From Fragmented Data to Smart Conversations in Energy Communities: The GAIA Approach to Cross-Domain IoT Integration

(Article begins on next page)

From Fragmented Data to Smart Conversations in Energy Communities: The GAIA Approach to Cross-Domain IoT Integration

1st Alessio Viticchié
AlphaWaves S.r.l.
Torino, Italy
a.viticchie@awaves.it

2nd Felice Cetrone
AlphaWaves S.r.l.
Torino, Italy
f.cetrone@awaves.it

3rd Roberto Puntorieri
AlphaWaves S.r.l.
Torino, Italy
r.puntorieri@awaves.it

4th Christian Camarda
Midori S.r.l.
Torino, Italy
christian@midori.it

5th Leonardo Napoli
Bosco Elettronica S.r.l.
Torino, Italy
leonardo.napoli@boscoelettronica.com

6th Edoardo Patti
Politecnico di Torino
Torino, Italy
edoardo.patti@polito.it

7th Alessandro Aliberti
Politecnico di Torino
Torino, Italy
alessandro.aliberti@polito.it

***Abstract*—In modern digital ecosystems, managing heterogeneous data sources is a significant challenge, particularly within Renewable Energy Communities (RECs), where multiple energy vectors, such as electricity, heating, and water, must be integrated seamlessly.**

The GAIA meta-platform addresses the persistent fragmentation of IoT ecosystems by enabling federated access, semantic harmonization, and cross-domain analytics across heterogeneous data silos. Designed to support both expert and non-expert users, GAIA combines modular data processing, a Python SDK, and an AI-driven conversational agent (i.e., GAIA Chat) to facilitate intuitive interaction with multi-source datasets. This paper presents the platform's architecture and functionalities, emphasizing its role in advancing data-driven services for RECs.

Finally, a real-world deployment demonstrates GAIA's ability to integrate energy and water data, enabling advanced use cases such as cross-domain anomaly detection and indirect consumption estimation. The results validate GAIA as a scalable, domain-agnostic infrastructure capable of supporting intelligent services in complex smart environments.

***Index Terms*—Internet of Things, Federated IoT Meta-Platform, Multi-Energy Vectors Management, Renewable Energy Communities, Climate Change, AI Agents**

I. Introduction

The global imperative for decarbonization and sustainable development has accelerated the adoption of novel paradigms within the energy sector, notably Renewable Energy Communities (RECs), Smart Grids, Smart Cities, and Multi-Energy Systems (MES). These systems fundamentally depend on real-time data derived from an extensive range of sensors and devices to optimize processes of energy generation, distribution, and consumption. Nevertheless, despite significant advancements in Internet of Things (IoT) technologies, the persistent issue of data fragmentation remains a critical bottleneck, arising from heterogeneous and isolated data management platforms [1].

IoT infrastructures often evolve as domain-specific silos, where specialized service providers manage and distribute raw or processed data based on their offerings. These silos typically employ proprietary protocols, distinct data models, and varied access policies, resulting in limited interoperability and complicating integrated analyses across different domains. For instance, an energy community may monitor electricity, water consumption, and meteorological data through disconnected systems, limiting the ability to perform joint analyses that could reveal usage patterns, inefficiencies, or optimization opportunities.

To address fragmentation across IoT ecosystems, the GAIA meta-platform adopts a federated architecture designed to integrate heterogeneous infrastructures. By leveraging semantic models and standardized interfaces, GAIA enables seamless access to both raw and processed data across silos, facilitating the development of cross-domain applications and enabling the construction of multi-domain datasets. To further support users, the platform includes GAIA Chat, an AI-driven conversational agent equipped with data analysis and visualization tools. Through GAIA Chat, even non-technical users can perform integrated analytics, overcoming the incompatibilities that typically hinder cross-domain insights.

Expanding on this foundation, GAIA extends its federated architecture and interoperability frameworks to RECs, Smart Grids, Smart Cities, and MES, supporting their specific operational needs through real-time data integration and cross-domain analytics, helping to enhance energy efficiency and local sustainability. Notably, the first real-world implementation was implemented within a REC, demonstrating its practical value in optimizing energy use, resource management, and community engagement.

