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)

1. AI-Powered Industrial Anomaly Detection: A Dual Approach with LLMs and Machine Learning

Ala Arman[1,*], Filippo Bianchini[1], Marco Calamo[1], Loredana Cristaldi[2], Emilia Lenzi[2], Matteo Marinacci[1], Davide Martinenghi[2], Luca Martiri[2], Massimo Mecella[1], Andrea Moschetti[2], Jacopo Rossi[1], Letizia Tanca[2]

[1]Department of Computer, Control and Management Engineering, Sapienza University of Rome,
Via Ariosto, 25, 00185 Rome, Italy
[2]Department of Electronics, Information and Bioengineering Politecnico di Milano,
Via Giuseppe Ponzio, 34, 20133 Milan, Italy
*Corresponding author. Email: arman@diag.uniroma1.it

Keywords
Anomaly Detection, Convolutional Neural Networks, Deep Learning, Large Language Models, Machine Learning, Quality Control, Retrieval-Augmented Generation.

1.1. The MICS Project

The MICS Project [1, 2] is a major national initiative uniting academia and industry to promote circularity in key sectors by developing data-driven models and methods that support the full lifecycle of industrial processes. see Table 1 for more details.

Project Full Name	*Made in Italy Circolare e Sostenibile (MICS)*
Duration	January 2023 – December 2025 (36 months)
Participants	12 Public and 13 Industrial partners
Funding Agency	Italian MUR (Ministry of University and Research), funded by *NextGenerationEU (PNRR)*
Total Investment	130+ million euros
Researchers Involved	Over 1000 from academia and industry
Active Sub-Projects	72 innovation projects (as of early 2024)
Cascade Funding	21.5 million euros
Funded Organizations	87 companies and 26 universities/research institutes
Key Contributors	*Marco Taisch* – MICS President. *Elisa Negri* – Scientific Coordinator. *Roberto Merlo* – Program Research Manager. *Federica Acerbi* - Scientific Content Coordinator. *Eva De Francesco and Anna Ettorre* - Project Managers.
Official Website	https://www.mics.tech/en/home/

Table 1
The MICS project summary (as of mid-2025)

1.2. The Proposed Approach

Anomaly detection is critical for reliable and cost-effective manufacturing. To address the limitations of traditional inspections, we propose a hybrid intelligent system that combines semantic reasoning with advanced data analysis. Large Language Models (LLMs) enhanced with Retrieval-Augmented Generation (RAG) interpret technical documents to provide real-time, context-aware guidance to inspectors. Simultaneously, machine learning and deep learning techniques, including Random Forest (RF) and Convolutional Neural Networks (CNNs), analyze high-resolution images to detect and classify defects, with built-in resilience to noisy or imperfect data. This dual-track architecture integrates structured and unstructured data, enabling informed, efficient decision-making and reducing human error in complex inspection scenarios.

Acknowledgements.

This study was carried out within the MICS (Made in Italy – Circular and Sustainable) Extended Partnership and received funding from Next-Generation EU (Italian PNRR – M4 C2, Invest 1.3 – D.D. 1551.11-10-2022, PE00000004). CUP MICS D43C22003120001.

Declaration on Generative AI

The authors have not employed any Generative AI tools.

References

[1] N. Berti, A. Arman, P. Esmaili, M. Zeynivand, D. Battini, D. Bianchini, L. Cristaldi, L. B. De Giuli, A. Galeazzo, G. Gruosso, A. L. Bella, F. Leotta, L. Martiri, E. Masero, M. Mecella, P. Plebani, P. Rocco, L. Salmaso, R. Scattolini, G. A. Susto, L. Tanca, Sustainability and resilience in the mics spoke8 project: The role of the digital twin, in: 2024 IEEE International Conference on Metrology for eXtended Reality, Artificial Intelligence and Neural Engineering (MetroXRAINE), 2024, pp. 669–673.

[2] L. Cristaldi, P. Esmaili, G. Gruosso, A. L. Bella, M. Mecella, R. Scattolini, A. Arman, G. A. Susto, L. Tanca, The mics project: A data science pipeline for industry 4.0 applications, in: 2023 IEEE International Conference on Metrology for eXtended Reality, Artificial Intelligence and Neural Engineering (MetroXRAINE), 2023, pp. 427–431.

2. An Ontology-based Multidimensional Data Modeling

Domenico Lembo[1], Maurizio Lenzerini[1], Antonella Poggi[1], Federico Maria Scafoglieri[1,*],
Jacopo Brunetti[1], Roberta Radini[2], Michele Riccio[2], Valerio Santarelli[3]

[1]Sapienza Università di Roma, Rome, Italy
[2]ISTAT Italian National Institute of Statistics, Rome, Italy
[3]OBDA Systems, Rome, Italy
*Corresponding author. Email: scafoglieri@diag.uniroma1.it

Keywords
Data Warehousing, Ontologies, Semantic Technologies.

The Problem. Aggregate data, also known as macro-data, concern with information produced in summarized form from individual level data, also known as micro-data. Typically gathered from operational databases, and possibly integrated from various data sources, aggregate information is usually managed through Business Intelligence and Data Warehousing solutions [1, 2, 3, 4, 5, 6].

