Green Energy and Sustainability
https://doi.org/10.47248/ges2606030008

Original Research

A comprehensive framework for data management and AI-driven analytics in the energy sector

Alexandros Tsitsanis *, Fenareti Lampathaki

Suite5 Data Intelligence Solutions Limited, Limassol 3013, Cyprus;
Email: fenareti@suite5.eu

* **Correspondence:** Alexandros Tsitsanis; Email: alexandros@suite5.eu

Abstract

The notion of "Energy Data Spaces" is gaining traction as a revolutionary framework within the energy industry, promoting the seamless integration, sharing, and analysis of data among various stakeholders. Endorsed by the European Commission Energy Data Spaces are designed to establish a cohesive digital ecosystem where data from diverse sources, including energy producers, consumers, and grid operators, can be collaboratively leveraged to streamline energy management and boost operational efficiency. This paper delves into the core principles, structural design, and prospective applications of Energy Data Spaces, emphasizing their pivotal role in fostering innovation and sustainability in the energy sector. We present a reference implementation of an Energy Data Space, highlighting the key features that must be addressed, such as data privacy, security, and standardization challenges.

Our findings demonstrate that this framework successfully enables secure, interoperable data exchange across heterogeneous systems, as validated through our collaboration with the **Electricity Authority of Cyprus (EAC)**. The results indicate that providing stakeholders throughout the energy data value chain with an Analytics Catalogue of adaptable AI services significantly optimizes grid operations and demand-side management.

Keywords: Energy Data Space; AI analytics; energy applications; energy data value; data sharing

Cite This Article:
Tsitsanis A, Lampathaki F. A comprehensive framework for data management and AI-driven analytics in the energy sector. Green Energy Sustain. 2026;6(3):0008. https://doi.org/10.47248/ges2606030008

Received: 14 Oct 2025
Accepted: 29 Jun 2026
Published: 10 Jul 2026

1. Introduction

The energy sector is undergoing a profound transformation driven by rapid digitization, fundamentally altering the production, distribution, and consumption of energy [1,2]. Two pivotal trends are at the heart of this shift: the development of **Energy Data Spaces (EDS)** and the integration of **Artificial Intelligence (AI)** analytics. Energy data spaces facilitate the secure, decentralized exchange of data among various stakeholders, while AI analytics harnesses this data to enhance operational efficiency, predict demand, and integrate renewable energy sources. These innovations promise to make energy systems more efficient, sustainable, and resilient. However, their widespread adoption is met with numerous challenges that must be addressed to fully realize this vision.

In more detail, Energy Data Spaces are a groundbreaking concept designed to enable secure and seamless data exchange among diverse stakeholders in the energy sector, including utilities, grid operators, consumers, and service providers. The essence of these data spaces lies in their ability to integrate disparate data sources - such as smart meters, grid sensors, and renewable energy generation systems - into a unified, interoperable framework. Initiatives like the International Data Spaces Association (IDSA) [3] and the GAIA-X [4] project are driving current developments, focusing on creating standardized data infrastructure that ensures data sovereignty, security, and compliance with data protection regulations. These platforms lay the groundwork for managing data flows securely while preserving control over data ownership and privacy, which are critical in the highly regulated energy industry.

In the field of AI analytics, Energy AI analytics is emerging as a transformative force, leveraging **Artificial Intelligence (AI) and Machine Learning (ML)** to optimize energy systems, enhance efficiency, and integrate renewable energy sources. AI-driven models are now employed for real-time demand forecasting, predictive maintenance, and grid optimization [5-7]. Recent advancements include the use of deep learning techniques to predict energy consumption patterns, which can improve load balancing and reduce energy wastage [8]. Additionally, reinforcement learning has been utilized to optimize energy storage systems and distributed energy resources (DERs), ensuring more efficient energy distribution and reducing reliance on fossil fuels during peak demand. These AI-based approaches not only boost operational efficiency but also contribute to the sustainability objectives of many nations.

Despite these advancements, significant hurdles remain fully harnessing the potential of Energy Data Spaces and AI analytics. A primary challenge is the absence of universally accepted standards for data interoperability across energy systems, which impedes cross-border and cross-sector data sharing. While standards like IEC CIM have facilitated interoperability in specific areas like substation automation, a comprehensive, standardized framework for the entire energy sector is still lacking. Cybersecurity is another major concern, particularly with the increasing frequency of cyberattacks targeting energy infrastructure. Ongoing research is exploring advanced encryption techniques, federated learning, and secure multi-party computation to safeguard data integrity while ensuring privacy. Moreover, the implementation of AI analytics in energy systems faces challenges related to data quality, the availability of large-scale, high-resolution datasets, and the need for algorithms that can adapt to the variability of renewable energy sources like wind and solar, which are influenced by unpredictable factors such as weather. Ensuring the cybersecurity and privacy of AI-driven energy systems is also crucial as these systems become increasingly interconnected and data-driven.

