi di Torino*

Room: Aula Magna (Cavallerizza)

Climate Data IT: towards a National Data Hub for Climate Science

Sandro Fiore, Paola Nassisi, Ludovica Sacco, Gabriele Padovani, Fabrizio Antonio, Donatello Elia, Italo Epicoco, Nicola Giuseppe Marchioro, Marco Robol, Konstantinos Zefkilis, Flavio Vella, Carolina Sopranzetti, Andrea Alessandri, Annalisa Cherchi, Franco Catalano, Gianmaria Sannino, Gabriella Scipione, Marco Puccini, Erika Coppola, Ivan Giroto, Marco Reale, Stefano Salon, Antonella Galizia, Marcello Iotti, Hilary J Oliver, Valentine Anantharaj, Silvio Gualdi

Variational Inference for the Partial Credit Model

Paolo Frazzetto, Nicolò Navarin, Alessandro Sperduti

Urban Region Embeddings from Service-Specific Mobile Traffic Data

Giulio Loddi, Chiara Pugliese, Francesco Lettich, Chiara Renso, Fabio Pinelli

Analyzing and visualizing open urban data to study city accessibility and segregation

Mirko Lai, Anna Sapienza, Salvatore Vilella, Massimo Canonico, Federica Cena, Giancarlo Ruffo

Data Science and AI for Good III

September 11, 12:00–13:00

Session chair: Domenico Beneventano, *Università degli Studi di Modena e Reggio Emilia*

Room: Aula Multifunzionale (Cavallerizza)

A Deep Learning-based Anomaly Detection System for Railway Equipment

Michele Bevilacqua, Walter Didimo, Alessandro Eugeni, Fabrizio Montecchiani, Giovanni Spinelli

Aircraft data analysis and predictive maintenance: The experience of the 'Arrowhead-fpv' project

Gianluca Boleto, Gianluca Sommariva, Luca Oneto, Stefano Rovetta, Alberto Calvo, Luca Mortorano, Claudio Porretti, Davide Anguita

Automated Patent Landscaping via Citation Expansion and Semantic Filtering for Twin (Green and Digital) Patents

Francesca Ghinami

On effective ways to warn people against misleading contents

Luigi Catuogno, Emanuel Di Nardo, Clemente Galdi, Gennaro Ragucci

A Service-Oriented Architecture for the Digital Thread of Smart Products

Devis Bianchini, Massimiliano Garda, Anisa Rula

SLO-aware orchestration and autoscaling for microservice based applications

Orazio Tomarchio, Angelo Marchese

Fairness and Applications of Data Science

September 11, 14:00–14:45

Session chair: Valerio Bellandi, *Università degli Studi di Milano*

Room: Aula Magna (Cavallerizza)

Reconciling Statistical and Causal Metrics of Fairness in Machine Learning on Data-Driven Systems

Chiara Criscuolo, Davide Martinenghi, Jing Huang

Supporting human evaluation for the Fundamental Rights Impact Assessment of an AI-based hiring system

Roberta Savella, Roberto Trasarti, Francesca Pratesi, Marco Di Cristo

From Wearable Data to Computational Notebooks

Anna Khristodulo, Giancarlo Mascetti, Giorgio Delzanno, Marta Bassi, Nicola Minuto

The ego network of words

Kilian Ollivier, Chiara Boldrini, Andrea Passarella, Marco Conti

Data Science for Healthcare II

September 11, 14:45–15:30

Session chair: Nicola Bena, *Università degli Studi di Milano*

Room: Aula Magna (Cavallerizza)

A Federated Learning Architecture for Prostate MRI Image Segmentation

Alice Bovio, Mario Barile, Francesco Pallotta, Lorenzo Pede, Alessandro Maiocchi, Marco Alì, Fatemeh Darvizeh, Deborah Fazzini, Francesco Lacavalla, Massimo Banzi, Gabriele Gianini, Filippo Berto, Ruslan Bondaruc, Ernesto Damiani, Saman Fouladi

A secure AI-enabled platform to support the automated Alzheimer's disease diagnosis