Data aggregation is usually carried out by referring to the so-called *multidimensional model* [4], where events of interest for the analysis are represented as logical cubes. These cubes are characterized by *dimensions* (e.g., time or space), which correspond to the aspects of the business along which one wants to perform aggregation. Dimensions may be associated to *hierarchies* specifying different *levels* of aggregation (a.k.a. dimensional attributes [4]), and by *measures*, which are properties of the event on which to make calculations (e.g., sums or averages), and that can be used as business performance indicators (e.g., income of a shop, number of enrollments in a school). Operations performed on data cubes (also called OLAP operations) include the increment or decrement of the level of aggregation, called roll-up and drill-down, respectively, or the selection of a portion of events in the multidimensional space, called slice-and-dice.

Goal. In this project we devidse a new approach for modeling and manipulating aggregate data, which is based on the use of OWL2 ontologies [7] that provide a rigorous formalization of both the application domain and the multidimensional model. The overall ontology that we devise makes it explicit the way in which macro-data are obtained from micro-data, by exploiting *views* over the domain ontology [8], which are first-class citizens in our model. Data cubes and hierarchies are indeed seen as constructed from the (SPARQL) queries associated to the views, which allow cubes dimensions, cubes measures and hierarchy levels, to be instantiated from the answers to such queries. This is a distinguishing feature of our approach, considered that other models for multidimensional data (e.g., [9, 10]) do not formalize this aspect, and methodologies for data warehouse design do not provide declarative means to specify the connection between micro- and macro-data, which is usually hidden in ETL procedures, and thus it is difficult to understand and reconstruct, e.g., for data provenance and/or lineage.

Approach. Our work is currently focused on the development of services to support both design- and run-time activities related to the production, distribution and integration of aggregate data. Such services are defined according to a formal semantics that extends the Metamodeling Semantics proposed in [11]. This semantics allows us to reason over various representation layers, i.e., the meta-level formalizing the multidimensional model, the actual data cubes designed for the analysis of the business trends, the domain ontology and the views bridging it to the cubes. A fundamental service in this scenario is query answering. Such service is indeed at the basis of several more complex functionalities, such as integration of aggregate data sets [12], possibly linked to the ontology through mappings as in OBDM [13, 14], and publishing of linked open data. Interestingly, queries in our framework may smoothly combine together elements belonging to the various mentioned levels. We finally remark that we are currently working on the implementation of software components, integrated in the OBDM tool Mastro [15, 16], that realize the multidimensional ontology-based approach devised in this project.

Acknowledgements.

Lembo, Lenzerini, and Poggi received support from the PNRR MUR project PE0000013-FAIR; Lembo's work was also funded by the POLAR project (Code 2022LA8XBH, CUP: B53D23013100006). Scafoglieri's research was entirely and exclusively supported by the PNRR MUR project PE0000013-FAIR.

Declaration on Generative AI

The authors have not employed any Generative AI tools.

References

[1] W. H. Inmon, Building the Data Warehouse, second ed., John Wiley & Sons, 1996.

[2] B. Devlin, Data Warehouse: From Architecture to Implementation, Addison Wesley Publ. Co., 1997.

[3] M. Jarke, M. Lenzerini, Y. Vassiliou, P. Vassiliadis (Eds.), Fundamentals of Data Warehouses, Springer, 1999.

[4] M. Golfarelli, S. Rizzi, Data Warehouse Design: Modern Principles and Methodologies, McGraw-Hill, 2009.

[5] R. Kimball, M. Ross, The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling, 3 ed., Wiley, 2013.

[6] M. Jarke, M. A. Jeusfeld, C. Quix, P. Vassiliadis, Architecture and quality in data warehouses, in: Proc. of CAiSE, volume 1413 of *LNCS*, Springer, 1998, pp. 93–113.

[7] B. C. Grau, I. Horrocks, B. Motik, B. Parsia, P. Patel-Schneider, U. Sattler, Owl 2: The next step for owl, Journal of Web Semantics 6 (2008) 309–322.

[8] M. Console, G. De Giacomo, M. Lenzerini, M. Namici, et al., Intensional and extensional views in dl-lite ontologies, in: Thirtieth International Joint Conference on Artificial Intelligence, International Joint Conferences on Artificial Intelligence, 2021, pp. 1822–1828.

[9] R. Cyganiak, D. Reynolds, The RDF Data Cube Vocabulary, W3C Recommendation, W3C, 2014. Available at https://www.w3.org/TR/vocab-data-cube/.

[10] The official site for the SDMX community, https://sdmx.org/, 2023.

[11] M. Lenzerini, L. Lepore, A. Poggi, Metamodeling and metaquerying in OWL 2 QL, AIJ 292 (2021) 103432.

[12] R. Fagin, P. G. Kolaitis, D. Lembo, L. Popa, F. Scafoglieri, A Framework for Combining Entity Resolution and Query Answering in Knowledge Bases, in: Proceedings of the 20th International Conference on Principles of Knowledge Representation and Reasoning, 2023, pp. 229–239. URL: https://doi.org/10.24963/kr.2023/23. doi:10.24963/kr.2023/23.

[13] M. Lenzerini, Managing data through the lens of an ontology, AI Magazine 39 (2018) 65–74.

[14] D. Lembo, L. Yunyao, L. Popa, K. Qian, F. Scafoglieri, et al., Ontology mediated information extraction with mastro system-t, in: CEUR WORKSHOP PROCEEDINGS, volume 2721, 2020, pp. 256–261.

