Building a European-Level Energy Registry

Mahsa Faraji Shoyari[1], Jessy Matar[1] and Markus Duchon[1]
[1]fortiss GmbH, Munich, Germany
E-mail(s): farajishoyari@fortiss.org, matar@fortiss.org, duchon@fortiss.org

Abstract: The increasing integration of renewable energy across Europe demands a scalable and harmonized digital infrastructure to manage large volumes of energy-related data. This paper introduces the concept of a European Energy Registry (EER), aimed at establishing a framework that connects data providers, ensures semantic and structural integration of heterogeneous datasets, and exposes validated datasets to users. We examine the technical challenges in developing such a platform and outline a pipeline that leverages big data management, machine learning, and secure IT infrastructure. Key components include accurate data registration from diverse stakeholders, verification and correction mechanisms, and a robust, role-based user access system enhanced with multi-factor authentication and General Data Protection Regulation (GDPR) compliant security protocols. An API-driven data access model is proposed to enable seamless retrieval, filtering, and analysis of energy data. Automated validation and cleansing tools will support data quality and regulatory compliance. The underlying infrastructure will accommodate scalable databases, pipelines, and AI-driven analytics to provide predictive insights. Cloud-native deployment and APIs will ensure high availability, interoperability, and integration with external platforms. By addressing these requirements, the EER will support Europe's energy transition plan. This work presents a comprehensive technical blueprint for a secure, interoperable, and future-ready energy registry.

Keywords: Energy Registry, Data Management, Machine Learning, Interoperability

DOI: 10.18154/RWTH-CONV-254884. Published in the conference proceedings of the 36. Forum Bauinformatik 2025, Aachen, Germany,
© 2025 The copyright for this article lies with the authors. This publication, except for quotations and otherwise indicated parts, is licensed under a Creative Commons Attribution 4.0 International (CC BY 4.0) license.

1 Introduction

Europe's transition to a decentralized and sustainable energy system has led to a surge in renewable energy integration. While supporting climate goals and market liberalization, this shift generates vast, fragmented, and heterogeneous datasets across national systems. Inconsistent standards and siloed data hinder cross-border collaboration, planning, and decision-making [1].

To address these challenges, a scalable, secure, and interoperable digital infrastructure is needed, one that supports seamless data exchange, integration, and validation across national and organizational boundaries. Recent advancements in big data management, machine learning (ML), and cloud-native technologies offer powerful tools to meet these requirements [2], [3].

A notable step in this direction is the NEED ecosystem [4], which lays the groundwork for a robust and flexible energy data space in the context of municipal and regional energy planning. NEED integrates heterogeneous data sources via semantic models, employs synthetic data generation techniques, and exposes standardized APIs through a knowledge graph-based infrastructure. It emphasizes findability, accessibility, interoperability, and reusability (FAIR) data principles, data virtualization, and tool interoperability, demonstrating a scalable and modular approach to planning centric data integration.

Building upon the architectural and conceptual foundations of NEED, our work extends this vision to the European level by proposing a European Energy Registry, a cross-border digital infrastructure designed to standardize and streamline the access of energy data across member states. The EER aims to serve as a foundational component of the EU's energy data space, enabling trustworthy and interoperable data sharing among regulators, planners, researchers, and market actors.

Our proposed framework is aligned with the principles of Gaia-X and introduces additional energy-domain-specific capabilities. This work's key contributions include the introduction of a virtualization layer to enable on-demand, call-based data access, enhancing flexibility and reducing data duplication (1); the deployment of a workflow orchestration layer comprising automated data pipelines, cloud-native services, and monitoring mechanisms (2); the development of an OGC standard compliant spatial data access layer to support infrastructure planning and geospatial analytics (3); the integration of identity management, federated connectors, and secure access protocols to ensure data reliability and trust (4); the implementation of semantic interoperability tools such as ontologies, metadata catalogs, and knowledge graphs to link heterogeneous data sources meaningfully (5); and the use of synthetic data generation to address data availability and gaps where necessary (6).