Claudio A. Ardagna, Veronica Buttaro, Michelangelo Ceci, Marco Luzzara, Antonio Pellicani, Gianvito Pio, Antongiaco Polimeno

Health Data Sharing for Secondary Use: a Critical Overview

Alessio Famiani, Ruggero G. Pensa

Automatic Evaluation of Cancer Reduction during Radiotherapy

Valerio Bellandi, Stefano Siccardi

Keynote Speeches

Angela Bonifati: Property Graphs to the Rescue for Declarative Causal Analysis

10 September, 09:00–10:00

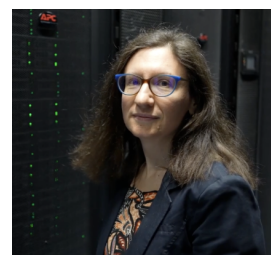
Room: Aula Magna (Cavallerizza)

Abstract

Graphs are powerful abstractions for modeling relationships across data and enabling complex data science tasks. In this talk, I will highlight powerful declarative graph operations and delineate their counterparts in causal inference, to support data-driven, personalized decision-making across several scientific domains. I will present our work to align causal analysis with property graphs—the foundation of modern graph databases—by rethinking graph models to incorporate hypernodes, structural equations, and causality-aware query semantics. By unifying graph databases with causal reasoning, causal tasks such as interventions and counterfactuals are mapped to property graph manipulation and transformation, combining expressiveness with computational efficiency.

Bio

Angela Bonifati is a Distinguished Professor of Computer Science at Lyon 1 University and at the CNRS Liris research lab, where she leads the Database Group. She is also an Adjunct Professor at the University of Waterloo in Canada from 2020 and a Senior member of the French University Institute (IUF) from 2023. Her current research interests are on several aspects of data management, including graph databases, knowledge graphs, data integration and their applications to data science and artificial intelligence. She has co-authored more than 200 publications in top venues of the data management field, including five Best Paper Awards, two books and an invited paper in ACM Sigmod Record 2018. She is a recipient of an ERC Advanced Grant 2024 dedicated to leading researchers in Europe. She is the youngest recipient of the prestigious IEEE TCDE Impact Award 2023 and a co-recipient of an ACM Research Highlights Award 2023. She is the General Chair of VLDB 2026 and was the Program Chair of IEEE ICDE 2025, ACM Sigmod 2022 and EDBT 2020. She is currently an Associate Editor for the Proceedings of VLDB Vol. 19 and for IEEE TKDE and ACM TODS. She is the Chair of the ACM Sigmod Executive Committee (2025-2029) and was the President of the EDBT Executive Board and Association (2020-2024). She is a member of the IEEE Technical Committee on Data Engineering (2024-2029) and a member of the PVLDB Board of Trustees (2024-2029).



Horst Simon and Emilio Porcu: ADIA Lab and Three Types of Climate Data Networks

11 September, 09:30–10:30

Room: Aula Magna (Cavallerizza)

Abstract

This presentation begins with a brief introduction to ADIA Lab, an independent research institute based in Abu Dhabi. ADIA Lab's mission is to advance fundamental and applied research in computational and data science, with a focus on addressing complex real-world challenges across domains such as climate, finance, and health. By fostering interdisciplinary collaborations and developing scalable algorithms and models, the lab seeks to drive innovation at the intersection of theory and practice. As an example of the projects underway at ADIA Lab, we then present our research on climate networks, structured around three complementary paradigms: (a) networks of data, where connections between geographic nodes are derived from statistical relationships, commonly referred to as Tsonis networks; (b) climate data over networks, where climate variables are defined on fixed topologies such as river basins or atmospheric grids—termed geophysical networks; and (c) networks for data, which leverages machine learning and statistical models grounded in network theory to analyze and interpret climate information. Special emphasis is placed on the first two frameworks. We examine how these network types are constructed, the insights they offer for understanding climate variability and model output, and their implications for climate governance. Finally, we discuss how integrating these perspectives can inform more robust analytical tools and policy strategies in the face of climate change.