In this paper, we delve into the dual challenges associated with Energy Data Spaces and AI analytics in energy. We explore issues such as data interoperability, privacy, security, and governance that hinder collaboration and data sharing among stakeholders. Simultaneously, we address the hurdles faced by AI analytics, including data quality, access to comprehensive datasets, and the development of advanced algorithms tailored to the complexities of energy systems. The combination of these challenges can limit the effective integration of AI-driven insights into the energy sector. We then propose a framework for data management and analytics, offering a potential solution and a testbed environment to foster a more intelligent, data-driven energy ecosystem.

The remainder of this paper is organized as follows: Section 2 provides a state-of-the-art background and identifies current research gaps; Section 3 describes the proposed **Energy Data Space (EDS)** framework and addresses barriers to data sharing; Section 4 details the AI analytics layer and its application categories; Section 5 presents the real-world validation through demonstration sites; and Section 6 provides our conclusions and next steps.

2. A State-of-the-art AI Analytics Services in the Energy Sector

The analytics framework is designed to foster a culture of data-sharing and analytical thinking within the energy sector, empowering stakeholders across the energy data value chain with valuable insights to enhance their operations. Central to this initiative is the Energy Analytics Catalogue, which offers a comprehensive suite of flexible analytics services tailored to meet the specific needs of the energy domain. These services are structured to serve dual purposes:

- **Delivering Ready-to-Use Insights:** Providing immediate insights into common challenges faced by the sector, enabling stakeholders to make informed decisions swiftly.
- **Accelerating Customized Service Development:** Facilitating the rapid development of tailored analytics solutions to address the unique requirements of individual stakeholders.

The analytics services within the energy sector [9], as proposed by this framework, are categorized based on their application:

- **Occupant Behaviour and Comfort Profiling:** This category addresses the analysis of energy consumption patterns influenced by consumers' thermal and visual preferences. By understanding these patterns, energy providers can optimize comfort while reducing energy waste.
- **Energy Demand Prediction:** This service focuses on forecasting energy demand at various levels, from individual buildings to entire grids, and across different timeframes. Accurate demand prediction is crucial for efficient energy management and grid stability.
- **Energy Generation Prediction:** Here, the analytics aim to forecast energy generation from various sources, including renewable and traditional power plants, at different scales and time periods. This helps in balancing supply with demand and optimizing the integration of renewable energy sources.
- **Flexibility Forecasting:** This involves identifying segments of energy demand or generation that can be adjusted—either reduced, increased, or shifted—within a specified timeframe. Such forecasting is essential for demand response programs, energy storage optimization, and overall grid flexibility.

These categories are not rigidly defined, and more complex problems may span multiple categories. Often, solving a more advanced issue requires combining multiple baseline analytics solutions.

Background Overview

The current research landscape relevant to the analytics problem is rich with evolving methodologies and innovative approaches that aim to tackle complex challenges in data analysis and predictive modeling. Recent advancements indicate a significant shift towards integrating multi-source data, including real-time transactional records, unstructured social media data, and comprehensive sensor inputs, to construct more nuanced and accurate analytical models. Additionally, there is a growing emphasis on addressing critical issues such as data privacy, the ethical implications of AI decision-making, and the need for model interpretability to ensure transparency and trustworthiness. This dynamic research environment strives to harness cutting-edge algorithms and technological innovations to address pressing problems in the energy sector, and thus an exhaustive background analysis per model is provided.

The implementation involves utilizing diverse input data types, including raw and preprocessed datasets, to train and validate analytical models. Key steps include:

- **Data Preprocessing:** Addressing noise, missing values, or inconsistencies in raw data to ensure data quality.
- **Feature Extraction:** Identifying and extracting relevant features from the data to enhance model performance.
- **Model Selection:** Choosing appropriate algorithms based on the problem's nature, considering factors like interpretability, computational efficiency, and model complexity.
- **Model Configuration:** Adjusting parameters and hyperparameters during training to optimize performance.
- **Evaluation Metrics:** Assessing model performance using established metrics such as accuracy, precision, recall, F1 Score, and confusion matrix.

The technology stack includes:

- **Data Manipulation:** Pandas for handling tabular data, NumPy for numerical operations.
- **Analytics:** Scikit-learn for Machine Learning algorithms and tools.
- **Deep Learning:** TensorFlow and PyTorch for developing and training complex neural network models.

Key Challenges:

- **Data Quality:** Ensuring input data is representative and accurate to avoid incorrect predictions.
- **Generalization:** Models should perform well on unseen data, reflecting real-world effectiveness.
- **Model Complexity:** Balancing model complexity to prevent overfitting.
- **Scalability:** Addressing challenges related to the model's ability to scale efficiently with increasing data volume or features.

By categorizing these analytics services, the framework not only addresses immediate operational needs but also sets the stage for future innovations in energy management, promoting a data-driven approach to energy efficiency and sustainability.

3. A Comprehensive Framework for Data Management and Analytics

In response to the challenges and requirements outlined above, we introduce a reference big data management platform, termed the Data Space [10,11]. This platform integrates advanced data collection and harmonization services, data sharing and trading services, and is further complemented by an advanced analytics framework. This framework is enhanced by AI-driven analytics capabilities, which enable the generation of insights and derived data through the integration of near real-time building and grid level information with data from various external sources. The different layers of this framework are presented in detail below.