This paper presents a comprehensive overview of the GAIA Meta-Platform, with a focus on its architectural design, the novel functionalities introduced, and its practical implementations across relevant application domains. Specifically, we identify and describe core platform components, interoperability frameworks, and tangible services enabled by GAIA. Through these discussions, we demonstrate how GAIA establishes a unified, scalable, and secure foundation that fosters innovation in decentralized, citizen-focused, and data-driven energy systems.

The rest of the paper is organized as follows: Section II

This publication is part of the project NODES which has received funding from the MUR-M4C2 1.5 of PNRR founded by the European Union - NextGenerationEU (grant agreement no. ECS00000036).

979-8-3315-9515-9/25/$31.00 ©2025 IEEE

reviews related work on IoT interoperability, semantic integration, and natural language interfaces. Section III introduces the GAIA architecture. Section IV details the GAIA Core Processing Engine. Section V focuses on GAIA Chat and its role in cross-domain analytics. Section VI presents a real-world use case within a REC. Section VII discusses key insights and future directions. Finally, Section VIII provides the concluding remarks.

II. Related Works

Over the past decade, IoT has attracted considerable research interest across domains such as energy [2], mobility [3], healthcare [4], and Smart Cities [5]. These applications have demonstrated IoT's potential to enhance efficiency, monitoring, and user interaction, while also revealing persistent fragmentation across data platforms and infrastructures.

To mitigate interoperability challenges, several initiatives have proposed the federation of heterogeneous IoT platforms. Frameworks like FIESTA-IoT [6] and INTER-IoT [7] support the integration of diverse testbeds through semantic meta-directories and standardized APIs, enabling cross-domain use cases such as smart cities and building automation. Further developments, including symbIoTe [8] and Data Spine [9], introduced layered architectures and federated service buses for transparent access across industrial silos. Platforms such as VirIoT [10] provide virtualized access via broker interfaces compliant with standards like oneM2M and NGSI-LD. Despite these advancements, real-world implementations enabling cross-domain analytics remain limited. Most platforms prioritize data interoperability but fall short of delivering unified analysis across different domains. A significant gap also persists in the availability of combined datasets, particularly those integrating data from multiple utility domains such as electricity and water. As highlighted in [11], there are few available datasets that concurrently include both water and electricity consumption. This lack of integrated data represents a missed opportunity for more comprehensive analyses, as the correlation between electricity and water usage can provide deeper insights into consumption patterns and potential optimization strategies.

Alongside advancements in platform interoperability, Natural Language Interfaces (NLIs) have significantly evolved over the past decade, transitioning from simple rule-based query systems to expressive and dynamic tools for interacting with data-intensive platforms [12]. This transformation has been driven by developing Large Language Models (LLMs) [13], which now exhibit advanced capabilities such as contextual reasoning, tool usage, and structured task execution. These advancements have led to the emergence of LLM-based autonomous agents [14]. By leveraging the expressive and dynamic nature of LLMs, modern NLIs empower users to perform complex, multi-step analytical queries and generate visual outputs through natural language, thereby significantly reducing the cognitive and technical barriers typically associated with data analysis [15], [16].

In summary, while existing platforms address interoperability to varying extents, they often fall short when it comes to unified, cross-domain analytics. Based on these considerations, GAIA's scientific novelty emerges from its ability to integrate not only semantic harmonization and modular data processing, but also human-centered interaction through natural language. This distinctive combination enables stakeholders to analyze and interpret heterogeneous data sources, thus bridging the gap between fragmented data ecosystems and coherent, actionable insights.

III. GAIA Meta-Platform

The GAIA Meta-platform tackles the growing complexity and fragmentation of modern IoT data ecosystems by offering a unified framework for the integration and management of heterogeneous data across multiple domains. Designed with accessibility at its core, GAIA supports expert users, such as data scientists, and non-expert stakeholders, such as citizens and public administration, through intuitive interfaces and tailored services. By leveraging GAIA's services, stakeholders can derive real-time insights, optimize resource utilization, and deploy advanced, data-driven solutions across multiple IoT domains. Additionally, GAIA abstracts domain-specific data models, streamlining the development of single-domain applications. This decoupling enables smooth integration or substitution of data sources without requiring changes to application-level logic, thereby enhancing both flexibility and scalability.