[15] Mastro - The OBDM Engine, https://obdm.obdasystems.com/mastro/, 2023.

[16] L. Lepore, M. Namici, G. Ronconi, M. Ruzzi, V. Santarelli, D. F. Savo, et al., Monolith: an obdm and knowledge graph management platform, in: CEUR WORKSHOP PROCEEDINGS, volume 2456, CEUR-WS, 2019, pp. 173–176.

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.

4. HEREDITARY: HetERogeneous sEmantic Data Integration for the guT-brAin inteRplaY

Gianmaria Silvello[1,*]

[1]University of Padua, Department of Information Engineering, Via Gradenigo 6/b, Padova, Italy
*Corresponding author. Email: gianmaria.silvello@unipd.it

Keywords
Horizon Europe Project, Data Integration, Gut-Brain Interplay, Knowledge Extraction, Knowledge Representation.

The Project

The **HEREDITARY** project is a European research initiative funded under the Horizon Europe programme. It aims to build an integrated digital infrastructure for precision medicine using multimodal data. The University of Padua (UNIPD) coordinates the project.

- **Project Acronym:** HEREDITARY
- **Project Duration:** 1 January 2024 – 31 December 2027
- **Total Duration:** 48 months
- **EU Contribution:** €9,988,833.75
- **Funding Agency:** European Commission, Horizon Europe programme
- **Project Website:** https://hereditary-project.eu/

Work Packages Overview. The HEREDITARY project is organized into nine Work Packages (WPs), each addressing a key dimension of the project and led by a dedicated partner. **WP1**, coordinated by Università degli Studi di Padova (UNIPD), ensures overall project management, including coordination, implementation, and timely delivery. **WP2**, led by Università degli Studi di Torino (UNITO), defines clinical use cases and provides curated clinical, research, and environmental data through a federated infrastructure. **WP3**, under Aalborg Universitet (AAU), develops the semantic integration platform supporting multimodal and multilingual data analysis.

Building on this, **WP4** (Haute École Spécialisée de Suisse Occidentale – HESSO) implements an analytics and learning platform to extract insights from heterogeneous data sources using advanced AI techniques. **WP5**, led by Technische Universität Graz (TUGRAZ), focuses on visual analytics, enabling users to explore and interpret data interactively. **WP6** (Observa Associazione) enhances societal impact through citizen engagement, public communication, and policy-oriented activities.

WP7, managed by Katholieke Universiteit Leuven (KU Leuven), ensures compliance with legal, ethical, and regulatory standards, particularly concerning data protection and AI governance. **WP8**, led by Fundación Empresa Universidad Gallega (FEUGA), defines strategies for exploitation, innovation, and dissemination to promote long-term sustainability. Finally, **WP9**, also led by UNIPD, oversees adherence to the ethical requirements of the project.

Consortium Partners and Roles. The HEREDITARY consortium brings together leading institutions across Europe and beyond. **UNIPD** acts as *Project Coordinator* and leads WP1 and WP9 (https://www.unipd.it). **ONTOTEXT (ONTO)** is *Exploitation Manager* (https://www.ontotext.com), while **FEUGA** coordinates WP8 and manages intellectual property (https://www.feuga.es). **OBSERVA** leads WP6 (https://www.observanet.it), and **KU Leuven** is responsible for WP7 (https://www.kuleuven.be).

UNITO and **AAU** lead WP2 and WP3, respectively (https://www.unito.it, https://www.aau.dk), with **HESSO** heading WP4 (https://www.hes-so.ch) and **TUGRAZ** in charge of WP5 (https://www.tugraz.at). Other key contributors include **SURF BV** (https://www.surf.nl), **Radboud University Medical Centre (RUMC)** (https://www.radboudumc.nl), **CRG-CERCA** (https://www.crg.eu), **UNL** (https://www.unl.pt), **EUpALS** (https://www.eupals.eu), **European Brain Council (EBC)** as *Quality and Risk Manager* (https://www.braincouncil.eu), **University of Colorado (UCD)** (https://www.colorado.edu), **EMBL** (https://www.embl.org), and **CNAG** (https://www.cnag.eu), each contributing specialized expertise across the scientific, technical, and societal aspects of the project.

The **Project Coordinator** is Gianmaria Silvello from UNIPD, and the **Scientific and Technical Manager** is Manfredo Atzori (UNIPD-HESSO).

Scientific vision and goals

The HEREDITARY project aims to develop a secure and distributed system for linking multimodal health data, such as electronic health records, genomic data, medical imaging, and environmental data. By leveraging secure supercomputing environments and federated learning, data remains localized, respecting privacy and regulatory standards like GDPR. This infrastructure facilitates collaborative analysis without compromising sensitive health information, advancing medical research and improving patient outcomes. In addition, the project focuses on developing semantics-aware learning methods to integrate multimodal and genomics data, enhancing health outcomes. Using advanced AI techniques and Ontology-Based Data Access (OBDA), HEREDITARY creates unified data representations for complex queries and predictive analytics. These methods aim to provide deeper insights into the gut-brain axis and neurodegenerative diseases, ultimately contributing to the development of personalized medicine and healthcare solutions. Furthermore, the project empowers decision-making and strengthens citizen trust through an interactive data-driven platform for visual analytics. This platform enables researchers, clinicians, and policymakers to analyze complex health data using advanced visualization tools. By integrating explainable AI and engaging the public in the research process, HEREDITARY promotes transparency, fosters trust, and enhances public awareness, supporting informed decision-making for better health outcomes.