3.1. An innovative approach in the field of Energy Data Space

The Energy Data Space environment is a fundamental element of the overarching framework, emphasizing data governance and the integration of processing mechanisms as outlined in the literature.

The primary function of Data Governance within the Energy Data Space is to enable the collection of data from a diverse array of distributed energy and data sources on the prosumer side, facilitating their seamless exchange with energy value chain actors to realize added value services and optimize energy and flexibility management.

While the technical potential of such environments is clear, a primary barrier to implementation remains the traditional operational customs of the energy sector and the subsequent reluctance of grid operators or energy producers to share their data. This behavior is often driven by concerns over data sovereignty, competitive advantage, and the perceived risks of transitioning from established, closed operating techniques. To persuade these actors and counteract these barriers, the proposed framework shifts the focus from 'data ownership' to 'data usage control'. By ensuring a highly secure and sovereign data management framework, stakeholders are empowered to set specific, machine-enforceable access policies, ensuring they retain absolute control over their assets even after they are shared within the ecosystem. It is important to clarify that while this framework is designed to support the broader energy data value chain, including carriers such as gas and thermal energy, the primary focus of this paper and the subsequent validation is the electricity value chain. This focus is justified by the advanced digital maturity of the electricity sector, characterized by the widespread availability of high-frequency smart meter data and DSO telemetry.

The platform is designed to streamline the collection of data from various, yet dispersed, data assets through multiple ingestion methods, including batch, API-based, and real-time collection of data with varying volumes and velocities. Configuring appropriate data mapping mechanisms is essential to harmonize available datasets to a Common Energy Data Model. A configuration environment should be provided to users, allowing them to define cleaning rules for handling any incorrect, incomplete, inaccurate, irrelevant, or missing data parts. Anonymization features will enhance the plug-and-play nature of the solution, complemented by data owner-centric access policy enforcement rules. This ensures the establishment of a highly secure and sovereign data management framework around residential prosumers' data, addressing concerns regarding data privacy and confidentiality when making data available to traditional energy value chain stakeholders.

The Data Space should also prioritize providing advanced data exploration and retrieval features. Customized API creation for streamlining data retrieval and increasing its availability to a variety of access-granted stakeholders, including traditional energy market stakeholders, data scientists, and ICT application developers, enables prosumers to participate in energy market transactions in a data-driven manner. Incorporating data exploration mechanisms that enable visual exploration of data assets uploaded and ingested into the Data Space allows end-users to search for and discover datasets of interest for further use in business applications residing on top of the Data Space. Enabling retrieval of selected datasets through dynamically configured APIs based on querying parameters defined by end-users is also a key feature.

By considering these high-level specifications, the conceptual architecture overview is presented in **Figure 1**.

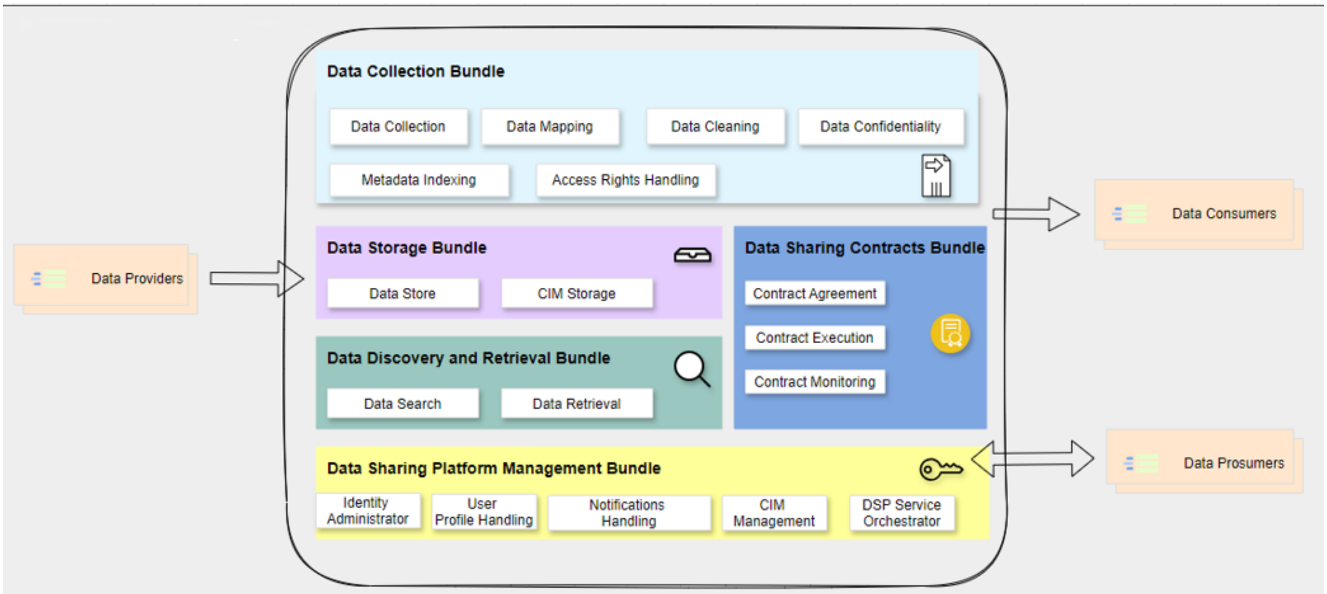

Figure 1. Data Space UML component diagram.

