consorzio interuniversitario nazionale per l'informatica

DATA SCIENCE | CINI NATIONAL LAB

UNIVERSITÀ DEGLI STUDI DI BARI ALDO MORO

ITA DATA 2026

The 5th Italian Conference on Big Data and Data Science

November 10–12, 2026
Bari, Italy

Sponsored by

MEDITECH COMPETENCE CENTER

Centro Regionale Information Communication Technology CeRICT SCRL

FINCANTIERI

ENDURANCE

avant

Supported by

INAF USC-C

INAF ISTITUTO NAZIONALE DI ASTROFISICA

eng

ITADATA2026

Welcome Message from General and Program Chairs

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

The Italian Conference on Big Data and Data Science (ITADATA) is the annual event of the CINI National Laboratory on Data Science aimed at bringing together Italian researchers and professionals from academia, industry, and government working in the big data and data science area, as well as in 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 operate. Actual business processes increasingly require timely management and analysis of large amounts of data, as well as guaranties for low-latency access to the data itself. In this context, data science plays a fundamental role: defining and developing the enabling technologies to support value creation from data in multidisciplinary areas while 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.

ITADATA2026 is the fifth edition of the Conference, hosted at the Department of Computer Science, University of Bari Aldo Moro. This year, in addition to the main research track and the presentation track, ITADATA2026 also includes a discussion paper track, with a summary of already-published papers or papers presenting ongoing work, as well as an industrial paper track. Two selected distinguished keynote speeches, by Richard Lloyd Stevens and Ernesto Damiani, 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 two panels complete the program.

Since 2022, ITADATA is confirming the growing interest in data science topics. The conference attracted submissions from national and international researchers from different disciplines. More specifically, the conference received a total of 59 submissions (+18% from 2025). Following a rigorous review process, each submission has been 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 28 regular papers, 9 discussion papers, 7 industrial papers, and 9 presentations.

ITADATA2026 has been made possible by the efforts of several volunteers who dedicated time and energy to 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 high quality of submissions made the selection of the papers difficult. This process 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 ITADATA2026, hoping that you find it both stimulating and enjoyable.

Finally, we would like to thank all sponsors contributing to the conference: Meditech Competence Center, CeRICT, Fincantieri, Avant, and Endurance, as well as our supporters INAF, USC-C, and Engineering.

General chairs

Claudio A. Ardagna
Università degli Studi di Milano
Michelangelo Ceci
Università degli Studi di Bari Aldo Moro
Donato Malerba
Università degli Studi di Bari Aldo Moro

Program chairs

Nicola Bena
Università degli Studi di Milano
Salvatore Distefano
Università degli Studi di Messina
Antonella Longo
Università degli Studi del Salento
Gianvito Pio
Università degli Studi di Bari Aldo Moro

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, Università degli Studi di Milano
Donato Malerba, Università degli Studi di Bari Aldo Moro
Luigi Romano, Università degli Studi di Napoli "Parthenope"

General Chairs

Claudio A. Ardagna, Università degli Studi di Milano
Michelangelo Ceci, Università degli Studi di Bari Aldo Moro
Donato Malerba, Università degli Studi di Bari Aldo Moro

Program Chairs

Nicola Bena, Università degli Studi di Milano
Salvatore Distefano, Università degli Studi di Messina
Antonella Longo, Università degli Studi del Salento
Gianvito Pio, Università degli Studi di Bari Aldo Moro

Program Chairs – Industrial Track

Barbara Alicino, Lutech
Marco Anisetti, Università degli Studi di Milano
Davide Dalle Carbonare, Engineering
Pietro Leo, IBM Italia

Demo/Tutorial Chairs

Claudia Diamantini, Università Politecnica delle Marche
Davide Martinenghi, Politecnico di Milano
Vincenzo Pasquadibisceglie, Università degli Studi di Bari Aldo Moro

Workshop Chair

Giuseppina Andresini, Università degli Studi di Bari Aldo Moro

Website Co-Chairs

Antonio Pellicani, Università degli Studi di Bari Aldo Moro
Marco Luzzara, Università degli Studi di Milano

Publication Chairs

Alex Della Bruna, Università degli Studi di Milano
Aneela Nasim, Università degli Studi di Milano
Paolo Mignone, Università Telematica Pegaso

Publicity Chair

Aneela Nasim, Università degli Studi di Milano

Hackathon Chairs

Fabrizio Montecchiani, Università degli Studi di Perugia (Hackathon Chair & Lab Hackathon and Challenge Delegate)
Antonio Maratea, Università degli Studi di Napoli Parthenope (Hackathon Chair)
Luca Amoroso, Università degli Studi di Napoli Parthenope (Challenge Design Chair)
Filippo Berto, CERN (Technical Infrastructure Chair)
Marco Luzzara, Università degli Studi di Milano (Technical Infrastructure Co-Chair)