Bio

Dr. Horst D. Simon is an internationally recognized expert in high-performance computing and computational science, with over four decades of experience in parallel algorithms and large-scale numerical methods. After completing his Ph.D. in Mathematics at UC Berkeley, he held leadership roles across academia (Stony Brook, UC Berkeley), industry (Boeing, SGI), and national research labs (NASA Ames, Lawrence Berkeley Lab. At Berkeley Lab, he directed NERSC and served as Deputy Lab Director, earning two prestigious Gordon Bell Prizes and co-editing the biannual TOP500 list of supercomputers. Since 2023, Dr. Simon has been the founding Director of ADIA Lab in Abu Dhabi, spearheading cutting-edge research in computational and data science. He continues to advance scalable algorithms that address complex scientific and societal challenges.



Professor Emilio Porcu is a distinguished expert in spatial-temporal statistics and data science. He earned his Ph.D. in statistics in 2005 and became a full professor by 2012 while holding Chair positions at Newcastle University and Trinity College Dublin. Since August 2020, he has served as Professor of Statistics & Data Science at Khalifa University in Abu Dhabi, and he continues as an adjunct professor at Trinity College Dublin. Dr. Porcu leads a research group specializing in space-time covariance modeling, producing nearly 180 papers and advancing theory and application in areas such as Gaussian random fields on non-Euclidean domains, covariance functions on spheres and networks, and scalable kernel-based methods. He is also a Research Fellow at ADIA Lab, contributing to innovative data science initiatives in climate, complex systems, and computational statistics.



BigHPC

The fruitful collaboration between the CINI National Laboratory on Data Science and the CINI National Laboratory on "HPC: Key Technologies and Tools" continues also this year with the third edition of BigHPC, a special track dedicated to the enabling technologies for data science, including parallel, distributed, network-based, and quantum computing. More information about BigHPC can be found at <https://www.itadata.it/2025/bighpc2025>.

BigHPC 2025 Chairs: Alessia Antelmi (*University of Turin*), Massimo Cafaro (*University of Salento*), Italo Epicoco (*University of Salento*), Marco Pulimeno (*University of Salento*)

When: 10 September, 11:30-18:00

Where: Aula multifunzionale (Cavallerizza)

Schedule

September 10, 11:30–13:00

Room: Aula multifunzionale (Cavallerizza)

11:30-11:45 | **Opening**

11:45-12:45 | **Keynote:** Prof. Biagio Cosenza

- *Title*
Beyond Performance: Software Approaches for Energy-Efficient HPC
- *Abstract*
Energy efficient computing is a major challenge for exascale computing. Power constraints, rising electricity costs, and the diminishing efficiency benefits of Moore's Law have exacerbated this challenge and increased the need for energy-efficient technologies. This talk will present state-of-the-art software approaches aimed at minimizing the energy consumption of HPC applications. We will discuss some common misconceptions about energy optimization and provide insights into understanding when optimizing for energy differs from optimizing for performance. We will see how different approaches in HPC can target energy optimization from different perspectives, such as job schedulers, compilers, and operating systems. We will see how recent research targets fine-grained solutions that span the software stack (SYnergy [SC23]), including programming models, compilers, runtime systems, and extensions to phase-based approaches [IPDPS25]. Finally, the talks conclude with an overview of emerging trends in fields including carbon-efficient computing.

September 10, 14:00–16:00

Room: Aula multifunzionale (Cavallerizza)

14:00-14:25 | **dqsweep: Parameter Sweeps for Benchmarking Distributed Quantum Computing Applications**

Quang-Phong Tran (IIT-CNR), Claudio Cicconetti (IIT-CNR)*, Marco Conti (IIT-CNR), and Andrea Passarella (IIT-CNR)

14:25-14:50 | **The Quantum Bottleneck: An Analysis of Data Balancing in QML for Security**

Mansur Ziatdinov (UNIME)*, and Salvatore Distefano (UNIME)

14:50-15:05 | **Exploring an implementation of quantum learning pipeline for support vector machines**