- **Data Collection:** This component is responsible for facilitating the timely and successful retrieval of data from various sources. It enables the acquisition of related data from DER connectors and other sources via data retrieval APIs or batch files provided by external parties. Once retrieved, the data is temporarily stored in data lakes, where it awaits further processing and analysis.
- **Data Mapping:** This component within the Data Governance is tasked with processing and transforming ingested data to align it with the common information model. It involves retrieving the configuration file for each data collection task from the Data Management Platform Orchestrator and performing the necessary transformations. The data, once transformed to conform with the CIM, is then stored in the Data Storage & Indexing module for efficient access and utilization.

- **Data Cleaning:** Utilizing Machine Learning (ML) algorithms, this component detects data outliers or missing information and subsequently fills these gaps with meaningful values inferred from a combined analysis of data owners' rules and historical data. This process can include manual completion by data owners if necessary, ensuring the integrity and accuracy of the dataset for reliable use.
- **Data Confidentiality:** This component facilitates experimentation with different data anonymization scenarios, allowing for the selection of the most effective method that balances privacy protection with the utility of anonymized datasets. This is achieved through the implementation of robust anonymization protocols, such as k-anonymity and noise addition techniques, ensuring that sensitive prosumer information remains protected while maintaining the dataset's statistical relevance for analytics.
- **Metadata Indexing:** This component ensures that all data, resulting from data collection, are systematically organized and categorized. Upon importing data, data providers must define a complete profile for each data asset. This includes providing a title for the data asset, as well as specifying its characteristics, such as the data type, format, language, temporal coverage, and temporal resolution units. These details are captured via a user-friendly interface, using a series of drop-down lists for ease of selection.
- **Access Rights Handling:** This component oversees the handling of access policy features, such as Attribute-based Access Control (ABAC). ABAC provides the necessary cybersecurity layer by ensuring that data access is granted only when a specific set of dynamically verified attributes (e.g., user role, purpose of use, and environmental conditions) are met, effectively protecting against unauthorized access or misuse.
- **Data Search:** This component enables advanced search functionality across the collected datasets by leveraging metadata annotations provided by the data provider. It supports quick and specific access to datasets, facilitates searching within datasets using filters, allows for efficient retrieval of data from the central storage module, and offers sample datasets for experimentation and testing purposes.
- **Data Retrieval:** This component builds upon the data discovery functionality offered by the Data Search Component, enabling users to execute requests and retrieve data efficiently. This component allows users to refine their requests and quickly assess whether the results meet their requirements. Specifically, it includes:

    1. Request Execution: Processing user's requests and returns relevant results, accompanied by additional information to help users easily identify the most suitable data assets. Users have the flexibility to manage and retrieve data assets individually or as part of a bundle.
    2. Search History, Request Updates, and Re-Execution: Users can save and re-execute past requests, making it simple to refine searches and access updated results as needed. This ensures users can easily revisit previous searches, modify them, and retrieve new data.

- **Contract Agreement:** This component is responsible for facilitating the creation and negotiation of data sharing contracts between data consumers and providers. Once a consumer identifies data assets of interest, they submit a request, which triggers the contract creation process. Providers can then draft a contract with predefined and customizable terms, including reimbursement details and specific conditions. The component supports negotiation between the parties, allowing adjustments to the contract terms before final agreement.

- **Contract Execution:** This component is responsible for the seamless implementation of the contract's terms following its mutual agreement and signing by both parties. It ensures that the data consumer fulfils the payment obligation, as specified in the contract, by processing the agreed payment method, such as cryptocurrency (e.g., ETH).
- **Contract Monitoring:** Oversees the entire lifecycle of a contract, ensuring compliance with its terms and conditions, including data transfer, usage, and reimbursement. It provides tools for monitoring and auditing the fulfilment of contractual obligations, alerting the parties involved if any issues arise. In cases of non-compliance, the component can initiate corrective actions to address them. Additionally, it manages contract renewals or modifications when necessary.

Along with the aforementioned functionalities, a series of overlay services is provided to support the overall functionality of the Data Space.

- **Identity Administrator:** This component is a crucial element of the Data Space Management Bundle, offering multiple layers of security, including identity verification and robust registration and authentication processes for platform users.
- **User Profile Handling**: This component allows users to manage and modify their personal details within the platform. Users can access their profile page and update essential information, such as their first and last name, and change their password by entering new credentials in the provided fields.
- **Notifications Handling:** This component allows users to receive notifications about key events related to their data collection pipelines. These notifications provide real-time updates on the progress of ongoing pipelines, highlighting significant milestones.
- **CIM Management:** Allows administrators to oversee and update the GridAI CIM within the platform, ensuring they meet the changing needs of data providers. It offers administrators the tools to manage and review the structure and content of the CIM, while ensuring that any modifications are tracked and updated appropriately. Administrators are notified when new concepts are suggested, facilitating prompt updates and maintaining the model's relevance.
- **Data Sharing Platform Service Orchestrator:** This component orchestrates and manages the execution of services related to data gathering, cleaning, anonymizing, handling, storing, searching, and provisioning. It interfaces with all system modules, facilitates message exchanges, allocates resources for data processes, and monitors the operational status of each module. Additionally, it acts as a logging system, capturing the various states and activities to ensure transparency and traceability.