Program Committee

Giuseppina Andresini, Università degli Studi di Bari Aldo Moro
Valerio Bellandi, Università degli Studi di Milano
Nadia Bennani, INSA Lyon
Alessio Cascione, *Università di Pisa*
Michelangelo Ceci, Università degli Studi di Bari Aldo Moro
Stefano Cirillo, *Università degli Studi di Salerno*
Alessia Citarella Auriemma, *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*
Andrea D'Angelo, Università degli Studi dell'Aquila
Simon Dahdal, *Università degli Studi di Ferrara*
Fabiola De Marco, *Università degli Studi di Salerno*
Alex Della Bruna, Università degli Studi di Milano
Luigi Di Biasi, *Università degli Studi di Salerno*
Edoardo Di Caro, *Università degli Studi di Ferrara*
Emanuel Di Nardo, *Università degli Studi di Napoli "Parthenope"*
Roberto Esposito, Università degli Studi di Torino
Sandro Fiore, *Università degli Studi di Trento*
Paolo Frazzetto, Universitat de València
Massimiliano Garda, *Università degli Studi di Brescia*
Chirine Ghedira-Guegan, University Lyon 3
Daniele Granata, Università degli Studi di Napoli Parthenope
Andrea Loddo, Università degli Studi di Cagliari
Marco Luzzara, Università degli Studi di Milano
Samira Maghool, Università degli Studi di Milano
Antonio Maratea, Università degli Studi di Napoli Parthenope
Andrea Mastropietro, Lamarr Institute, Nara Institute of Science and Technology
Fabrizio Montecchiani, Università degli Studi di Perugia
Vincenzo Moscato, Università degli Studi di Napoli "Federico II"
Nicolò Navarin, Università degli Studi di Pavia
Serena Nicolazzo, Università degli Studi del Piemonte Orientale "Amedeo Avogadro"
Antonino Nocera, Università di Padova
Antonio Pellicani, Università degli Studi di Bari Aldo Moro
Domenico Potena, Università Politecnica delle Marche
Chiara Rucco, Università degli Studi del Salento
Motaz Saad, Università degli Studi del Salento
Anna Sapienza, Università degli Studi del Piemonte Orientale "Amedeo Avogadro"
Giandomenico Solimando, Università degli Studi di Salerno
Domenico Talia, Università degli Studi della Calabria
Orazio Tomarchio, Università di Catania
Riccardo Torlone, Università degli Studi Roma Tre
Genoveva Vargas-Solar, CNRS
Gennaro Zanfardino, Università degli Studi dell'Aquila
Mansur Ziiatdinov, Università degli Studi di Messina

Logistics & Social Events

ITADATA2026 is held at the Dipartimento di Informatica, Università degli Studi di Bari Aldo Moro, Campus Universitario "Ernesto Quagliariello," Via Edoardo Orabona 4, 70125 Bari, Italy. The conference will take place in **Aula Magna, Aula A and Aula B** (ground floor). The registration desk will be located at the **main entrance** of the Department, while coffee breaks and lunches will be held at **Aula C** (ground floor).



Social Event

The ITADATA 2026 Social Event will take place on **10th November starting at 17:30** with the Welcome Cocktail Reception in **Aula C, Department of Computer Science**.

Guided Tour

The tour will start at **18:30** with a guided walking tour of **Bari's Old Town (Bari Vecchia)**, allowing participants to discover its historic streets, landmarks, and traditions. The tour will last until **20:00**.

Social Dinner

The Social dinner will take place on **11th November starting at 20:30** at **JR Hotels Grande Albergo delle Nazioni (Lungomare Nazario Sauro 7, Bari)**, overlooking Bari's seafront.

Program Overview

Legend				
Plenary	ITADATA	Hackathon	Tutorial	Workshop

Day 1 - November 10, 2026

<table>
<tr><th>Time</th><th>Aula Magna</th><th>Aula A</th><th>Aula B</th></tr>
<tr><td>9:00-10:45</td><td>Hackathon</td><td>No Event</td><td>Workshop “Scientific HPC in the pre-Exascale era”</td></tr>
<tr><td>10:30-11:00</td><td colspan="3">Welcome Coffee</td></tr>
<tr><td>11:00-11:15</td><td>Greetings</td><td>No Event</td><td rowspan="3">Workshop “Scientific HPC in the pre-Exascale era”</td></tr>
<tr><td>11:15-12:00</td><td>Benchmarking and Data Governance</td><td rowspan="2">Tutorial “CSMining: Mining and Pooling Synthetic Cognitive Data (SCD) via LLMs”</td></tr>
<tr><td>12:00-13:00</td><td>Big Data and Data Science Applications</td></tr>
<tr><td>13:00-14:00</td><td colspan="3">Lunch</td></tr>
<tr><td>14:00-15:00</td><td>Security, Privacy, and Beyond I</td><td rowspan="2">Tutorial “Machine Learning Models for Mobility and Transportation Dynamics”</td><td rowspan="4">No Event</td></tr>
<tr><td>15:00-15:30</td><td rowspan="2">Distributed Processing and Workload Management</td></tr>
<tr><td>15:30-15:55</td><td rowspan="2">Interacting with LLMs and More</td></tr>
<tr><td>15:55-16:00</td><td>No Event</td></tr>
<tr><td>16:00-16:30</td><td colspan="3">Coffee Break</td></tr>
<tr><td>16:30-17:00</td><td>Award</td><td rowspan="2">No Event</td><td rowspan="2">No Event</td></tr>
<tr><td>17:00-17:15</td><td>Hackathon</td></tr>
<tr><td>17:30-18:30</td><td colspan="3">Welcome Cocktail</td></tr>
</table>