Resources

All public deliverables released by the project are available on the website at https://hereditary-project.eu/deliverables/ and in Zenodo at https://zenodo.org/communities/hereditaryproject/records.

The scientific publications related to the project are available at the URL https://hereditary-project.eu/publications/ and they are updated at a monthly basis.

During the initial phase of the project (spanning from month 1 to month 18), the HERITAGE consortium generated more than 50 publications spanning the project topics, including **Knowledge Graphs and Data Quality** [1, 2, 3, 4, 5, 6, 7], **Data Annotation** [8, 9], **Ontologies** [10, 11], **Deep Learning in biomedicine** [12, 13, 14, 15, 16], **Information Extraction and Evaluation** [17, 18, 19, 20, 21], **Data Integration** [22, 23], **Synthetic Data** [24], **Visualization** [25, 26], **Citizen Science** [27], and **Terminology** [28, 29, 30, 31, 32, 33].

Acknowledgments.

This work is partially supported by the HEREDITARY Project, as part of the European Union's Horizon Europe research and innovation programme under grant agreement No GA 101137074.

Declaration on Generative AI

The authors have not employed any Generative AI tools.

References

[1] S. Marchesin, G. Silvello, Efficient and reliable estimation of knowledge graph accuracy, in: Proc. of the VLDB Endowment, volume 17, 2024. doi:10.14778/3665844.3665865.

[2] S. Marchesin, G. Silvello, O. Alonso, Veracity estimation for entity-oriented search with knowledge graphs, in: Proc. of the 33rd ACM International Conference on Information and Knowledge Management (CIKM 2024), 2024. URL: https://doi.org/10.1145/3627673.3679561.

[3] S. Marchesin, G. Silvello, O. Alonso, Utility-oriented knowledge graph accuracy estimation with limited annotations: A case study on dbpedia, in: Proc. of the AAAI Conference on Human Computation and Crowdsourcing (HCOMP 2024), volume 12, 2024, pp. 105–114. URL: https://doi.org/10.1609/hcomp.v12i1.31605.

[4] F. Shami, S. Marchesin, G. Silvello, Fact verification in knowledge graphs using llms, in: Proc. of the The 48th International ACM SIGIR Conference on Research and Development in Information Retrieval, 2025. doi:10.1145/3726302.3730142.

[5] S. Marchesin, G. Silvello, Credible intervals for knowledge graph accuracy estimation, in: Proc. ACM Manag. Data 3, 3 (SIGMOD), Article 142, 2025. doi:10.1145/3725279.

[6] S. Marchesin, G. Silvello, Binomial confidence intervals for knowledge graph accuracy estimation (extended abstract), in: 33rd Symposium On Advanced Database Systems (SEBD 2025), 2025.

[7] P. Buneman, D. Dosso, M. Lissandrini, G. Silvello, H. Sun, Can we measure the impact of a database?, Communications of the ACM 68 (2025) 69–76. doi:10.1145/3704723.

[8] O. Irrera, S. Marchesin, G. Silvello, Metatron: advancing biomedical annotation empowering relation annotation and collaboration, BMC Bioinformatics 25 (2024). URL: https://doi.org/10.1186/s12859-024-05730-9.

[9] O. Irrera, S. Marchesin, F. Shami, G. Silvello, Doctron: A web-based collaborative annotation tool for ground truth creation in ir, in: Proc. of the The 48th International ACM SIGIR Conference on Research and Development in Information Retrieval, 2025. doi:10.1145/3726302.3730286.

[10] G. Faggioli, L. Menotti, S. Marchesin, A. Chiò, A. Dagliati, M. de Carvalho, M. Gromicho, U. Manera, E. Tavazzi, G. M. Di Nunzio, G. Silvello, N. Ferro, An extensible and unifying approach to retrospective clinical data modeling: The brainteaser ontology, Journal of Biomedical Semantics (2024). URL: https://doi.org/10.1186/s13326-024-00317-y.

[11] M. Cazzaro, I. G. Gut, L. Menotti, M. Rueda, G. Silvello, Hero-genomics: An ontology for integration and access of multicenter genomic data, in: roc. of the 16th International SWAT4HCLS Conference - Semantic web Applications and Tools for Health Care and Life Sciences (SWAT4HCLS 2025), 2025.

[12] F. Del Pup, A. Zanola, L. F. Tshimanga, P. E. Mazzon, M. Atzori, Selfeeg: A python library for self-supervised learning in electroencephalography, Journal of Open Source Software 9 (2024) 6224. doi:10.21105/joss.06224.

[13] A. Polejowska, F. Ayatollahi, A. S. O. Erdogan, F. Ciompi, A. Boleij, Spirochetosis detection in colon histopathology images via fine-tuning and boosting techniques using foundation models, in: Medical Imaging with Deep Learning, 2024. URL: https://ceur-ws.org/Vol-3643/paper2.pdf.

[14] L. F. Tshimanga, A. Zanola, S. Facchini, A. L. Bisogno, L. Pini, M. Atzori, M. Corbetta, Behavioral clusters and lesion distributions in ischemic stroke, based on nihss similarity network, Journal of Healthcare Informatics Research (2025).