In summary, the Energy Data Space environment aims to streamline the collection, mapping, and secure exchange of diverse energy data between prosumers and energy value chain stakeholders. Key features include data mapping to a Common Energy Data Model, data cleaning, and confidentiality to ensure privacy and anonymization. The proposed framework also offers advanced data exploration and retrieval tools, enabling users to configure queries and APIs for tailored access to data. This fosters secure, intelligence-driven participation in energy markets and supports value-added services.

3.2. An advanced AI analytics framework for the energy sector

The data analytics layer [10], as described in this section, provides a suite of analytics services in the form of baseline intelligence services to support operational and business optimization functions across the energy value chain. It leverages raw data from sensors, metering devices, and other sources within the energy system and related ecosystems to extract valuable insights and intelligence. These insights are crucial for Energy Service Providers (ESPs) to optimize their processes and enhance their service offerings.

A common barrier to the adoption of advanced analytics frameworks is the potential requirement for specialized data analytics expertise. To address this challenge, the proposed framework is designed to support users with different levels of technical knowledge. Building upon the EDS functionalities for data collection, mapping, cleaning, confidentiality, users are provided with prepared and interoperable datasets suitable for analytics tasks. Pre-processing operations such as data quality improvement, missing-value handling, outlier treatment, format standardization, and alignment to the Common Information Model are performed through the underlying Data Space services. Once these steps are completed, the Energy Analytics Catalogue provides ready-to-use analytics services for common energy use cases such as load forecasting, anomaly detection, flexibility estimation, and descriptive analytics. In parallel, the analytics layer enables users to design and execute pipelines through configurable components, reusable functions, and predefined analytics models, while visualization and monitoring services provide clear feedback on results and pipeline execution. In this way, both technical and non-technical stakeholders can exploit advanced analytics services according to their operational needs without requiring extensive data science expertise. Furthermore, the integration of Interactive Data Visualization components allows users to explore complex patterns and forecast results through intuitive interfaces, significantly lowering the technical barrier to entry for diverse energy actors.

Similar to the core Energy Data Space implementation, the Data Analytics Services Bundle, as presented in **Figure 2**, comprises several components, each with specific functionalities:

- **AI Pipelines design:** This component helps users build, configure, and execute data analysis workflows tailored specifically for energy data. It offers a user-friendly interface that enables data scientists to easily engage with complex data processing tasks.
- **Data Exploration**: This component offers an intuitive and interactive platform for users to visually analyze selected data assets, enhancing their ability to uncover valuable insights. By engaging in this service, users can apply filters and focus on specific data subsets. The service provides a wide range of visualization tools that clearly display data relationships, trends, and patterns.
- **Data Analytics Execution**: This component handles the execution of data analysis jobs based on configurations created by the AI Pipelines Design. It enables users to run diverse data processing workflows by combining various blocks, which serve as wrappers for analytics models defined in the pipeline.

- **Interactive Data Visualization**: This component plays a crucial role in the Energy Data Analytics Services by enabling users to visually explore insights from specific analytics results. This component allows users to create custom visualizations that highlight interesting data of the analytics pipelines they have designed and executed.

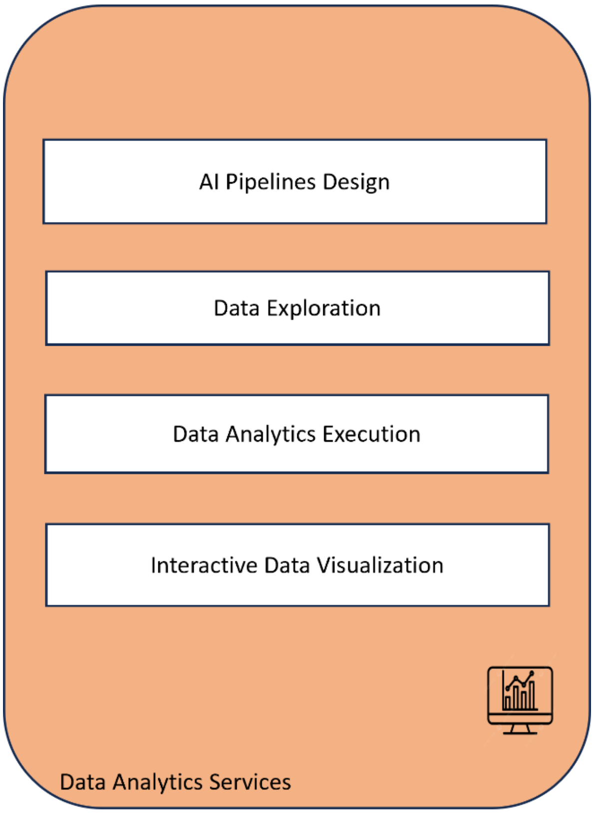

Figure 2. Data analytics Services UML component diagram.