Day 2 - November 11, 2026

<table>
<tr><th>Time</th><th>Aula Magna</th><th>Aula A</th><th>Aula B</th></tr>
<tr><td>9:00-10:00</td><td>Keynote Speech: Richard Lloyd Stevens</td><td colspan="2" rowspan="2">No Event</td></tr>
<tr><td>10:00-11:00</td><td>Panel: “Data-Driven AI: Understanding Risks, Trust, and the Future of Intelligent Systems”</td></tr>
<tr><td>11:00-11:30</td><td colspan="3">Coffee Break</td></tr>
<tr><td>11:30-12:15</td><td rowspan="2">Workshop “Building Resilient Data Ecosystems”</td><td>Database and Querying</td><td rowspan="2">Tree-Based Approaches, Online Learning, and More</td></tr>
<tr><td>12:15-12:45</td><td>Security, Privacy, and Beyond II</td></tr>
<tr><td>12:45-13:45</td><td colspan="3">Lunch</td></tr>
<tr><td>13:45-14:45</td><td rowspan="2">Workshop “Building Resilient Data Ecosystems”</td><td>Tutorial “Adversarial Machine Learning in Large Language Models”</td><td>No Event</td></tr>
<tr><td>14:45-15:30</td><td>Federated Learning and Anomaly Detection</td><td>Big Data and Data Science in Urban Applications, and Clustering</td></tr>
<tr><td>15:30-15:45</td><td colspan="3">Coffee Break</td></tr>
<tr><td>15:45-16:30</td><td>Panel “Building Trustworthy and Resilient Data Ecosystems for European Society”</td><td colspan="2">No Event</td></tr>
<tr><td>18:30-20:00</td><td colspan="3">Guided Tour of Bari’s Old Town</td></tr>
<tr><td>20:30</td><td colspan="3">Gala Dinner</td></tr>
</table>

Day 3 - November 12, 2026

<table>
<tr><th>Time</th><th>Aula Magna</th><th>Aula A</th><th>Aula B</th></tr>
<tr><td>9:30-10:30</td><td>Keynote Speech: Ernesto Damiani</td><td colspan="2">No Event</td></tr>
<tr><td>10:30-11:00</td><td colspan="3">Coffee Break</td></tr>
<tr><td>11:00-12:00</td><td>Industry I</td><td rowspan="2">Tutorial “YSocial: Building, Running, and Analyzing LLM-Powered Social Media Digital Twins”</td><td>Knowledge Augmentation and Ontologies</td></tr>
<tr><td>12:00-13:00</td><td>Industry II</td><td>Workshop “Risks and Unintended Harms of Generative AI Systems”</td></tr>
<tr><td>13:00-14:00</td><td colspan="3">Lunch</td></tr>
<tr><td>14:00-14:35</td><td>RAG</td><td rowspan="2">Tutorial “Uncertainty and Learning from Data: Statistical Perspectives”</td><td rowspan="4">Workshop “Risks and Unintended Harms of Generative AI Systems”</td></tr>
<tr><td>14:35-15:00</td><td rowspan="2">Healthcare and Data Imputation</td></tr>
<tr><td>15:00-15:15</td><td rowspan="2">No Event</td></tr>
<tr><td>15:15-15:30</td><td>No Event</td></tr>
</table>

Program Details

Benchmarking and Data Governance

November 10, 11:15–12:00
Session chair: Giancarlo Ruffo, *Università degli Studi del Piemonte Orientale "Amedeo Avogadro"*
Room: Aula Magna

Benchmarking Embedding Models for ESG Data
Motaz Saad, Veronica Creti, Ivan Gentile, Kianna Kazemi, Antonella Longo

Auditing and Guardrailing Large Language Models for Ethical Issues and Bias
Chiara Criscuolo, Mathyas Giudici, Marco Brambilla, Davide Martinenghi

Data Governance Project Valuation: a risk-based model for the economic valuation of data governance initiatives
Mauro Tuvo

Big Data and Data Science Applications

November 10, 12:00–13:00
Session chair: Antonio Pellicani, *Università degli Studi di Bari "Aldo Moro"*
Room: Aula Magna

How Sensitive Are Clustering Models to Data Errors? An Error Sensitivity Profile Approach
Andrea Maurino, Fariha Maqbool

From waste to resource: How the VALOR project uses data management and AI to recover critical raw materials
Stefano Rovetta, Muhammad Mohsin, Rosario Capponi, Gianluca Dini

Context-Aware Explainable Artificial Intelligence for Multimodal Marine Environmental Monitoring
Pietro Patrizio Ciro Aucelli, Francesco Camastra, Angelo Ciaramella, Emanuel Di Nardo, Alessio Ferone, Antonio Maratea, Antonino Staiano

A Self-Hosted Modular RAG Architecture for Academic Student Support: Design, Deployment, and Evaluation
Giulia Giglioni, Michele Mariotti, Gabriele Costante, Fabrizio Montecchiani