[15] F. D. Pup, M. Atzori, Toward improving reproducibility in neuroimaging deep learning studies, Frontiers in Neuroscience (2024). doi:10.3389/fnins.2024.1509358.

[16] A. Polejowska, A. Boleij, F. Ciompi, Histopathobiome – integrating histopathology and microbiome data via multimodal deep learning, in: Proceedings of the MICCAI Workshop on Computational Pathology, volume 254 of *Proceedings of Machine Learning Research*, PMLR, 2024, pp. 203–213.

[17] G. M. Di Nunzio, E. Gallina, F. Vezzani, Unipd@simpletext2024: A semi-manual approach on prompting chatgpt for extracting terms and write terminological definitions, in: CEUR-WS, 2024. URL: https://ceur-ws.org/Vol-3740/#paper-313.

[18] G. M. Di Nunzio, F. Vezzani, V. Bonato, H. Azarbonyad, J. Kamps, L. Ermakova, Overview of

the clef 2024 simpletext task 2: Identify and explain difficult concepts, in: CEUR-WS, 2024. URL: https://ceur-ws.org/Vol-3740/#paper-306.

[19] A. Nentidis, G. Katsimpras, A. Krithara, M. Krallinger, M. Rodriguez Ortega, N. Loukachevitch, A. Sakhovskiy, E. Tutubalina, G. Tsoumakas, G. Giannakoulas, A. Bekiaridou, A. Samaras, G. M. Di Nunzio, N. Ferro, S. Marchesin, L. Menotti, G. Silvello, G. Paliouras, BioASQ at CLEF2025: The thirteenth edition of the large-scale biomedical semantic indexing and question answering challenge, in: Proc. of the 47th European Conference on Information Retrieval (ECIR 2025), 2025.

[20] M. Martinelli, G. Silvello, V. Bonato, G. M. Di Nunzio, N. Ferro, O. Irrera, S. Marchesin, L. Menotti, F. Vezzani, Overview of GutBrainIE@CLEF 2025: Gut-Brain Interplay Information Extraction, in: CLEF 2025 Working Notes, 2025.

[21] M. Martinelli, Advancing cross-document relation extraction with hybrid retrieval and knowledge-augmented reasoning, in: 33rd Symposium On Advanced Database Systems (SEBD 2025), 2025.

[22] M. Cazzaro, Design and development of a polystore system for heterogeneous biomedical data, in: 33rd Symposium On Advanced Database Systems (SEBD 2025), 2025.

[23] A. Zanola, F. D. Pup, C. Porcaro, M. Atzori, Bidsalign: a library for automatic merging and preprocessing of multiple eeg repositories, Journal of Neural Engineering (2024).

[24] F. M. Trudslev, M. Lissandrini, J. M. Rodriguez, M. Bøgsted, D. Dell'Aglio, Priveval: a tool for interactive evaluation of privacy metrics in synthetic data generation, in: Proc. VLDB Endow., 2025.

[25] S. Lengauer, P. Waldert, T. Schreck, Droplets: A marker design for visually enhancing local cluster association, in: IEEE VIS 2024 Bio+MedVis Challenge, 2024.

[26] B. Kantz, K. Innerebner, P. Waldert, S. Lengauer, E. Lex, T. Schreck, Onset: Ontology and semantic exploration toolkit, in: Proc. of the The 48th International ACM SIGIR Conference on Research and Development in Information Retrieval, 2025.

[27] G. Pellegrini, C. Lovati, Stakeholders' engagement for improved health outcomes. a research brief to design a tool for better communication and participation, Frontiers in Public Health 13 (2025). doi:10.3389/fpubh.2025.1536753.

[28] V. Bonato, G. M. Di Nunzio, F. Vezzani, A novel approach to semic analysis: Extraction of atoms of meaning to study polysemy and polyreferentiality, MDPI Languages (2024). URL: https://www.mdpi.com/2226-471X/9/4/121. doi:10.3390/languages9040121.

[29] F. Vezzani, G. M. Di Nunzio, A. Salgado, R. Costa, When LMF and TMF meet: towards a Unified Markup Framework (UMF), John Benjamins (2024).

[30] F. Vezzani, R. Costa, Variation in psychopathological terminology. a case study on body-dismorphic disorder, John Benjamins (2024). doi:10.1075/term.00078.vez.

[31] R. Costa, M. Ramos, M. Canelas, A. Mouro, Exploring terminological collocations in biomedical texts, in: 4th International Conference on Multilingual digital terminology today. Design, representation formats and management systems (MDTT) 2025, 2025.

[32] G. M. Di Nunzio, Consumer-centered technology-assisted review: Insights and challenges from the clef ehealth tasks, in: Semmelweis Medical Linguistics Conference 2025, 2025.

[33] A. Mouro, M. Canelas, M. Ramos, R. Costa, Big diagnoses, small humans: Making neuro concepts child's play through plain language, in: CS4Health 2025 Conference, 2025.

5. PRIN 2022 HOMEY project: objectives and current results

Antonio Nocera[1], Emanuele Storti[2,*], Paolo Napoletano[3]

[1]DIII, Università di Pavia, via A. Ferrata 5, 27100, Pavia, Italy
[2]DII, Università Politecnica delle Marche, via Brecce Bianche, 60131 Ancona, Italy
[3]DISCo, Università degli Studi di Milano-Bicocca, Viale Sarca 336, 20126 Milan, Italy
*Presented, corresponding author. Email: e.storti@univpm.it

Keywords
Industry 5.0, Big Data Stream, Knowledge Graph, Anomaly detection, Human-machine interaction, IoE.