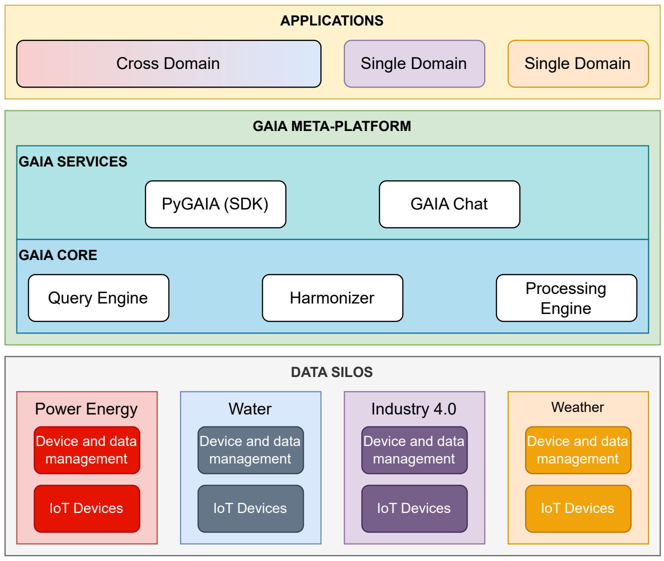

Fig. 1. Overview of the GAIA Meta-platform architecture, illustrating its layered structure: i) the Data Silos Layer, which includes domain-specific silos (e.g., Energy, Water, Industry 4.0, Weather); ii) the GAIA Meta-platform, which includes the macro components—GAIA Core, handling interoperability and data federation via modular microservices, and GAIA Services, offering high-level interfaces for external interactions; and iii) the Applications Layer, enabling vertical and cross-domain services based on harmonized data.

The logical architecture leveraging the GAIA Meta-platform is organized into three distinct layers (Fig. 1), arranged from bottom to top as follows: *Data Silos Layer*, *GAIA Meta-platform*, and *Applications Layer*. Each layer is assigned well-defined responsibilities to ensure modularity and scalability. It is important to emphasize that GAIA itself constitutes solely a meta-layer within this architectural framework.

- *Data Silos Layer*: This layer consists of isolated IoT infrastructures, each operating within a specific application domain (e.g., energy, environment, industrial systems). A typical Data Silo is composed of two main components: a network of *IoT Devices*, responsible for sensing and acquiring raw data from the physical environment, and a *Device and Data Management* unit, which processes, stores, and manages internal access

to that data. These silos are designed to function autonomously and typically lack mechanisms for cross-domain interoperability.
- *GAIA Meta-platform*: Acting as an operational backbone, this layer provides functionalities for data registration, validation, harmonization, and query management. It interfaces with the underlying Data Silos Layer by retrieving data through each silo's Device and Data Management unit, which exposes access to the internally managed data generated by IoT Devices. This layer also includes user-oriented tools and interfaces, such as the PyGAIA SDK for developers and GAIA Chat for intuitive, natural language interaction
- *Applications Layer*: This layer encompasses external, independent applications that consume the harmonized data provided by GAIA Meta-platform. Applications can operate either in a cross-domain or single-domain manner. For example, a cross-domain application may integrate and correlate data from multiple silos in the lower architectural layer—for instance, combining water and electricity meter data to identify consumption patterns and provide recommendations for optimizing overall resource usage. In contrast, a single-domain application may focus on the post-processing and analysis of data from a single silo, such as performing Non-Intrusive Load Monitoring (NILM) [17] and energy disaggregation exclusively based on electricity consumption data.

This layered architectural approach fosters a clear separation of concerns, promoting a structured yet flexible environment capable of supporting a wide range of IoT application scenarios and accommodating evolving requirements.

A. Functional Overview of the GAIA Meta-platform

GAIA Meta-platform is structured following the microservices software design pattern mainly providing two macro components: *GAIA Core* and *GAIA Services*, as illustrated in Fig. 1. These components collectively provide comprehensive functionalities to manage, harmonize, and expose IoT data efficiently.

1) The GAIA Core: serves as the central intelligence of the Meta-platform, orchestrating key functions related to data interoperability, federation, and access. It is composed of modular microservices that manage the entire data integration lifecycle—from provider registration to query execution and semantic harmonization. The Core also incorporates formal descriptors that define data structures, access mechanisms, and semantic relationships, ensuring consistent interpretation and integration across heterogeneous sources.