Security, Privacy, and Beyond I

November 10, 14:00–15:00
Session chair: Fabiola De Marco, *Università degli Studi di Salerno*
Room: Aula Magna

Security Assurance for Modern Big Data APIs
Marco Anisetti, Enea Manzi

The Impact of the Cyber Resilience Act on IoT Devices
Martin Jaatun

Distributed Quantum Redundancy: A Framework for Delayed Selective Redemption of Encrypted Clones
Gabriele Gianini, Claudio Ardagna, Marco Anisetti, Paolo Calvi, Nicolò Tosin, Stelvio Cimato, Ernesto Damiani

Targeted Augmentation for Intent Routing in Joint Search and Recommendation
Ajla Rustemi, Cristina Muntean, Raffaele Perego

Distributed Processing and Workload Management

November 10, 15:00–15:55
Session chair: Michelangelo Ceci, *Università degli Studi di Bari "Aldo Moro"*
Room: Aula Magna

A Neurosymbolic Residual-based Model for GPU Demand Forecasting
Lorenzo Colombi, Francesco Resca, Simon Dahdal, Elena Bellodi, Mauro Tortonesi, Cesare Stefanelli

From Data to Praxis: Fragility, Resilience, and Antifragility in Data-Intensive Systems
Salvatore Distefano

An Observability-Driven Architecture with a Two-Stage Decision Process for Predictive Autoscaling in Kubernetes
Angelo Marchese, Francesco Romeo, Orazio Tomarchio

Reversibility in Knowledge Model Editing: A First Taxonomy
Manuela Sanguinetti, Maurizio Atzori

Interacting with LLMs and More

November 10, 15:30–16:00
Session chair: Salvatore Contino, *Università degli Studi di Palermo*
Room: Aula A

Data-Driven Resilience
Ali Aghazadeh Ardebili, Massimo Stella

Mind Reading or Misreading? LLMs on the Big Five Personality Test
Francesco Di Cursi, Chiara Boldrini, Andrea Passarella, Marco Conti

Database and Querying

November 11, 11:30–12:15
Session chair: Fabrizio Montecchiani, *Università degli Studi di Perugia*
Room: Aula A

A Computation-Friendly, GQL-compliant Compressed Static Graph Database
Lorenzo Bellomo, Giuseppe Cianci, Luca de Rosa, Paolo Ferragina, Michelangelo Pucci

Adaptive-Elbow Graph Traversal for Schema Retrieval in Text-to-SQL
Fabio Picariello, Salvatore Cipolla, Susanna Bonura, Vito Morreale

GRACE: Parallel Best-of-N Multi-Agent Analysis of Closed-Form Questions over Dataframes with Guardrail Selection and Adaptive Retry
Fabio Picariello, Salvatore Cipolla, Susanna Bonura, Vito Morreale

Tree-Based Approaches, Online Learning, and More

November 11, 11:30–12:45
Session chair: Gianvito Pio, *Università degli Studi di Bari "Aldo Moro"*
Room: Aula B

A Semantic-Driven Retrieval-Augmented Generation Framework for Technical Document Assistance in Airport Ground Operations
Aiman Al Masoud, Marco Arazzi, Davide Gabrielli, Antonino Nocera, Matteo Zucchetti

Explainable-by-Design Personal Protective Equipment Recognition with Prototype-Based Decision Trees
Giustino Claudio Miglionico, Fabrizio Ruffini, Pietro Ducange, Carlo Vallati, Francesco Marcelloni

Behaviour Cloning for Online Learning of Driver Habits
Antonio Daniele Ferraro, Massimo Cafaro, Italo Epicoco

Pseudo-Trees for Top-k Simple Shortest Paths: Construction and Structural Properties
Gianlorenzo D'Angelo, Andrea D'Ascenzo, Mattia D'Emidio, Gabriele Di Stefano, Marta Innamorati

PATH4U: A Knowledge-Driven Expert System for Context-Aware Tourist Itinerary Recommendations
Mariangela Graziano, Luigi Colucci Cante, Beniamino di Martino

Security, Privacy, and Beyond II

November 11, 12:15–12:45
Session chair: Devis Bianchini, *Università degli Studi di Brescia*
Room: Aula A

Bridging Model Explainability and Data Provenance for Interactive Pipeline Auditing
Pasquale Leonardo Lazzaro, Paolo Missier, Riccardo Torlone

Integrity of peer-to-peer distributed LLM inference under malicious nodes
Mert Cihangiroglu, Antonino Nocera

Context-Aware Software Vulnerability Prioritisation: A Case Study
Steve Taylor, Panos Melas, Mike Surridge, Manuel Leone, Paolo De Lutiis, Ravishankar Borgaonkar, Martin Jaatun

Federated Learning and Anomaly Detection

November 11, 14:45–15:30
Session chair: Lorenzo Colombi, *Università degli Studi di Ferrara*
Room: Aula A

QUASAR-FL: A Quality-Aware Framework for Secure Federated Learning via Homomorphic Encryption
Raffaele Aurucci, Stefano Cirillo, Giuseppe Polese, Giandomenico Solimando

Federated Digital Twins on Data Spaces: An Architectural Framework
Ferdinando Bosco, Vincenzo Croce, Marcantonio La Franca