The Data Analytics Services Bundle is designed to seamlessly integrate with the Energy Data Space environment, leveraging the secure, standardized, and harmonized data available within the data space. This integration ensures that analytics services adhere to the same data privacy and security standards as the Energy Data Space, ensuring that sensitive data is protected during analysis. This is crucial given the increasing interconnectivity of smart grids and IoT devices, which introduce more entry points for potential data breaches. Users maintain control over their data, with access policies enforced to prevent unauthorized use or sharing, aligning with the principles of data governance frameworks that define clear roles and responsibilities for data management. Additionally, the analytics framework can interface with various data sources and systems, promoting a holistic view of the energy ecosystem. This is essential for addressing the challenges posed by the lack of industry-wide interoperability standards, which currently hinder the full potential of consumer flexibility solutions.

Data Storage is a core bundle for both the Energy Data Space and the Data Analytics Services Bundle, responsible for persistently storing and archiving the large volumes of data generated. Transformed data from the data handling process, as well as the analytics results data (together with the associated metadata), are stored in a non-relational (NoSQL) database, chosen for its scalability and optimized management of diverse datasets. Analytics and AI models defined and configured are stored in a separate database environment to support the execution of the analytics services. A data indexing engine is utilized to facilitate efficient search and retrieval of the data. The definition of the Common Information Model (CIM) and its details are an inherent part of the Data Space. To enhance the security of the data, the credentials to access the platform are stored in a dedicated environment. More details about the Data Storage Bundle of the overall platform are provided in **Figure 3** below, where the different functional elements comprising the Data Storage component are highlighted.

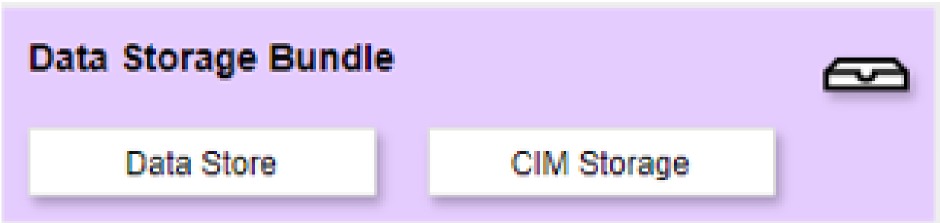

Figure 3. Data storage layer of the proposed framework.

The Data Storage Bundle ensures **reliable and resilient data storage** through diverse solutions, each tailored to specific data types and retrieval methods. It further enhances data protection with robust **Backup and Recovery Services**. These services are meticulously created, managed, monitored, and supported for applications, files, and servers. This comprehensive approach safeguards critical data against security breaches or disasters, minimizing business disruption by enabling rapid restoration. The selection of appropriate large-scale backup and recovery solutions is paramount to upholding data safety and security.

- **Data Store:** The Data Store component is responsible for securely and reliably storing a wide variety of data and their associated metadata. It incorporates all relevant storage and indexing tools, focusing on the following:
    1. Data storage: Safely storing data assets and their related metadata to ensure availability across all bundles and components.
    2. Log data storage: Storing log-related information for platform operations, including data from users and organizations, as well as administrative data necessary for the data space's seamless functioning
- **CIM Storage:** This component is responsible for securely storing and managing the GridAI CIM. This component ensures that the CIM is properly versioned, accessible, and aligned with all platform data activities and is responsible for:
    1. CIM Versioning: Ensures that users always access the most up-to-date version of the GridAI CIM.
    2. Data Model Definition: Stores the structure and definitions of the CIM, including relevant concepts and fields that standardize data across the platform.

3. Security and Integrity: Protects the CIM from unauthorized changes, ensuring the model remains accurate and reliable for all platform users and data operations.

4. Real Life Testing, Validation and Next Steps

GridAI [12] introduces a proposed end-to-end solution for managing and analyzing big data, primarily or secondarily related to the electricity domain and DSO Operations. This data will come from diverse sources such as historical energy demand and generation data from open sources, weather data, energy market data, and other relevant datasets. The solution is designed to assist Distribution System Operators (DSOs) in effectively and interoperably managing the data they own. It will also enable the integration of external data required for driving operational decisions (e.g., weather data, energy market data) and combine this data to enhance internal intelligence. Ultimately, this will support data-driven decision-making for the resilient operation of the distribution grid by leveraging accurate AI-based demand and generation forecasts across varying spatial and temporal resolutions. In GridAI, extracting valuable intelligence from the distribution grid data assets, carrying important electricity-related information requires a novel approach.

The demonstration phase [13] of the GridAI solution adopts an iterative, user-centric approach that ensures both technological robustness and practical applicability. This phase encompasses a structured sequence of activities aimed at validating GridAI in real-world conditions while ensuring continuous alignment with stakeholder needs, regulatory compliance, and economic viability. The process begins with a detailed elicitation of requirements, grounded in the explicit needs of end-users. These are enriched by internal knowledge derived from prior projects and collaborative efforts with Distribution System Operators (DSOs) and other relevant stakeholders across the energy value chain. Critical regulatory frameworks—including the EU Data Act and GDPR—as well as business drivers and economic considerations are systematically integrated into the design process. This enables early identification and mitigation of associated risks.