Mario Bifulco (UNITO), and Luca Roversi (UNITO)

15:05-16:00 | **Panel: What's Next for BigHPC? Parallel, Distributed, and Quantum Horizons**

Claudio Cicconetti (IIT-CNR), Biagio Cosenza (UNISA), and Sandro Fiore (UNITN)

September 10, 16:30–18:00

Room: Aula multifunzionale (Cavallerizza)

16:30-16:55 | **Hiding Latencies in Network-Based Image Loading for Deep Learning**

Francesco Versaci (CRS4)*, and Giovanni Busonera (CRS4)

16:55-17:20 | **The impact of ECS logic on parallel performance in agent-based model simulations**

Daniele De Vinco (UNISA)*, Antonio Ambrosio (UNISA), Francesco Foglia (UNISA), Carmine Spagnuolo (UNISA), and Vittorio Scarano (UNISA)

17:20-17:45 | **Federated Learning for Distributed Weather Forecasting: A Practical Approach on Real Multidimensional Georeferenced Data**

Giuseppe Fiorillo (UNIParthenope)*, Attilio Di Vicino (UNIParthenope), Luigi Galluccio (UNIParthenope), and Raffaele Montella (UNIParthenope)

17:45-18:00 | **Benchmarking Federated Learning Frameworks Scalability**

Gianluca Mittone (UNITO), Samuele Fonio (UNITO)*, Robert Birke (UNITO), and Marco Aldinucci (UNITO)

Workshops

ITADATA2025 features an exciting program of 3 workshops. Detailed information can be found at <https://itadata.it/program> and <https://www.itadata.it/workshop>.

Scientific HPC in the pre-Exascale era (2nd edition)

Workshop chairs: Valentina Cesare, *INAF*, Alberto Vecchiato, *INAF*, Gianluca Mittone, *University of Turin*, Bruno Casella, *University of Turin*

9 September, 9:00–17:15

Room: Sala Blu (Rettorato)

In this epoch of technological evolution, data concerning problems of different scientific areas are steeply increasing in volume, requiring hundreds of PB of storage (Big Data). Specifically, these data are extensive in both single file size and number of files in many fields, such as astronomy, cosmology, biology, and meteorology. Maintaining a proper performance trend towards pre-Exascale systems requires a specific codesign between hardware and software, exploiting High-Performance Computing (HPC) techniques. On the hardware side, increasingly heterogeneous architectures with multiple nodes and accelerators linked with high-bandwidth bridges to the single node are required. On the software side, applications have to be written with programming languages that allow portability among diverse architectures while not losing performance and minimizing the time required by the programmer to adapt the application. Other important aspects are the maintainability of the numerical stability of a problem solution while increasing the system size and the number of computational resources and the requirement of a "green" solution, that is, the ability to build infrastructures and applications to compute operations with Big Data volumes without excessively increasing the energetic consumptions.

Secure Data Architectures and Post-Quantum Strategies: Technologies for Resilient ICT Systems

Workshop chairs: Daniele Granata, *Università degli Studi di Napoli "Parthenope"*

10 September, 11:30–18:00

Room: Sala Blu (Rettorato)

Ensuring data security in current and future ICT systems requires coordinated efforts across cryptographic research, software engineering, and institutional support. This workshop presents integrated strategies that address emerging threats, such as those posed by quantum computing, and practical challenges in secure software deployment. Experimental results on post-quantum TLS and digital signatures show the performance trade-offs of adopting quantum-resistant algorithms in real-world settings. These findings complement efforts to improve container security in CI/CD pipelines through automated threat analysis and enforcement mechanisms. The workshop also highlights the role of public funding and regional initiatives—such as those led by CERICT—in enabling collaborative research and innovation. This also draws attention to challenges at the network edge—where distributed, resource-constrained systems expand the attack surface and demand tailored security for edge-cloud environments. Together, the contributions provide a structured view of how secure architectures can be designed and deployed in scalable, future-ready environments.