Experimenting with Federated Learning for Multivariate Time Series Classification
Bruno Casella, Lorenzo Sciandra

BACON in TELEMETRY context
Rosella Omana Mancilla

Big Data and Data Science in Urban Applications, and Clustering

November 11, 14:45–15:30
Session chair: Gabriele Gianini, *Università degli Studi di Milano-Bicocca*
Room: Aula B

Personalizing Hyper-Proximity: A User-Centric Web Platform for Customized Urban Accessibility Metrics
Giancarlo Ruffo, Mirko Lai, Anna Sapienza, Federica Cena

From Urban Big Data to Generative Digital Twins: Research Activities at the M.O.D.A.L. Lab
Marzia Canzaniello, Sara Amitrano, Stefano Izzo, Daniela Annunziata, Fabio Giampaolo, Francesco Piccialli

Large-Scale Interpolation of Georeferenced Urban Sensing Data
Carmine Colarusso, Ida Falco, Eugenio Zimeo

ONSET: Online Causal Detection of Urban Events from Mobile-Subscriber Traffic
Luca Fantin, Chiara Pugliese, Fabio Pinelli, Bruno Zamengo

Industry I

November 12, 11:00–12:00
Session chair: Davide Dalle Carbonare, *Engineering*
Room: Aula Magna

DT4Energy: Open AI-Driven Digital Twins for Energy Systems (DT4EC Use Case)
Ferdinando Bosco, Marcantonio La Franca, Manlio Serio, Luigi Forioso, Tania Sari Bonaventura, Pasquale Leone, Michele Fuschillo, Lorenzo Cristofori, Angelo Triveri, Vincenzo Croce

Agentic AI for ERP Tender Analysis: Design and Industrial Validation of a Multi-Agent Assistant in the Microsoft Ecosystem
Daniele Angeloni, Fabrizio Montecchiani, Daniele Selvi

DocFlow AI: a Secure, Model-Agnostic Platform for Next-Gen Industrial Document Processing
Riccardo Martoglia, Piero De Giacomo, Mirko Orsini, Sonia Bergamaschi

Empowering building energy management systems with foundational time-series models
Angelo Impedovo

Knowledge Augmentation and Ontologies

November 12, 11:00–12:00
Session chair: Giandomenico Solimando, *Università degli Studi di Salerno*
Room: Aula B

Towards LLM-Driven Knowledge Graph Construction for the Semantic Representation of Formulas
Claudia Diamantini, Alex Mircoli, Domenico Potena, Cristina Rossetti, Emanuele Storti

Combining Ontology-Guided Retrieval and Fine-Tuned LLMs for Practical Experiment Assistance in Scientific Domains
Celeste Sguera, Antonio Pellicani, Michelangelo Ceci

Ontology-Enhanced Retrieval-Augmented Generation for Knowledge-Intensive AI Systems
Ada Bagozi, Devis Bianchini, Massimiliano Garda, Michele Melchiori, Stefano Picchi, Anisa Rula

A Knowledge-Guided Architecture for Structuring and Exploring Conversational Data in Assisted Care
Massimo Callisto De Donato, Flavio Corradini, Flavio Gerosa, Francesco Finucci, Niccolò Vito Vacca

Industry II

November 12, 12:00–13:00
Session chair: Barbara Alicino, *Lutech*
Room: Aula Magna

From heterogeneous Open Data to decision support: a modular Platform for smart rural area management
Francesca D'Agresti, Antonio Filograna, Alessio Carachino, Tjaša Heričko, Muhamed Turkanović, Elena Not, Eleonora Mencarini

Near-Real-Time AI Assistance for Swiss Ambulance Emergency-Call Handling: An Industrial Pilot
Federico Frati, Alessio Schiavo, Alessio Bechini, Pietro Ducange, Francesco Marcelloni

On-Premises Natural Language Access to Incident Knowledge Graphs: An Industrial Experience Report
Marco Palumbo, Alessio Schiavo, Pietro Ducange, Francesco Marcelloni, Nicola Tonellotto, Andrea La Mantia

RAG

November 12, 14:00–14:35
Session chair: Chiara Boldrini, *CNR - IIT*
Room: Aula Magna

Language-Aware Multilingual Retrieval-Augmented Generation via Local-to-Global Retrieval
Gianluca Bonifazi, Christopher Buratti, Stefano Cirillo, Michele Marchetti, Federica Parlapiano, Giulia Quaglieri, Giandomenico Solimando, Davide Traini, Domenico Ursino, Luca Virgili

Merging LLMs, Geospatial data and RAG
Tonio De Paolis, Massimo Cafaro, Italo Epicoco

MIRAGE-GNN: Multimodal Imaging Representation And Graph Embedding
Salvatore Contino, Luca Cruciata, Dario Morreale, Roberto Pirrone

Healthcare and Data Imputation

November 12, 14:35–15:15
Session chair: Corrado Loglisci, *Università degli Studi di Bari "Aldo Moro"*
Room: Aula Magna

Text2Triage: Fine-Tuning Clinical NER Models on Synthetic Emergency Speech Data
Mariano Barone, Francesco Di Serio, Vincenzo Moscato, Gian Marco Orlando, Marco Perillo, Marco Postiglione, Giuseppe Riccio, Antonio Romano, Diego Russo