Subsequent efforts focus on the detailed design and specification of GridAI’s technological components and services. This involves the continuous refinement of functional and non-functional requirements, supporting the iterative development of a fully integrated prototype. Proven technologies are enhanced and incorporated to optimize GridAI’s overall performance and ensure compatibility with operational, regulatory, and business objectives.

Demonstration activities are deployed across a diverse and representative set of environments, designed to test the full spectrum of GridAI’s innovations and use cases. This broad coverage enhances the replicability of the solution and is complemented by a structured demonstration and cross-replication plan, followed by a dedicated replication campaign to further validate its applicability in varied contexts.

A cornerstone of the validation strategy is the collaboration with the Electricity Authority of Cyprus (EAC), the national Distribution System Operator (DSO), which serves as the primary demonstration partner. This strategic partnership enables hands-on experimentation within a real operational environment, as shown in **Figure 4** offering a realistic and representative setting for the rigorous testing and validation of GridAI’s functionalities. The demonstration will leverage a diverse subset of EAC’s metering infrastructure, incorporating data from various metering endpoints, including conventional and smart meters installed at customer premises, photovoltaic (PV) generation systems, and grid connection points. This comprehensive data input enables the evaluation of GridAI under heterogeneous operating conditions and supports the validation of key functionalities such as load forecasting, anomaly detection, and descriptive analytics. The inclusion of data from distributed energy resources, such as PV systems, further enhances the relevance and granularity of the validation, ensuring that GridAI can effectively support real-time grid monitoring and decision-making in a modern, increasingly decentralized energy landscape.

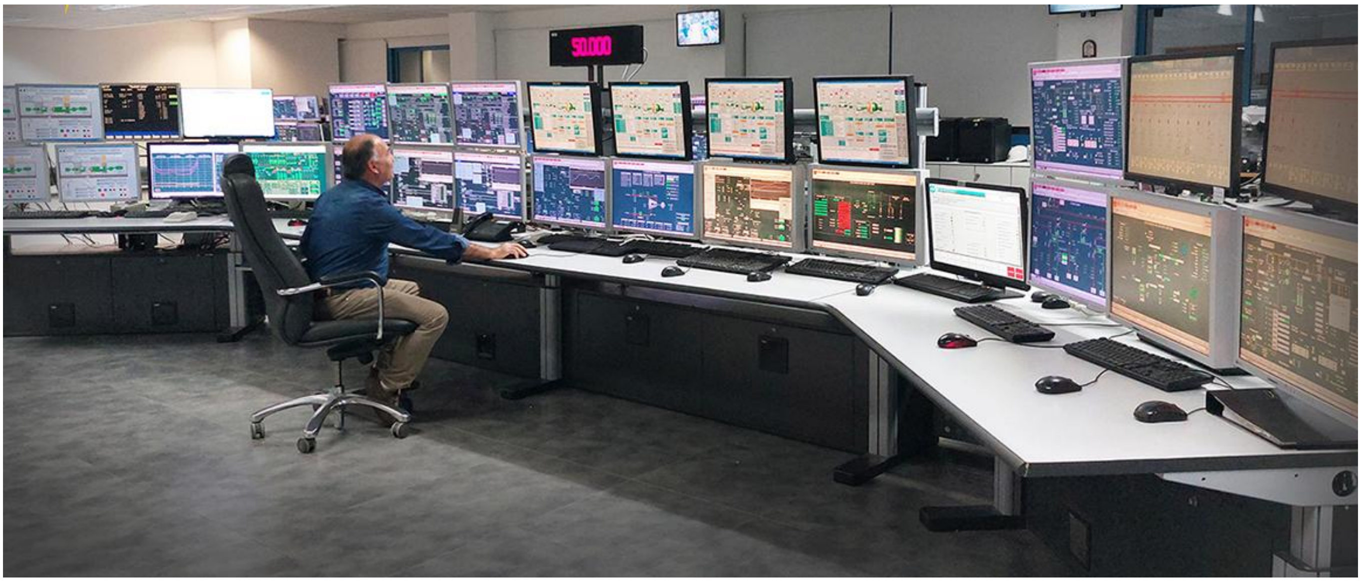

Figure 4. Demonstration sites and activities – EAC Scada System.

In parallel to the live demonstration activities, unsupervised beta testing is performed to assess GridAI’s technical performance, operational efficiency, and scalability. A critical element of this process is the validation of the AI pipelines through extensive training using publicly available distribution grid and metering datasets. These trained models support the generation of descriptive analytics at both grid and substation levels, delivered via purpose-built dashboards designed to provide meaningful insights to system operators and stakeholders.

Collectively, these demonstration activities ensure that GridAI emerges as a reliable, scalable, and market-ready solution, validated under realistic conditions and informed by continuous feedback from key industry actors.

5. Conclusions & Future Directions

This paper has explored the motivation, design, and implementation of an integrated Data Space implementation tailored to the energy sector's needs. The platform is designed to provide a comprehensive solution for organizations seeking to optimize their operations and extract valuable insights from their data. Key features include:

- Enhanced Data Access: Users can access real energy data and enrich it with external sources, ensuring efficient and transparent data management.
- Secure Data Sharing: An intuitive marketplace facilitates standardized data access, promoting secure data sharing among stakeholders.
- Insightful Analytics: The platform offers pre-trained models and custom-built algorithms for direct analytics, empowering users with actionable insights.