5.1. General information

HOMEY (A Human-centric IoE-based Framework for Supporting the Transition Towards Industry 5.0) is a PRIN 2022 collaborative project by Università di Pavia (UNIPV), Università Politecnica delle Marche (UNIVPM), and Università degli Studi di Milano-Bicocca (UNIMIB), funded by the European Union - Next Generation EU, mission 4 component 1 (code: 2022NX7WKE, CUP: F53D23004340006).
Duration: from 28/09/2023 to 28/02/2026.
Url: https://homey-prin22.unipv.it/. Repository: https://github.com/Homey-Prin22.
Contributors from UNIPV: Antonino Nocera (PI), Marco Ferretti, Claudio Cusano, Tullio Facchinetti, Marco Arazzi; from UNIVPM: Emanuele Storti (sub-PI, research unit coordinator), Paola Pierleoni, Monica Marconi Sciarroni, Domenico Ursino; from UNIMIB: Paolo Napoletano (research unit coordinator), Simone Bianco, Raimondo Schettini, Gianluigi Ciocca, Gabriele Galimberti, Sergio Verga.

5.2. Project Objectives

Industry 5.0 is a novel paradigm identifying the transition from traditional industries towards smart, human-centric, and green-aware industrial ecosystems. In this context, the HOMEY project proposes a comprehensive framework that leverages the Internet of Everything (IoE), an evolution of the IoT interconnecting devices, people, data and processes, to create intelligent, adaptive, and human-centric industrial environments. The project is structured around three main objectives: O.1) Design and implementation of an industrial IoE framework that enables seamless interaction among humans, machines, and data. Semantic-based approaches for data management, access and monitoring ensure interoperability, while context-aware mechanisms for data extraction aim to provide each worker with a personalized view of relevant information. Security is guaranteed by role-based access and privacy-preserving anomaly detection mechanisms. O.2) Definition and design of human-centric immersive digital working environment through Augmented Reality and zero-touch interfaces. Wearable devices with sensors and lightweight Machine Learning enable gesture control and real-time feedback for an intuitive, ergonomic experience. O.3) Personalized recommendation for task execution and team building. The system will assess risk levels for tasks and perform optimal tasks assignment based on the monitored worker's stress level, effort of the task and other organizational constraints.

5.3. Current status and intermediate results

The initial outputs for O.1 are related to data management and include the definition of the metadata model of the IoE network, represented as a Knowledge Graph (KG) based on the SemIoE ontology [1]. SemIoE[2] is an OWL2 ontology built by integrating several modules (e.g., SSN/SOSA, BOT, ORG) designed to provide a structured and standardized description of entities (agents, roles, smart devices, locations, access rights, preferences) and their relations, thereby supporting semantic interoperability at IoE level. Built on top of it, a micro-service architecture delivers both basic functionalities, such as authentication and authorization, and advanced capabilities. The Data Gathering platform [2] is

[2] Ontology specification is available at https://w3id.org/semioe.

responsible for collecting heterogeneous data streams from a variety of sources, including traditional IoT sensors, computationally capable smart objects, wearable devices and IT modules such as BPM systems. It supports customized stream pre-processing (e.g., filtering, decryption, decompression), data stream collection, data post-processing (e.g., aggregation) and routing to appropriate DBMSs for persistent storage. Stream processing policies are governed by metadata stored in the KG, which describes the streams and their generators. Semantic-based Monitoring and Querying services allow users to access both real-time and historical data by formulating request using the KG terminology. These services enforce context-aware and role-based access control to ensure secure access only to data relevant to the user's location and assigned tasks. The Anomaly Detection module [3] is responsible for monitoring the behavior of devices deployed within the system and includes a privacy-preserving delegation mechanism. This enables self-monitoring and self-healing capabilities for the IoE. The solution employs a Gated Recurrent Unit (GRU) model to identify the most likely sequences of communication packets and detect anomalies based on the model's prediction errors. Additionally, to support collaboration among IoE devices in collectively training behavioral models via Federated Learning, the module integrates a homomorphic encryption mechanism, ensuring secure synchronization among agents.

The O.2 explores two complementary innovations in human-machine interaction within the context of Smart Industry 5.0. The first focuses on secure teleoperation of a robotic arm via IMUs worn by the user [4]. As a first output, a biometric authentication system based on logistic regression ensures access control, achieving an average Equal Error Rate of 8.89%, while task recognition with random forest reaches a macro F1-score of 75.60%. The second innovation adapts an arm gesture recognition system to run entirely on a consumer-grade Wear OS smartwatch, eliminating the need for cloud processing. The system runs efficiently on the edge, with only a slight drop in accuracy due to limited data.

An outcome of O.3 consists in the definition of a task recommendation module, which dynamically reallocates activities among workers. The module balances efficiency and sustainability through a flexible and periodic negotiation process, allowing workers to refuse an activity if it exceeds a sustainable stress level, as monitored via wearable devices [5]. The system is modeled using Mixed Integer Linear Programming (MILP) with a hierarchical objective function, aimed at first maximizing the number of assignments and then minimizing the cost due to reassignments, levels of stress and possible overtimes.