From Prediction to Event-Centered Discovery: Longitudinal Complement Proteomics in Islet Autoimmunity
Domenico Rossi

On the Transformer-based Data Imputation in (Dynamic) Relational Databases
Gianvincenzo Alfano, Sergio Greco, Lucio La Cava, Tariq Mahmood, Irina Trubitsyna

Keynote Speeches

Richard Lloyd Stevens: From Data Spaces to AI Factories: Data Labs and Europe's Green Deal Data Space

November 11, 09:00–10:00
Room: Aula Magna

Abstract

Europe is assembling the infrastructure layer of its AI ambitions, and the pieces are beginning to connect. Common European Data Spaces supply operational data sharing under agreed rules; EuroHPC AI Factories, nineteen of them with thirteen antennas, supply compute, platforms and expertise. Between the two sits a layer only recently named: Data Labs, introduced in the Data Union Strategy of November 2025 as the place where data is pooled, filtered, labelled, governed and made legally usable, with the first Data Labs to be established within the AI Factories themselves. This talk examines that emerging three-layer architecture from the perspective of the Big Data Value Association and the discussions at the European Big Data Value Forum 2026 in Galway, whose theme, building Europe's AI and data backbone, captures the shift from policy design to delivery and investment. It then grounds the architecture in a concrete case: SAGE, the project building the Green Deal Data Space. Funded under the Digital Europe Programme with a budget of some EUR 16 million and a consortium of more than forty partners coordinated by IDC Italia, SAGE federates fragmented environmental data across biodiversity, climate adaptation, circular economy and zero pollution, validating it through ten pilot use cases and establishing the operating entity intended to sustain the data space beyond the project's end in 2028. The talk closes on what remains unresolved: semantic interoperability at scale, incentives for sharing restricted data, and whether Europe's data layer will be ready when the compute is.

Bio

Richard Lloyd Stevens is Associate Vice President and Director of Government Consulting at IDC Italia, where he leads IDC's European Government Consulting practice. Based in Milan, Italy, he advises European institutions and national administrations on data spaces, federated data architectures, generative and agentic AI, and the digital infrastructure underpinning the European Green Deal. He works across a portfolio of Horizon Europe and Digital Europe projects, including SAGE, the project building the Green Deal Data Space, which IDC Italia coordinates. He has served continuously as an expert evaluator for the European Commission since 2014, across multiple Directorates-General and agencies, and takes part in the Big Data Value Association community.



Ernesto Damiani: The Epistemological Ladder: Descent for Scale, Ascent for Understanding in Data Science

November 12, 09:30–10:30
Room: Aula Magna

Abstract

Foundation models and large language models are entering data science workflows at speed, delivering scale and performance that often outstrip human experts. Yet not all inferences carry the same epistemic value. We formalize this stratification as a three-rung Epistemological Ladder: theory-driven, falsifiable reasoning at the top; interpretable machine learning on engineered features in the middle; and opaque black-box foundation models at the bottom. Descent for scale or efficiency is frequently the right decision; the danger lies in losing the instruments, institutions, and habits needed to climb back for understanding. This keynote presents seven criteria for epistemic ascent—provenance, external validation, intrinsic traceability, uncertainty quantification on the analyzed sample, robustness to analytic choices, logged human veto, and a traceable link to a falsifiable claim—and illustrates them with complementary examples from a published re-analysis of a terminated Alzheimer's trial and a calibrated synthetic simulation. Both achieve only partial ascent and fail in different ways: simulation supplies calibration unavailable from single trials; real-world re-analysis supplies provenance and cohort grounding. For data scientists, the practical implications are immediate: declare the rung of a claim, report the analyzed rather than the recruited sample, characterize discovered strata in the input variables, and prefer models whose interpretability is intrinsic rather than post-hoc. Beyond methodology, the talk reflects on a bigger risk: the quiet atrophy of the capacity to ask why—a loss that cannot be fully captured by papers alone, but must be felt if we are to guard the fragile, correctable reasons that constitute human knowledge.

Bio



Ernesto Damiani is Full Professor of Computer Science at the University of Milan, where he leads the SESAR Lab. He served as Senior Director of the Artificial Intelligence and Intelligent Systems Institute at Khalifa University, UAE, and as President of the Italian Inter-University Consortium for Informatics (CINI). His research spans artificial intelligence, machine learning, Big Data analytics, knowledge processing, and cybersecurity. Author of over 880 Scopus publications, he is an ACM Distinguished Scientist. Ernesto has received many awards, including an honorary doctorate from INSA Lyon and the 2025 IEEE TCCH Research and Innovation Award for contributions at the intersection of computer science and the humanities. In 2022, President Sergio Mattarella awarded him the rank of Officer of the Order of the Star of Italy for his role in international scientific collaboration.

Workshops

ITADATA2026 features an exciting program of 3 workshops. Detailed information can be found at `https://www.itadata.it/2026/workshops`.