The paper has outlined the different service bundles and architectural layers of the platform, demonstrating its integration with various software tools and applications relevant to buildings, communities, and energy market stakeholders. It is worth noting that the proposed big data and analytics framework is currently undergoing validation in four large demonstration sites across Spain, Greece, Switzerland, and Bulgaria. These sites encompass a diverse range of building types, including offices, commercial spaces, and residential buildings, involving various users, data sources, energy systems, and assets. The ongoing demonstrations aim to:

- Highlight Platform Capabilities: Showcase the platform's ability to handle real-world energy data and analytics challenges.
- Demonstrate Value: Provide tangible evidence of the platform's benefits in optimizing energy management and market participation.
- Support the Vision: Validate the overall vision of the proposed framework, demonstrating its potential to drive innovation and sustainability in the energy sector.

In conclusion, this paper has presented a robust framework for data management and Artificial Intelligence (AI) analytics in the energy sector, validated through real-world applications. The framework not only addresses current challenges but also sets the stage for future innovations in energy data management and analytics, promoting a data-driven approach to energy efficiency and sustainability.

Declarations

Availability of Data and Material

The pilot data supporting the findings of this study will not be shared because the datasets are preliminary, confidential, and intended solely for internal study validation.

Funding

The Project is co-funded by the European Regional Development Fund and the Republic of Cyprus through the Research and Innovation Foundation (RIF), under the project protocol number ENTERPRISES/ENERGY/1123/0018.

Competing Interests

The authors of the manuscript have declared that no competing interests exist.

Author Contributions

Conceptualization: A.T. and F.L.; Methodology: A.T. and F.L.; Software: A.T. and F.L.; Validation: A.T. and F.L.; Formal analysis: A.T. and F.L.; Investigation: A.T. and F.L.; Resources: A.T. and F.L.; Data Curation: A.T. and F.L.; Writing - Original

Draft: A.T.; Writing - Review & Editing: A.T. and F.L.; Visualization: F.L. and A.T.; Supervision: F.L.; Project administration: F.L.; Funding acquisition: A.T. and F.L.

Acknowledgments

The work presented in this paper is carried out in the context of the GridAI project (Project Protocol Number: ENTERPRISES/ENERGY/1123/0018), highlighting the key insights of the Data Analytics Services Bundle. The project demonstrates significant benefits through enhanced data management and sharing, emphasizing the development of an analytics layer that incorporates adaptable and comprehensive analytics services designed to meet the diverse and evolving needs of the energy sector, fostering a culture of data-driven decision-making and innovation.

References

1. Heinze T, Splieth P, Möller J, Mikut R. Towards energy data ecosystems: Challenges and potentials of data spaces in the energy domain. arXiv preprint arXiv. 2021;2107.01965.
2. Anwar MF, Guerrero JM, Muyeen SM. Smart grid and renewable energy integration using blockchain and AI. arXiv preprint arXiv. 2022;2202.10098.
3. International Data Spaces Association. International Data Spaces Association [Internet]. IDSA; [cited 2026 Jul 1]. Available from: https://internationaldataspaces.org/.
4. GAIA-X. GAIA-X: A Federated Data Infrastructure for Europe [Internet]. GAIA-X; c2026 [cited 2026 Jul 1]. Available from: https://gaia-x.eu/.
5. Ahmed IA, Asamoah PB. AI-driven predictive maintenance for energy infrastructure. Int J Res Sci Innov. 2024;11(9): 507-528. DOI
6. Pal G. IoT and AI for Smart Grid Optimization in Urban Areas. Soft Comput Fusion Appl. 2025;2(1):54-61. DOI
7. Mastoi MS, Wang D, Ma N, Hassan M, Shafiullah M, Bashir T, et al. AI-driven control and optimization for renewable energy integration in smart grids: Challenges, applications, and future research directions. Energy Strategy Rev. 2026;64:102049. DOI
8. Pérez MV, Sánchez D, Moreno MN, Martínez P. A systematic review of artificial intelligence in the energy sector: Applications and research trends. Renew Sustain Energy Rev. 2020;136;110535.
9. Zhang Z, Zhang C, Wang H, Li D. Digital twins in smart grid: A review of applications, challenges, and future directions. Renew Sustain Energy Rev. 2022;159:112239.
10. European Commission. Data Governance Act: Proposal for a Regulation of the European Parliament and of the Council on European data governance. Brussels: European Commission; 2020.
11. European Commission. EU Data Act: Proposal for a Regulation on harmonised rules on fair access to and use of data. Brussels: European Commission; 2022.
12. GridAI [Internet]. Limassol: GridAI [cited 2026 Jul 1]. Available from: https://gridai.eu/.
13. GridAI Consortium, GridAI D10 GridAI Requirements and Detailed Technical Design. GridAI Consortium; 2024. Project Protocol Number: ENTERPRISES/ENERGY/1123/0018 ENTERPRISES/ENERGY/1123/0018Internal project documentation.