Finanziato
dall'Unione europea
NextGenerationEU

Declaration on Generative AI

The authors have not employed any Generative AI tools.

References

[1] M. Arazzi, A. Nocera, E. Storti, The semioe ontology: A semantic model solution for an ioe-based industry, IEEE Internet of Things Journal 11 (2024) 40376–40387.

[2] M. M. Sciarroni, M. Esposito, P. Pierleoni, E. Storti, Monitoring data streams in industry 5.0: a knowledge graph approach, in: 2024 IEEE 8th Forum on Research and Technologies for Society and Industry Innovation (RTSI), IEEE, 2024, pp. 566–571.

[3] M. Arazzi, S. Nicolazzo, A. Nocera, A fully privacy-preserving solution for anomaly detection in iot using federated learning and homomorphic encryption, Information Systems Frontiers (2023) 1–24.

[4] I. E. Stan, H. Amrani, P. Napoletano, D. D'Auria, Authenticated robotic teleoperation with task recognition, IEEE Consumer Electronics Magazine (2025).

[5] C. Diamantini, O. Pisacane, D. Potena, E. Storti, Personalized task reassignment in industry 5.0: A milp-based solution approach, in: Proceedings of the 27th International Conference on Enterprise Information Systems - Volume 2: ICEIS, INSTICC, SciTePress, 2025, pp. 813–820.

6. Supporting Energy Consumption Prediction: A Sustainable Approach

Zahra Ziran[1,*], Massimo Mecella[2], Francesco Muzi[1], Giuseppe Piras[1]

[1]Department of Astronautical, Electrical and Energy Engineering, Sapienza University of Rome, Rome, Italy
[2]Department of Computer, Control and Management Engineering, Sapienza University of Rome, Rome, Italy
*Corresponding author. Email: zahra.ziran@uniroma1.it

Keywords
Deep Learning, Energy Consumption, Machine Learning, Sustainability, Smart Buildings, Energy4Rome.

Project Context and Scientific Goals

Within the scope of the Sapienza Research project *"Automatic electrical microgrid management system through machine learning techniques"*, the broader goal is to develop intelligent systems for the automatic management of electrical microgrids through the application of data-driven methods. These systems must support efficient energy distribution while maintaining low computational overhead, making them suitable for deployment in embedded or real-time operational contexts. Within this framework, our work focuses on identifying which predictive models are most appropriate for such constrained environments by analyzing both their forecasting performance and their consumption of computational resources.

This project builds upon and extends the methodological foundation and experimental analysis established in research on sustainability-aware energy consumption prediction models [1]. The referenced study conducted a comparative evaluation of various machine learning and deep learning techniques applied to real-world residential energy datasets, underscoring the critical trade-offs between predictive accuracy and computational resource demands. Expanding on these contributions, the present study introduces the *Accuracy-Sustainability Trade-off Index* (ASTI), a novel metric that formalizes this trade-off into a unified evaluative criterion. ASTI facilitates principled model selection by integrating accuracy, memory footprint, and power consumption into a single measure, thus aligning predictive performance with the practical requirements of resource-constrained environments such as smart microgrids and edge-computing infrastructures.

In addition to proposing the ASTI metric and validating it through extensive experimentation, the research sets the foundation for future directions in sustainable energy management. These include the incorporation of more advanced and hybrid predictive architectures, the evaluation of model behavior under different climatic conditions and building types, and the implementation of adaptive strategies capable of tuning themselves to changing operational constraints. Furthermore, we aim to investigate how the relative importance of accuracy, power, and memory affects the ranking of models under different scenarios by conducting sensitivity analyses on ASTI's weighting parameters. Through this line of inquiry, the project contributes a principled and practical approach to integrating AI in the emerging field of smart and sustainable microgrid systems.

6.1. The Proposed Approach

To address the dual imperative of predictive accuracy and computational sustainability in microgrid environments, we propose a model selection framework grounded in the *Accuracy-Sustainability Trade-off Index (ASTI)*. ASTI is a novel, multi-dimensional metric designed to evaluate forecasting models by jointly considering their accuracy, memory footprint, and power consumption. By consolidating these criteria into a single quantitative score, ASTI enables rigorous, sustainability-aware comparison across a diverse set of machine learning (ML) and deep learning (DL) architectures. Beyond standard evaluation metrics such as R^2 and MSE, our approach incorporates empirical assessments of resource utilization to reflect deployment conditions typical of edge-based or embedded systems.

The conceptual development of ASTI is informed by earlier research that advocates for transparent, computationally efficient modeling practices in the context of energy management [2]. That work highlighted the viability of simple statistical models, showing that robust predictive performance can be achieved even with limited data and minimal computational burden. Building on this foundation, our approach extends the methodological horizon by providing a unified framework for assessing both traditional and advanced learning models under sustainability constraints. In doing so, we offer a means to formalize and operationalize the trade-offs often encountered in real-world applications—balancing accuracy with feasibility for long-term, resource-conscious deployment.

To empirically validate this methodology, we apply it to *Energy4Rome*, a rich dataset encompassing two years of detailed energy consumption data from four major residential complexes in Rome.[3] The dataset includes not only granular consumption records but also auxiliary contextual information such as utility bills, occupancy behavior, and architectural specifications. Within this experimental setting, a range of ML (e.g., SVR, Random Forest, XGBoost) and DL (e.g., LSTM, GRU, TCN) models are trained, optimized, and evaluated. Results indicate that although DL models exhibit marginally superior predictive accuracy, tree-based ML models—particularly XGBoost—achieve the most favorable balance according to the ASTI score. These findings underscore the practical utility of our proposed framework in guiding model selection for sustainable, performance-conscious energy forecasting in smart microgrid contexts.