The GAIA Core is structured around key components, each fulfilling specific roles, as follows:

- *Query Engine* manages the routing and execution of data requests. It translates user queries into actionable tasks, optimizes query routing, and coordinates data retrieval from relevant silos [18].
- *Harmonizer* standardizes and aggregates data retrieved from heterogeneous sources, returning structured and semantically harmonized data. This unified format subsequently supports further analytics performed by the Processing Engine [18].
- *Processing Engine* is designed to perform additional analytics and data processing tasks. Further details on this component are provided in Secction IV.

2) The GAIA Services: provide simplified, high-level interfaces that can be employed by the Application Layer to interact with the GAIA Core, enabling interoperability and facilitating access for external applications, while supporting user interactions across varying levels of expertise. In particular, the GAIA Services include the following components designed to bridge the Application Layer and the Core:

- *PyGAIA SDK* (Software Development Kit) provides a programmatic interface for requesting data from the GAIA Meta-platform and for configuring the Processing Engine to enable data transformation and processing operations. It constitutes the communication link between the GAIA Meta-platform and the Application Layer, ensuring standardized and efficient access to platform functionalities for the development of data-centric applications.
- *GAIA Chat* is an AI-driven conversational agent allowing natural language queries and providing graphical data visualizations. It exhibits key characteristics of LLM-based autonomous agents, particularly through its ability to interpret user goals, determine appropriate tools to invoke, and orchestrate data processing tasks accordingly.

Together, these services facilitate effective application development and deployment over the GAIA platform.

IV. Processing Engine: Modular and Generalizable Data Processing module

At the heart of the GAIA Core lies the Processing Engine, responsible for transforming semantically harmonized, multi-source data into structured, analysis-ready formats. It performs key preprocessing tasks such as temporal alignment, missing data imputation, anomaly detection, and feature normalization, organized into modular blocks that can be selectively activated and configured. Users define workflows programmatically via the PyGAIA SDK, with each block offering default settings for ease of use and advanced customization when needed. This flexible architecture ensures data quality, temporal consistency, and statistical robustness, forming a solid foundation for downstream analytics and Machine Learning (ML).

In the following, we provide an in-depth overview of each processing component and its contribution to flexible, user-defined workflows.

Temporal Alignment and Resampling: A key challenge in integrating data from multiple IoT sources is the asynchronous and irregular nature of their update patterns, as sensors often operate at varying sampling frequencies. To ensure temporal consistency, GAIA employs a temporal alignment module that reindexes data streams based on a configurable global timeline with fixed-interval timestamps.

Missing Data Detection and Imputation: Missing data is common in IoT systems due to connectivity issues, sensor faults, or configuration errors. To address this, the module implements a structured process combining detection and imputation strategies. Available imputation methods include forward/backward filling, mean substitution, and linear or polynomial interpolation.

Anomaly Detection: The Processing Engine applies both statistical and ML techniques to detect anomalies in time series data, signaling potential sensor faults or extreme events. The supported methods include Interquartile Range (IQR), Z-score analysis and Isolation Forest. If specified, detected anomalies can be treated as missing values and subsequently imputed to maintain data continuity and enhance reliability.

Feature Scaling: To ensure balanced feature representation and compatibility with ML models, all numeric variables can undergo feature scaling. The Processing Engine offers optional data scaling, with Min-Max scaling as the default method to standardize feature ranges. Additionally, users can select other feature scaling techniques according to their needs, including Absolute Maximum Scaling and Standardization (Z-score scaling).

V. GAIA Chat: AI-driven Conversational Agent for Cross-Domain Data Accessibility

GAIA Chat is an AI-driven conversational agent integrated within the GAIA Meta-platform, designed to facilitate access to and analysis of heterogeneous data through natural language interactions. Unlike traditional interfaces, GAIA Chat operates as an autonomous agent, capable of efficiently interpreting user goals, invoking appropriate tools when needed, and coordinating data processing tasks.

The system supports two primary user profiles that the user can select: expert and non-expert. Non-expert users benefit from predefined workflows and default settings that simplify complex data analysis tasks. They can interact with the platform through intuitive, high-level queries without needing technical knowledge, and GAIA Chat uses the PyGAIA SDK as a fundamental tool to manage and manipulate data. Conversely, expert users can access more granular control over data processing and analysis, specifying preprocessing blocks, fine-tuning parameters, and directly configuring the data processing pipeline through natural language commands. This flexibility ensures that both expert and non-expert users can leverage the full potential of the GAIA Meta-platform.

