Research Project Exhibition Track SEBD 2025

Francesca De Luzi[1], Flavia Monti[1] and Massimo Mecella[1,*]

[1]Department of Computer, Control and Management Engineering, Sapienza Università di Roma, Rome, Italy

This contribution contains the accepted papers of the Research Projects Exhibition, held in conjunction with the 33rd Symposium on Advanced Database Systems (SEBD 2025). This edition of SEBD took place in Ischia (Italy) between June 16th and 19th. We were delighted to contribute to this year's symposium with the first edition of the Research Projects Exhibition (RPE@SEBD'25).

The SEBD conference is renowned as the premier Italian venue for presenting innovative and rigorous research across the broad spectrum of advanced database systems. In line with this tradition, the Research Projects Exhibition was specifically organized to provide a dedicated track to showcase ongoing research projects (e.g., European PNRR or Horizon Europe projects, national or regional initiatives, or other funded research) in the context of advanced database systems and their applications.

The main objective of this initiative was to create a forum where authors could disseminate intermediate results, present the objectives and achievements of their projects, and receive constructive feedback on project proposals under development. The exhibition also offered a friendly environment for finding potential research partners, strengthening existing collaborations, and stimulating new ideas.

In this first edition of RPE@SEBD'25, we accepted 7 posters that were presented in a dedicated face-to-face session running throughout the symposium. The list of accepted papers is provided below, and each of them is presented in the following sections.

- AI-Powered Industrial Anomaly Detection: A Dual Approach with LLMs and Machine Learning, *Ala Arman, Filippo Bianchini, Marco Calamo, Loredana Cristaldi, Emilia Lenzi, Matteo Marinacci, Davide Martinenghi, Luca Martiri, Massimo Mecella, Andrea Moschetti, Jacopo Rossi, Letizia Tanca* (see Section 1)
- An Ontology-based Multidimensional Data Modeling, *Domenico Lembo, Maurizio Lenzerini, Antonella Poggi, Federico Maria Scafoglieri, Jacopo Brunetti, Roberta Radini, Michele Riccio, Valerio Santarelli* (see Section 2)
- Building National Data lakehouse Ecosystems for Environmental and Public Health: AnTeA and IDEAH, *Mario Cerroni, Francesca De Luzi, Tommaso Filippini, Valentina Fuscoletti, Marco Giustini, Raffaele Landi, Francesco Leotta, Luca Lucentini, Mattia Macrì, Camilla Marchiafava, Marco Marras, Daniela Mattei, Giampaolo Maugeri, Massimo Mecella, Alessio Pitidis, Marco Vinceti* (see Section 3)
- HEREDITARY: HetERogeneous sEmantic Data Integration for the guT-brAin inteRplaY, *Gianmaria Silvello* (see Section 4)
- PRIN 2022 HOMEY project: objectives and current results, *Antonio Nocera, Emanuele Storti, Paolo Napoletano* (see Section 5)
- Supporting Energy Consumption Prediction: A Sustainable Approach, *Zahra Ziran, Massimo Mecella, Francesco Muzi, Giuseppe Piras* (see Section 6)
- The S-PIC4CHU Project: Semantics-based Provenance, Integrity, and Curation for Consistent, High-quality, Unbiased Data Science, *Gianvincenzo Alfano, Ilaria Bartolini, Diego Calvanese, Paolo Ciaccia, Sergio Greco, Davide Lanti, Emilia Lenzi, Davide Martinenghi, Christian Molinaro, Marco Patella, Letizia Tanca, Riccardo Torlone, Irina Trubitsyna* (see Section 7)

We warmly thank all participants for their valuable contributions and active engagement. We also extend our sincere gratitude to the SEBD 2025 organizing committees for their support in making this event possible and memorable.

June 2025

Francesca De Luzi
Flavia Monti
Massimo Mecella

SEBD 2025: 33rd Symposium on Advanced Database Systems, June 16-19, 2025, Ischia, Italy
*Corresponding author.
deluzi@diag.uniroma1.it (F. De Luzi); monti@diag.uniroma1.it (F. Monti); mecella@diag.uniroma1.it (M. Mecella)
0000-0002-9896-2528 (F. De Luzi); 0000-0003-3349-7861 (F. Monti); 0000-0002-9730-8882 (M. Mecella)

3. Building National Data lakehouse Ecosystems for Environmental and Public Health: AnTeA and IDEAH

Mario Cerroni[2], Francesca De Luzi[3], Tommaso Filippini[1], Valentina Fuscoletti[2], Marco Giustini[2], Raffaele Landi[4], Francesco Leotta[3], Luca Lucentini[2], Mattia Macrì[3], Camilla Marchiafava[2], Marco Marras[4], Daniela Mattei[2], Giampaolo Maugeri[4], Massimo Mecella[3], Alessio Pitidis[2], Marco Vinceti[1]

[1]Department of Biomedical, Metabolic and Neural Sciences, Section of Public Health, University of Modena and Reggio Emilia, Modena, Italy
[2]Department of Environment and Health - Istituto Superiore di Sanità, Rome, Italy
[3]Sapienza Università di Roma, Rome, Italy
[4]National Strategic Pole and We-COM Company, Viterbo, Italy

Keywords
Digital Health Infrastructure, Datalake, Smart environmental monitoring.