Scientific HPC in the pre-Exascale era (3rd edition)

Workshop chairs: Valentina Cesare, *INAF*, Gianluca Mittone, *Università degli Studi di Torino*, Bruno Casella, *Università degli Studi di Torino*, Giovanni Naldi, *INAF*

10 November, 9:00–13:00

Room: Aula B

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.

Building Resilient Data Ecosystems: Insights from the PATH, ENDURANCE, AVANT, Connect 2 the Future and Calliope European Projects

Workshop chairs: Luigi Romano, *Università degli Studi di Napoli "Parthenope"*

11 November, 11:30–15:30

Room: Aula Magna

The increasing availability of large-scale data and advanced Artificial Intelligence technologies is transforming the way critical societal systems are designed, monitored, and managed. From healthcare pathways to essential public services, data-driven approaches are becoming key enablers of resilience, adaptability, and informed decision-making. However, realizing this potential requires trustworthy AI solutions, interoperable data ecosystems, secure information sharing mechanisms, and governance frameworks capable of ensuring transparency, privacy, and sustainability.

This workshop brings together researchers, practitioners, and stakeholders involved in the European projects PATH, ENDURANCE, AVANT, Connect 2 the Future and Calliope to discuss emerging approaches for building resilient data ecosystems. PATH focuses on the adoption of trustworthy AI technologies and interoperable health data infrastructures to support cancer care pathways and improve clinical decision-making. ENDURANCE addresses the resilience of critical infrastructures and essential services through advanced data management, risk assessment methodologies, digital twins, and AI-enabled decision support systems. AVANT provides an energy-domain perspective through DT4Energy, an open and modular platform for secure and interoperable Digital Twins, and DT4EC, its operational implementation for Energy Communities. The presentation will highlight data interoperability and the role of Data Spaces in integrating heterogeneous and distributed energy data, models and services across Digital Twin environments. Connect 2 the Future presents Fincantieri's study of a Federated Maritime Data Space, conceived as a trusted digital ecosystem for secure, sovereign, governed and interoperable data sharing among key stakeholders of the maritime sector, enabling the transition from fragmented point-to-point exchanges to reusable data products and scalable ecosystem collaboration. CALLIOPE addresses advanced research, monitoring and prevention across human, animal and environmental health, while fostering technology transfer, training and collaboration with industry. To achieve this, CALLIOPE provides a digital innovation ecosystem for One Health, integrating AI, IoT, 5G and data-driven technologies.

Through project presentations, technical discussions, and cross-domain perspectives, the workshop will explore how data spaces, AI technologies, and digital platforms can enhance resilience across different application domains. Particular attention will be devoted to interoperability, trustworthiness, governance, privacy-preserving analytics, and the role of data science in supporting complex decision-making processes.

The workshop contributes to the objectives of ITADATA 2026 by fostering dialogue between data scientists, AI researchers, healthcare experts, infrastructure operators, and policy stakeholders working towards trustworthy and resilient data-driven systems.

Risks and Unintended Harms of Generative AI Systems

Workshop organizers: Stefano Cirillo, *Università degli Studi di Salerno*, Eliana Pastor, *Politecnico di Torino*, Francesco Pierri, *Politecnico di Milano*, Serena Tardelli, *IIT-CNR*, Angelo Impedovo, *AI & ML specialist at Nesea, Bari*, Vincenzo Pasquadibisceglie, *Università degli Studi di Bari Aldo Moro*

12 November, 12:00–15:30

Room: Aula B

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

ITADATA2026 features two plenary panels with distinguished panelists. Detailed information can be found at `https://www.itadata.it/2026/program`.

Data-Driven AI: Understanding Risks, Trust, and the Future of Intelligent Systems

Panel chair: Donato Malerba, *Università degli Studi di Bari Aldo Moro*
Panelists: Annalisa Appice, *Università degli Studi di Bari Aldo Moro*, Luigi Romano, *Università degli Studi di Napoli "Parthenope"*, Barbara Alicino, *Lutech S.p.A., Davide Dalle Carbonare*, *Engineering S.p.A., Richard Lloyd Stevens*, *IDC Italia; Steering Board member, Big Data Value Association (BDVA)*
11 November, 10:00–11:00
Room: Aula Magna

Data-driven AI is transforming how intelligent systems are designed, deployed, and used, reshaping research, industry, and society. As these systems increasingly inform decisions and operate in complex environments, their dependence on data introduces risks that extend beyond predictive accuracy. Data quality, changing operating conditions, and the interaction between AI and other system components raise questions about reliability, transparency, explainability, and responsibility. Understanding these risks and establishing trust are essential to responsible adoption.

This panel brings together diverse perspectives to discuss security, privacy, fairness, robustness, and accountability of data-driven AI, and to examine how trust in intelligent systems can be assessed and sustained throughout their lifecycle. The discussion will explore emerging challenges and opportunities, considering how research, technological innovation, and governance can shape the future of reliable and responsible AI.