In addition to expertise-level customization, GAIA Chat allows users to select one or more areas of expertise, such as energy monitoring, water management, or environmental analysis. This selection tailors the chatbot's responses to the specific context, enhancing the relevance and precision of the analysis. The embedded knowledge in GAIA Chat is achieved through carefully crafted prompting and the integration of domain-specific insights, enabling the chatbot to effectively guide users through structured data retrieval and analysis.

As illustrated in Fig. 2, GAIA Chat consists of four different components:

- A *Large Language Model* serves as the core component responsible for reasoning, planning, and interpreting inputs to generate responses. GAIA Chat supports the integration of various LLMs, including both open-source and proprietary models.
- The *Coordination Layer* represents the core personality of the autonomous agent, encapsulating the identity and behavior of the chatbot. It serves as the orchestrator that ensures the agent maintains a consistent and context-aware interaction throughout the data analysis process. The Coordination Layer includes key components such as Instructions, Objectives, Expertise, and Domain-specific prompts, with the latter two being selected by the user.
- The *Predefined Tools* component extends the agent's capabilities beyond mere text generation, allowing it to interact with its environment through specialized resources. In the context of GAIA Chat, the primary tools are the PyGAIA SDK and a Visual Analytics library specifically designed within the GAIA framework, which enables the automatic generation of graphical outputs such as time series plots, correlation matrices, and statistical distributions, thereby enhancing data interpretability and supporting data-driven decision-making.
- The *External Tool Connector* enables the integration of application-specific tools into the GAIA Chat environment. Acting as a bridge between GAIA Chat and external applications, it allows the chatbot to interact with specialized tools beyond the predefined set.

This modular architecture enables the platform to adapt to various application domains and evolving analytical requirements.

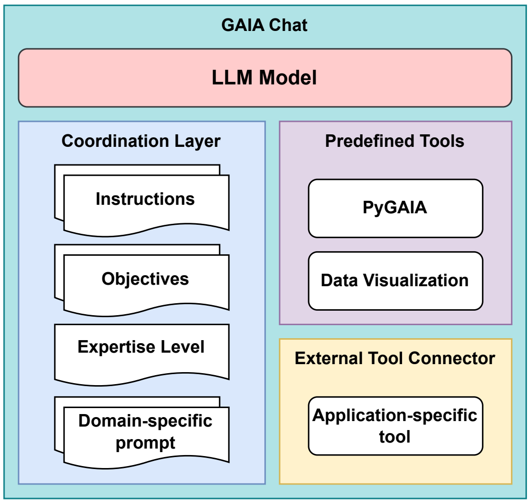

Fig. 2. Internal architecture of GAIA Chat, comprising four main components: a Large Language Model (LLM) for reasoning and planning; a Coordination Layer managing instructions, objectives, expertise, and domain prompts; Predefined Tools, such as the PyGAIA, enabling data access and interaction with the GAIA Core; and an External Tool Connector facilitating integration with application-specific tools for adaptive, domain-specific data analysis through natural language interaction.

VI. Real-World Use Case: Integrated Water and Electricity Monitoring in Alpine RECs

To illustrate the practical implementation of the GAIA meta-platform, we present its first real-world deployment within a REC, encompassing a heterogeneous set of users including 25 residential households, 2 restaurants, 2 office spaces, and 1 retail shop.

The primary objective of this deployment was to validate GAIA's capability to federate data from heterogeneous silos across different domains—specifically electricity and water—harmonize diverse datasets, and support integrated analytics for multi-vector resource consumption. This deployment reflects a deliberately heterogeneous mix of buildings and users—ranging from permanent residences and second homes to commercial activities with distinct usage profiles—thereby ensuring a comprehensive validation of the platform's interoperability, adaptability, and scalability in diverse real-world conditions.

In this deployment, stakeholders were also able to access and test GAIA Chat in an operational environment, validating its role in improving data accessibility and interpretability.

The federated dataset enabled by GAIA supports advanced analytical functions, including anomaly detection and indirect consumption estimation, demonstrating the platform's potential for cross-domain reasoning.