Risks and Unintended Harms of Generative AI Systems

Workshop chairs: Stefano Cirillo, *University of Salerno*, Eliana Pastor, *Politecnico di Torino*, Francesco Pierri, *Politecnico di Milano*, Serena Tardelli, *IIT-CNR*

11 September, 11:00–15:30

Room: Sala Blu (Rettorato)

As generative AI systems such as large language models and synthetic media generators become deeply integrated into big data ecosystems—spanning journalism, governance, healthcare, cybersecurity, and other critical sectors—, they introduce complex challenges and critical questions around trust, governance, and harm mitigation. This half-day workshop addresses the intersection of generative AI and big data from both technical and ethical perspectives. This workshop aims to bring together researchers, practitioners, and policymakers to discuss unintended consequences of generative AI, explore real-world failure scenarios, and collaboratively identify solutions. The workshop includes a keynote and/or an interactive panel with audience Q&A, and lightning talks selected through an open call. We invite lightning talk submissions that bring forward original insights, early-stage ideas, case studies, and provocations. The workshop will foster interactive discussion and interdisciplinary exchange, with the goal of shaping a research and policy agenda for responsible generative AI integration in data science and big data environments.

Panels

ITADATA2025 features a plenary panel with distinguished panelists. Detailed information can be found at <https://itadata.it/program>.

Data-centric AI: qualità dei dati per un'intelligenza artificiale responsabile

Panel chair: Claudia Diamantini, *Università Politecnica delle Marche*

10 September, 10:00–11:00

Room: Aula Magna (Cavallerizza)

Introduction: Donato Malerba, *Università degli Studi di Bari "Aldo Moro"*

Panelists: Riccardo Torlone, *Università degli Studi Roma Tre*, Angela Bonifati, *Université Claude Bernard Lyon 1*, Antonella Poggi, *Sapienza Università di Roma*, Silvia Ronchiadin, *Intesa Sanpaolo Innovation Center – AFC Digital Hub*, Davide Dalle Carbonare, *Engineering Ingegneria Informatica*

In the current context, where Artificial Intelligence is increasingly embedded in critical applications, data quality emerges as an enabling factor for developing reliable, fair, and transparent systems. This panel addresses the paradigm of Data-centric AI, which shifts the focus from model optimization to data engineering, exploring the importance of dataset curation, annotation, and continuous monitoring. Through academic and industrial contributions, the panel will discuss approaches, challenges, and tools to ensure data consistency, representativeness, and robustness, positioning quality as a lever for truly responsible AI. Special attention will be given to methodological implications, real-world applications, and prospects for collaboration between research and industry.

Demos/Tutorials

ITADATA2025 features an exciting program of 5 tutorials. Detailed information can be found at <https://itadata.it/program>.

Mining and learning on temporal networks

Organizers: Alessia Galdeman, *IT University of Copenhagen*, Manuel Dileo, *Università degli Studi di Milano*, Matteo Zignani, *Università degli Studi di Milano*, Sabrina Gaito, *Università degli Studi di Milano*

9 September, 11:15–13:00

Room: Aula Multifunzionale (Cavallerizza)

In many real-world systems, the relationships between entities are not static but evolve over time and across multiple dimensions. Examples include interactions in social media platforms, financial transactions, communication networks, and decentralized blockchain infrastructures. These systems are best modeled as temporal networks, where nodes and edges are associated with time-varying properties, and often as heterogeneous networks, where multiple types of nodes and relations coexist. Modeling these dynamics is crucial for uncovering the underlying mechanisms that govern the evolution of complex systems.

Open Data and Air Quality Data for Personalized Route Planning in Smart Cities

Organizers: Laura Po, *University of Modena and Reggio Emilia*, Federica Rollo, *University of Modena and Reggio Emilia*

9 September, 14:00–16:00

Room: Aula Multifunzionale (Cavallerizza)

Urban mobility planning is evolving towards a more sustainable, health-conscious, and data-driven approach. This hands-on tutorial will guide participants through a practical methodology to integrate heterogeneous open data sources (GTFS, OpenStreetMap, air quality APIs) into actionable insights for personalized active mobility routing. Attendees will learn to model multi-layer transport networks, design adaptable routing algorithms, and evaluate trade-offs between safety, pollution exposure, and quality of experience specifically for pedestrians and cyclists.