Acknowledgments.

This work has been supported by the Sapienza Research project “Automatic electrical microgrid management system through machine learning techniques” (CUP: B89J21032850001).

Declaration on Generative AI

The authors have not employed any Generative AI tools.

References

[1] G. Piras, F. Muzi, Z. Ziran, Open tool for automated development of renewable energy communities: Artificial intelligence and machine learning techniques for methodological approach, Energies 17 (2024) 5726. doi:10.3390/en17225726, DOI: https://doi.org/10.3390/en17225726.

[2] Z. Ziran, M. Mecella, F. Leotta, A simplified and sustainable approach for energy prediction, in: Intelligent Environments 2024: Combined Proceedings of Workshops and Demos & Videos Session, IOS Press, 2024, pp. 104–113. doi:10.3233/AISE240022, DOI: https://doi.org/10.3233/AISE240022.

[3] Energy4Rome dataset available at: https://github.com/zahraziran/Energy4Rome

7. The S-PIC4CHU Project: Semantics-based Provenance, Integrity, and Curation for Consistent, High-quality, Unbiased Data Science

Gianvincenzo Alfano[1], Ilaria Bartolini[2], Diego Calvanese[3], Paolo Ciaccia[2], Sergio Greco[1], Davide Lanti[3], Emilia Lenzi[4], Davide Martinenghi[4], Christian Molinaro[1], Marco Patella[2], Letizia Tanca[4], Riccardo Torlone[5], Irina Trubitsyna[1]

[1]University of Calabria, DIMES, Rende (CS), Italy
[2]Alma Mater Studiorum University of Bologna, DISI, Bologna, Italy
[3]Free University of Bozen-Bolzano, Faculty of Engineering, Bolzano, Italy
[4]Politecnico di Milano, DEIB, Milano, Italy
[5]Roma Tre University, DICITA, Roma, Italy

Project Data

Acronym: S-PIC4CHU
Duration: February 2025 – January 2027 (24 months)
Funding agency: Min. dell'Univ. e della Ricerca (MUR), Bando PRIN 2022 – Scorrimento
Project code: 2022XERWK9
Budget: € 210 694 (Funding: € 169 057)
Keywords: Data Science, data preparation data quality, semantics, ontologies, inconsistency, incompleteness, knowledge graphs, provenance, explanation, bias

Project Summary

The effectiveness of data-driven solutions, in Data Science as well as in Machine Learning, clearly depends on the quality and interpretability of the underlying data. Unfortunately, real-world data is often incomplete, inconsistent, biased, or lacks adequate semantic description. Traditional data preparation workflows typically rely on ad hoc methods, limited automation, and minimal consideration of domain knowledge, resulting in inefficiencies and unreliable analytical outcomes.

The S-PIC4CHU project proposes embedding semantics at the core of each stage of the process: a novel architecture for data preparation, grounded in a semantics-based methodology that supports provenance tracking, integrity enforcement, and fairness assessment. This paradigm shift is based on the design and implementation of a semantically-aware Data Preparation Pipeline (DPP), integrated with a corresponding Semantic Transformation Pipeline (STP): each data transformation step is semantically annotated through mappings to ontologies and knowledge graphs, enabling enhanced traceability, transparency, and reasoning over data.

A major contribution of the project is the formalization and implementation of semantic enrichment techniques that provide domain-aware annotations to both structured and multimedia data and support advanced operations like semantic imputation of missing values, preference-based resolution of inconsistencies, and detection and mitigation of bias in datasets. Importantly, the project tackles fairness not merely as a downstream property of algorithmic outputs, but as a core feature of input data, thus addressing societal concerns related to discrimination and ethical decision-making in AI systems.

The methodology will be validated through two concrete use cases from different domains: healthcare and sustainable urban development. In collaboration with the Policlinico Universitario A. Gemelli (Rome), the project will address challenges in preparing complex medical data for predictive modeling and decision support. Simultaneously, the project will work with the IMM Design Lab at Politecnico di Milano to support policy-making processes in urban planning through the integration of heterogeneous environmental and social datasets.

S-PIC4CHU is expected to yield several impactful outcomes: open-source software tools implementing the proposed methodologies, scientific publications targeting top-tier venues in data management and artificial intelligence, and educational resources for training the next generation of data scientists. The project also emphasizes outreach and engagement with public institutions and private stakeholders to foster the adoption of fairness-aware and semantically-grounded data processing pipelines.

In line with the objectives of the SEBD Research Project Exhibition, S-PIC4CHU represents a forward-looking initiative with the potential to influence both theoretical research and practical applications in the field of data science. By addressing fundamental issues related to data quality, interpretability, and fairness, the project contributes to the development of trustworthy and socially responsible data-driven technologies.

Acknowledgments.

This work is supported by the Italian Ministry of University and Research (MUR) PRIN 2022 grant 2022XERWK9 “S-PIC4CHU - Semantics-based Provenance, Integrity, and Curation for Consistent, High-quality, and Unbiased data science”.

Declaration on Generative AI

The authors have not employed any Generative AI tools.