A. Federated Data Silos: Power Energy and Water

Two primary data silos were federated into the GAIA meta-platform during this deployment:

- *The Power Energy Silo* provides access to raw and disaggregated electricity consumption data via authenticated RESTful APIs. The dataset includes appliance-level breakdowns derived through NILM techniques [17]. Each monitored unit was equipped with two IoT Devices, installed in the main electrical panel, sampling current and voltage at one-second intervals.
- *The Water Silo* offers real-time and historical water consumption data, accessible through secure RESTful API endpoints. Data is acquired via a single IoT Device per unit, connected to pulse-output water meters.

These silos constitute the foundation of a unified data layer. The resulting harmonized and semantically enriched dataset supports cross-vector resource analysis and informed decision-making within RECs.

B. Exploring Electricity and Water Consumption through GAIA's Cross-Domain Chat Interface

GAIA Chat was actively evaluated during the pilot as an intuitive interface for multi-domain data access via natural language interaction. Participants in the REC scenario engaged with the platform by selecting their domain of interest: electricity, water, or both, through predefined, domain-specific prompts.

Two concrete applications of this cross-domain approach were successfully demonstrated during the pilot. The first involves the indirect estimation of water consumption based on electricity usage data. As an illustrative example, Fig. 3 depicts the result of the user query:

> *"Show me the water consumption of the washing machine from April 1 to April 7."*

By analyzing both the duration and the amount of electrical energy consumed during washing machine operations, inferred from disaggregated electrical consumption data obtained through NILM, the platform estimated the corresponding water usage through a data-driven approach. To achieve this, the data was first aligned to ensure temporal consistency between electricity and water measurements. Subsequently, recurring patterns were identified to capture correlations between the energy consumed during specific washing cycles and the corresponding water usage. This approach allowed for the reconstruction of water consumption data even in the absence of direct metering.

The second example, shown in Fig. 4, demonstrates joint anomaly detection, configured by an expert user and based on Z-scores, computed across both electricity and water datasets. The Z-score is a statistical measure that quantifies how many standard deviations σ a value x deviates from the mean μ and is defined as:

$$Z = \frac{x - \mu}{\sigma}$$

The corresponding query used to generate the visualization was as follows:

> *"Generate a scatter plot of electricity and water Z-scores for the past 4 months, identifying as anomalies the points where the absolute electricity Z-score exceeds 1 while simultaneously the water Z-score exceeds 2. Display anomalies in red and annotate the anomalous points with dates in DD-MM format."*

This customizable approach allows expert users to fine-tune the anomaly detection sensitivity, facilitating the identification of significant deviations that occur concurrently in both domains, revealing potential joint irregularities in electricity and water usage, such as simultaneous peaks, mismatched high or low values, or unexpected low readings indicating sensor faults.

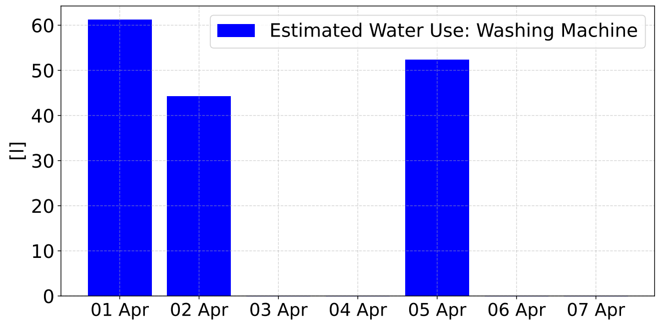

Fig. 3. Estimated daily water consumption of a real-world washing machine inferred from electricity usage data through cross-domain data analysis.

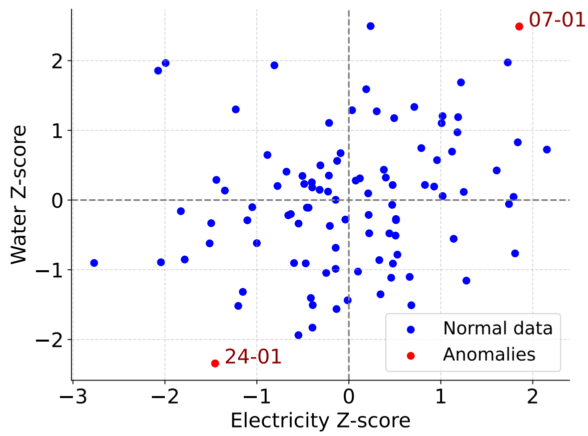

Fig. 4. Joint anomaly detection based on electricity and water data. In this specific instance, anomalies (in red) were identified when the user-defined Z-score threshold was exceeded.