Implementing Trusted European Data Spaces: From Strategy to Practice

Organizers: Antonella Longo, *University of Salento*, Angelo Martella, *University of Salento*, Cristian Martella, *University of Salento*

10 September, 14:00–16:00

Room: Aula Magna (Cavallerizza)

In its European Strategy for Data, the European Commission emphasized that citizens, businesses, and public authorities can be empowered to make better decisions through the effective use of data. To unlock the full potential of the vast and growing volume of data generated daily across Europe, it is essential to build trust in the ways data are collected, accessed, and used. In this context, the development of a robust European Data Spaces ecosystem is seen as a key opportunity to ensure better access to data and its responsible use. A Data Space is a structured framework that facilitates secure and trustworthy data sharing within a defined data ecosystem. It provides clear rules and technical means for participants to exchange, trade, and collaborate on data assets in a manner that complies with relevant legal and regulatory requirements, while guaranteeing fair treatment for all stakeholders (Source: Data Spaces Support Centre (DSSC) website). The establishment of Common European Data Spaces across key sectors enables the seamless sharing and reuse of data both within and across domains. This contributes to the emergence of a competitive and dynamic European data market, underpinned by the EU's strong legal foundations in personal data protection, intellectual property rights, safety, and cybersecurity. In turn, this facilitates greater demand for data-driven products and services across the continent. By promoting standardized technologies and legal frameworks, Data Spaces aim to overcome barriers to data sharing among organizations and foster the creation of innovative, value-added services. Through a focus on interoperability and data sovereignty, Data Spaces play a pivotal role in advancing collaboration and innovation across the European data economy. The aim of this tutorial is to provide an overview of the current landscape of Data Spaces initiatives and to introduce selected open-source platforms that support their practical deployment.

Anti-Financial Crime

Organizer: Silvia Ronchiadin, *Intesa Sanpaolo Innovation Center – AFC Digital Hub*

11 September, 11:00–13:00

Room: Aula Magna (Cavallerizza)

The tutorial aims to explore the activities of Intesa Sanpaolo Anti-Financial Crime Digital Hub, and is structured as follows.

- Dario Moncalvo, “Anti Financial Crime: “Practices and R&D priorities”
- Anti-Financial Crime Digital Hub use cases:
 - Riccardo Crupi, “Improving Efficiency of Detection Analysis – ImEDA”
 - Flavio Giobergia, “Anomaly Detection of Suspicious Financial Accounts”
 - Alan Perotti, “Operationalising Explainability”
 - Roberto Esposito, “Federated Learning for disambiguating account owners in bank transactions”

Intelligent Statistics: Applied Methods in Healthcare, Conflict Monitoring, and High-Frequency Trading

Organizers: Silvia Salini, *Università degli Studi di Milano*

Presenters: Irene Siragusa, *University of Palermo*, Marco Tagliapietra, *University of Turin*, Luca Macis, *University of Turin*, Jacopo Chiapparino

11 September, 11:00–12:00

Room: Aula Multifunzionale (Cavallerizza)

The tutorial aims to showcase recent contributions in the field of data analysis, where statistical modeling and machine learning intersect to address real-world challenges. Emphasis will be placed on the interpretability and practical implications of the proposed techniques, particularly in the analysis of high-dimensional, complex, and multisource datasets. Participants will gain insights into how advanced statistical tools and AI-driven models can enhance understanding and decision-making across domains such as healthcare, finance, and social dynamics. It includes the following presentations:

- Irene Siragusa, Marco Speciale, Roberto Pirrone, Antonella Plaia, “Automated Clinical Coding models”
- Marco Tagliapietra, Luca Macis, Elena Siletta, Paola Pisano, “Composite Indicators of Global Unrest: A Quantitative Approach to Instability”
- Jacopo Chiapparino, Paola Cerchiello, “LLM-Based Document Segmentation for Enhanced RAG Performance”