Building Trustworthy and Resilient Data Ecosystems for European Society

Panel chair: Luigi Romano, *Università degli Studi di Napoli "Parthenope"*
11 November, 15:45 - 16:30
Room: Aula Magna

How can Europe turn fragmented data infrastructures into trustworthy ecosystems that strengthen societal resilience? This panel synthesizes the approaches and cross-domain lessons emerging from the workshop "Building Resilient Data Ecosystems," connecting perspectives from healthcare, critical infrastructures, energy, and maritime services. The discussion spans five complementary application areas: trustworthy AI and interoperable health data infrastructures for cancer care; risk assessment and decision support for critical infrastructures and essential services; federated Digital Twins for energy systems and communities; sovereign data sharing across the maritime value chain; and integrated human, animal and environmental health through digital innovation ecosystems. Across these domains, a common insight emerges: resilience depends not only on advanced analytics, but also on the ability to make heterogeneous data, models and services interoperable within accountable governance frameworks. Data Spaces, privacy-preserving approaches and Digital Twins offer complementary foundations for collaboration, while explainability, data sovereignty and integration into operational workflows are essential conditions for adoption. Building on this synthesis, the panel examines shared challenges in balancing interoperability with local control, translating AI outputs into actionable decisions, and sustaining collaboration among research institutions, industry, public authorities and healthcare stakeholders. Particular attention is given to how technical architectures and governance arrangements can jointly support preparedness, coordinated responses and adaptation. The panel brings these perspectives together to identify common research priorities, transferable practices and opportunities for cross-domain collaboration, outlining an agenda for trustworthy, sustainable and resilient data ecosystems across Europe.

Demos/Tutorials

ITADATA2026 features an exciting program of 5 tutorials. Detailed information can be found at `https://www.itadata.it/2026/program`.

CSMining: Mining and Pooling Synthetic Cognitive Data (SCD) via LLMs

Presenters: Massimo Stella, *Università degli Studi di Trento*, Ali Aghazadeh Ardebili, *Università degli Studi di Trento*
10 November, 11:15–13:00
Room: Aula A

This tutorial proposal is a reproducible data-science perspective on LLM-generated cognitive traces. Instead of treating synthetic responses as isolated anecdotes, we show how to mine these traces as large-scale, metadata-rich datasets; how to pool them into analytically meaningful groups; and how to visualise and interpret them with caution. This makes the tutorial relevant to big-data processing, data governance, trustworthy AI, policy-sensitive analytics, and vertical applications of data science in education, mental health, and social discourse.

Machine Learning Models for Mobility and Transportation Dynamics

Presenter: Riccardo Cappi, *Università di Padova*
10 November, 14:00–15:30
Room: Aula A

With the rapid development of smart city infrastructures and the growing tendency of municipalities to install sensors across the urban environment, Machine Learning (ML) models are rapidly becoming a central tool for understanding and predicting human mobility and transportation dynamics. In this tutorial, we walk through state-of-the-art ML models to address three mobility-related core tasks, highlighting how the recent paradigm of physics-informed learning can add value in terms of both predictive accuracy and interpretability. We complement the lecture with a live coding session built on a newly released city-scale dataset derived from traffic sensors in Padua, Italy, demonstrating how to build static and dynamic spatio-temporal graphs and how to correctly train and evaluate a baseline forecaster.

Adversarial Machine Learning in Large Language Models

Presenter: Giuseppina Andresini, *Università degli Studi di Bari Aldo Moro*
11 November, 13:45–14:45
Room: Aula A

Large Language Models (LLMs) have recently demonstrated significant potential across several application domains. On the other hand, as with all AI-based models, LLMs are vulnerable to adversarial attacks, which involve deliberately manipulating input data to achieve malicious goals. The tutorial will provide an overview of how adversarial learning algorithms are categorized by goal, timing, and knowledge. The tutorial will also provide an overview of several state-of-the-art adversarial algorithms, ranging from pixel manipulations to prompt manipulations.

YSocial: Building, Running, and Analyzing LLM-Powered Social Media Digital Twins

Presenters: Giulio Rossetti, *ISTI-CNR*, Roberto Trasarti, *CNR*
12 November, 11:00–13:00
Room: Aula A

This tutorial introduces YSocial, an open-source LLM-powered social media digital twin for designing, running, and analyzing controlled simulations of online social dynamics, recommender effects, misinformation, and agent interactions. Participants will learn how to configure scenarios, run simulations, analyze generated data with ySights, and explore the creation of reusable YSocial datasets through the SoBigData infrastructure.

Uncertainty and Learning from Data: Statistical Perspectives

Presenters: Raffaele Mattera, *Università degli Studi della Campania "Luigi Vanvitelli"*, Giancarlo Manzi, *Sapienza Università di Roma*, Elena Siletti, *Università degli Studi di Torino*, Jacopo Chiapparino
12 November, 14:00–15:00
Room: Aula A

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 multi-source 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:

- Raffaele Mattera, Rosanna Verde, “Learning from interval-valued data: recent perspectives”
- Francesca De Battisti, Marcella Lorenzini, Elena Siletti, “From Historical Records to Research Data: Designing a Reproducible Workflow for Early Modern Notarial Sources”
- Francesca De Battisti, Giancarlo Manzi, “En attendant Godot: A dataset for modeling waiting times for academic publications”
- Jacopo Chiapparino, Alessandro Bitetto, Paola Cerchiello, “Toward a Dynamic Data-Driven Environmental Index: Extending DDEPI via DFM and DeepAutoencoders”