These use cases highlight how cross-domain reasoning enhances situational awareness, supports proactive interventions and reduces the need for redundant sensing infrastructure. GAIA's capacity to deliver actionable insights aligns with the broader objectives of sustainability, self-consumption, and informed community participation.

VII. Discussion & Next Steps

The deployment of the GAIA Meta-platform within the REC setting has substantiated its foundational principles and demonstrated its capacity to federate heterogeneous data silos, delivering harmonized, cross-domain insights essential for intelligent service delivery.

A distinguishing feature underscored during the pilot was the introduction of GAIA Chat, an AI-driven conversational agent designed to democratize access to complex datasets. GAIA Chat enables both expert and non-expert users to interact with multi-domain data using intuitive queries, thereby lowering technical barriers and enhancing the interpretability of analytical outputs. The system leverages LLMs, predefined tools and application-specific tools to carry out advanced data analysis. The validation of GAIA Chat in an operational environment underscores its role as a key enabler of human-centered, transparent, and efficient data interaction, which is critical for fostering engagement and facilitating informed decision-making among REC stakeholders.

While the REC context provided a valuable practical setting, GAIA's domain-agnostic architecture makes it broadly applicable across other smart environments, including urban infrastructure, industrial operations, environmental monitoring, and mobility systems.

Looking ahead, several strategic enhancements are planned to elevate GAIA's capabilities and expand its applicability:

- *Enhanced Data Processing Modules:* Ongoing developments aim to extend the Processing Engine with advanced functionalities such as categorical encoding, threshold-based alerting, and automatic feature generation (e.g., rolling statistics, temporal indicators, lagged variables, and cumulative metrics). Additionally, efforts are being made to support data export in multiple formats (e.g., `CSV`, `Parquet`).
- *Extended GAIA Chat Toolset:* Future developments will focus on enhancing the GAIA Chat Toolset by integrating image processing capabilities to analyze unstructured data, such as utility bills. This will enable the chatbot to extract relevant information from documents, which can be leveraged for Retrieval-Augmented Generation (RAG), allowing for more accurate and context-aware responses. Additionally, we are working on incorporating a time series forecasting tool featuring pre-trained models specifically designed to predict consumption trends by exploiting cross-domain data. These models will facilitate the analysis of interconnected variables (e.g., energy and water usage), thereby broadening the analytical scope of GAIA Chat and supporting proactive maintenance within Smart Communities.
- *Digital Twin Enablement:* By consolidating semantic standardization, federated access, and modular analytics, GAIA is positioned to serve as a foundational layer for Digital Twin architectures, facilitating predictive simulations and operational decision-making across sectors.

Through these advancements, GAIA is set to evolve into a pivotal, future-ready digital infrastructure that supports scalable, interoperable, and citizen-centric energy and sustainability ecosystems.

VIII. Conclusions

This paper has presented the GAIA Meta-platform as a robust, federated solution for cross-domain data integration, validated through its deployment within Renewable Energy Communities. By harmonizing heterogeneous data and providing an accessible agentic interface like GAIA Chat, the platform demonstrates significant potential to support both expert and non-expert stakeholders. With planned expansions in analytical tools, unstructured data handling, and Digital Twin capabilities, GAIA is poised to serve as a versatile and scalable infrastructure for data-driven innovation across smart environments.

References

[1] M. Noaman, M. S. Khan, M. F. Abrar, S. Ali, A. Alvi, and M. A. Saleem, "Challenges in integration of heterogeneous internet of things," *Scientific Programming*, vol. 2022, 2022.

[2] N. Hossein Motlagh, M. Mohammadrezaei, J. Hunt, and B. Zakeri, "Internet of things (iot) and the energy sector," *Energies*, vol. 13, no. 2, p. 494, 2020.

[3] S. Muthuramalingam, A. Bharathi, S. Rakesh Kumar, N. Gayathri, R. Sathiyaraj, and B. Balamurugan, "Iot based intelligent transportation system (iot-its) for global perspective: A case study," *Internet of things and big data analytics for smart generation*, pp. 279–300, 2019.

[4] Y. Yuehong, Y. Zeng, X. Chen, and Y. Fan, "The internet of things in healthcare: An overview," *Journal of Industrial Information Integration*, vol. 1, pp. 3–13, 2016.