3.1. Introduction

Environmental quality and human health are intrinsically linked. Recognizing this connection, both the World Health Organization and the European Union have prioritized efforts to advance public well-being and support innovation through environmental and health data. In Italy, the National Complementary Plan (PNC) has funded specific actions that integrate and enhance the National Recovery and Resilience Plan (PNRR), offering an opportunity to reform and innovate the management of environmental and public health data resources. In this context, two key digital platforms have been developed to support these goals: *(i)* AnTeA[1] (Dynamic Territorial Registry of Drinking Water), a platform for the acquisition, management, and analysis of data on water quality and supply in Italy, ensuring compliance with EU Directive 2020/2184 and supporting transparent water governance, and *(ii)* IDEAH (Integrated Database for Environment And Health), a national data lakehouse [1] integrating environmental and health data to support research, epidemiology, and policy development. Both projects are coordinated by Sapienza University of Rome with institutional and scientific partners: the Italian National Institute of Health (ISS), the University of Modena and Reggio Emilia, and We-COM, cloud enabler of the National Strategic Hub (PSN), which provides the technological infrastructure for sound and scalable implementation. This collaborative network ensures the development of modern and interoperable digital infrastructures focused on improving public health and environmental monitoring, through cooperation among national institutions, academia, and regional authorities.

3.2. AnTeA – Dynamic Territorial Registry of Drinking Water

AnTeA is a digital platform created to ensure standardized, transparent, and cooperative management of drinking water data in Italy, in line with Legislative Decree no. 18/2023 and the EU Drinking Water Directive. The project addresses the fragmentation of Italy's water sector - over 2,300 providers using disparate systems - by pursuing the following objectives:

- Data harmonization: AnTeA enables the integration of data on water sources, distribution systems, and water quality. It supports internal and external control reporting, incident tracking, risk assessment, and derogation management;
- Cooperative framework: AnTeA adopts a Request for Comments (RfC) process to engage institutional stakeholders (e.g., ISS–CeNSiA, ARERA, MASE, ISTAT, Regions, ASLs, EGATOs), ensuring shared governance and continuous improvement;

[1]https://www.iss.it/antea-il-progetto

- Interoperability and scalability: built on the National Strategic Hub (PSN), AnTeA leverages a secure cloud infrastructure for data reliability, availability, and exchange with European bodies and international institutions;
- Public transparency: the platform enhances citizens' right to information about water quality, contributing to public trust and informed environmental stewardship.

3.3. IDEAH – Integrated Database for Environment And Health

IDEAH is an initiative led by the ISS, developed within the framework of the SNPS (National System for the Prevention of Health from Environmental and Climate Risks), established by Legislative Decree no. 36/2022. It provides a centralized, scalable data lakehouse architecture that integrates heterogeneous environmental and health datasets across multiple territorial scales, from international to local. The platform enhance risk assessment, disease prevention, and policy-making through advanced analytics and interoperable data access, offering the following key features:

- Integrated data sources: IDEAH consolidates 40 environmental data sources, including terrestrial and satellite data (e.g., Copernicus missions Sentinel-2 and Sentinel-5P), and health data such as mortality, hospital discharge records, emergency room visits, and birth certificates;
- Privacy and security: compliance with national data protection regulations is ensured through anonymization, semi-anonymization techniques, and strong authentication mechanisms;
- User access and profiling: access is managed via SPID digital identity with role-based permissions. Researchers provide their professional background and research objectives, allowing IDEAH to tailor data access accordingly;
- Interactive dashboards: users can filter and explore datasets through dynamic graphs and maps, extracting specific geographic or temporal subsets;
- Flexible data analysis environment: a cloud-based JupyterLab-inspired interface allows users to work in R or Python, import custom containers, and load personal libraries or configuration files.

Declaration on Generative AI

The authors have not employed any Generative AI tools.

References

[1] A. A. Harby, F. Zulkernine, Data lakehouse: A survey and experimental study, Information Systems 127 (2025) 102460.