[5] A. S. Syed, D. Sierra-Sosa, A. Kumar, and A. Elmaghraby, "Iot in smart cities: A survey of technologies, practices and challenges," *Smart Cities*, vol. 4, no. 2, pp. 429–475, 2021.

[6] M. Serrano, A. Gyrard, M. Boniface, P. Grace, N. Georgantas, R. Agarwal, P. Barnagui, F. Carrez, B. Almeida, T. Teixeira *et al.*, "Cross-domain interoperability using federated interoperable semantic iot/cloud testbeds and applications: The fiesta-iot approach," in *Building the Future Internet through FIRE.* River Publishers, 2022, pp. 287–321.

[7] M. Ganzha, M. Paprzycki, W. Pawłowski, P. Szmeja, and K. Wasielewska, "Semantic interoperability in the internet of things: An overview from the inter-iot perspective," *Journal of Network and Computer Applications*, vol. 81, pp. 111–124, 2017. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1084804516301618

[8] S. Soursos, I. P. Zarko, and I. Book, "Symbiote: Symbiosis of smart objects across iot environments," in *Digitising the Industry—Internet of Things Connecting the Physical, Digital and Virtual Worlds.* River Publishers, 2016, pp. 303–307.

[9] R. A. Deshmukh, D. Jayakody, A. Schneider, and V. Damjanovic-Behrendt, "Data spine: a federated interoperability enabler for heterogeneous iot platform ecosystems," *Sensors*, vol. 21, no. 12, p. 4010, 2021.

[10] A. Detti, G. Tropea, G. Rossi, J. A. Martinez, A. F. Skarmeta, and H. Nakazato, "Virtual iot systems: Boosting iot innovation by decoupling things providers and applications developers," in *2019 Global IoT Summit (GIoTS)*, 2019, pp. 1–6.

[11] A. Di Mauro, A. Cominola, A. Castelletti, and A. Di Nardo, "Urban water consumption at multiple spatial and temporal scales. a review of existing datasets," *Water*, vol. 13, no. 1, p. 36, 2021.

[12] W. Zhang, Y. Wang, Y. Song, V. J. Wei, Y. Tian, Y. Qi, J. H. Chan, R. C.-W. Wong, and H. Yang, "Natural language interfaces for tabular data querying and visualization: A survey," *IEEE Transactions on Knowledge and Data Engineering*, 2024.

[13] M. U. Hadi, R. Qureshi, A. Shah, M. Irfan, A. Zafar, M. B. Shaikh, N. Akhtar, J. Wu, S. Mirjalili *et al.*, "A survey on large language models: Applications, challenges, limitations, and practical usage," *Authorea Preprints*, vol. 3, 2023.

[14] L. Wang, C. Ma, X. Feng, Z. Zhang, H. Yang, J. Zhang, Z. Chen, J. Tang, X. Chen, Y. Lin *et al.*, "A survey on large language model based autonomous agents," *Frontiers of Computer Science*, vol. 18, no. 6, p. 186345, 2024.

[15] M. Sun, R. Han, B. Jiang, H. Qi, D. Sun, Y. Yuan, and J. Huang, "A survey on large language model-based agents for statistics and data science," *arXiv preprint arXiv:2412.14222*, 2024.

[16] Y. Guo, D. Shi, M. Guo, Y. Wu, N. Cao, and Q. Chen, "Talk2data: A natural language interface for exploratory visual analysis via question decomposition," *ACM Transactions on Interactive Intelligent Systems*, vol. 14, no. 2, pp. 1–24, 2024.

[17] M. Castangia, A. A. Girmay, C. Camarda, and E. Patti, "Training nonintrusive load monitoring algorithms without supervision from submeters," *IEEE Transactions on Industrial Informatics*, vol. 20, no. 4, pp. 5440–5448, 2023.

[18] A. Viticchié, F. Cetrone, C. Camarda, V. Vassallo, L. Napoli, E. Patti, and A. Aliberti, "Gaia meta-platform: enabling multi-energy vectors data analysis via iot federation," in *2024 IEEE International Conference on Environment and Electrical Engineering and 2024 IEEE Industrial and Commercial Power Systems Europe (EEEIC/I&CPS Europe).* IEEE, 2024, pp. 1–6.