The rest of the paper is structured as follows: Section 2 reviews related work; Section 3 presents the EER architecture; Section 4 outlines discussion and conclusion.

2 Related Work

The European energy sector struggles with fragmented data systems and complex regulations. Recent policy efforts and technological advances offer pathways to a more integrated and sustainable infrastructure. The EU Data Act (2024) enhances data transparency and usability, while the GDPR enforces strict rules on personal data use, critical for demand-side and retail energy applications. GDPR compliance is essential for building trust in AI-powered energy solutions [5]. To address these challenges, Europe is fostering sovereign, interoperable, and federated data spaces that enable secure, lawful, and transparent data sharing. Multiple initiatives are shaping this ecosystem through shared architectures and governance models.

2.1 Gaia-X and the International Data Spaces Association

Gaia-X is a European initiative for a federated, secure, and interoperable data infrastructure. In partnership with the International Data Spaces Association (IDSA), it supports cross-sector data ecosystems through domain-specific data spaces. IDSA contributes to the Dataspace Protocol, a modular architecture for secure, standards-based data exchange, emphasizing data sovereignty and semantic interoperability [6]. The IDS Information Model defines roles and policies, while the IDS Connector manages identity, contracts, and secure communication. Together, Gaia-X and IDSA enable scalable and trusted data ecosystems [7], [8]. Energy data-X, a Gaia-X lighthouse project funded by BMWK, develops a federated energy data space supporting decentralized data exchange via standardized connectors and identity tools. Use cases include grid optimization and integration of flexible assets like EV chargers and heat pumps. OMEGA-X expands this vision at the European level, connecting energy domains-electricity, gas, mobility, smart cities through a secure, interoperable infrastructure. It promotes data sovereignty and cross-sector innovation with real-world pilots, supporting the energy transition and advancing a unified EU data space.

2.2 FAIR and Open Data

Germany's national research data infrastructure (Nationale Forschungsdateninfrastruktur - NFDI) fosters FAIR data practices like making data FAIR across scientific domains. The FAIR data spaces project expands this vision to industry, enabling secure, cross-sector data collaboration in line with Gaia-X principles. Use cases in health (NFDI4Health) and energy (NFDI4Energy [9]) demonstrate how distributed analytics and rich metadata can support innovation without compromising data sovereignty or privacy.

Complementing this, open data which is freely accessible and reusable by anyone promotes transparency, collaboration, and equitable access to knowledge, especially in resource constrained settings. While FAIR data doesn't necessarily mean "open," it ensures that data is shared under clear, well-defined conditions, balancing openness with the need to protect sensitive information.

Together, FAIR and open data approaches contribute to a federated, trustworthy, and inclusive data ecosystem [10].

While the above initiatives offer broad data-sharing frameworks, our blueprint addresses the European energy sector's complexity with advanced validation, AI analytics, and rich semantic metadata. This ensures a scalable, unified platform for efficient data management and the energy transition.

3 System Architecture of the EER

We present in this section, the core components that form the foundation of the proposed EER. We outline each component detailing the key building blocks of this framework starting from the southbound layer, followed by the middle and northbound layers (as seen in Figure 1 from bottom up).

3.1 Southbound Layer

The Southbound Layer serves as the primary interface for ingesting data from a wide range of heterogeneous distributed sources across the energy ecosystem. This layer must be designed to accommodate the diversity of formats, protocols, and update frequencies associated with different actors.

Aggregating different data sources forms the foundation for accurate analytics [11]. Among the primary energy data providers are Transmission System Operators (TSOs), who supply information on cross-border electricity flows, system frequency, and grid balancing operations. Distribution System Operators (DSOs) contribute detailed data at the local level, including voltage levels, transformer load status, and fault or outage events. Generation units, both renewable and conventional, feed the registry with production forecasts and output. At the consumption end, smart meters installed across residential, commercial, and industrial sites provide high-resolution usage data, enabling precise load profiling. In addition to these core contributors, energy market operators and exchanges provide

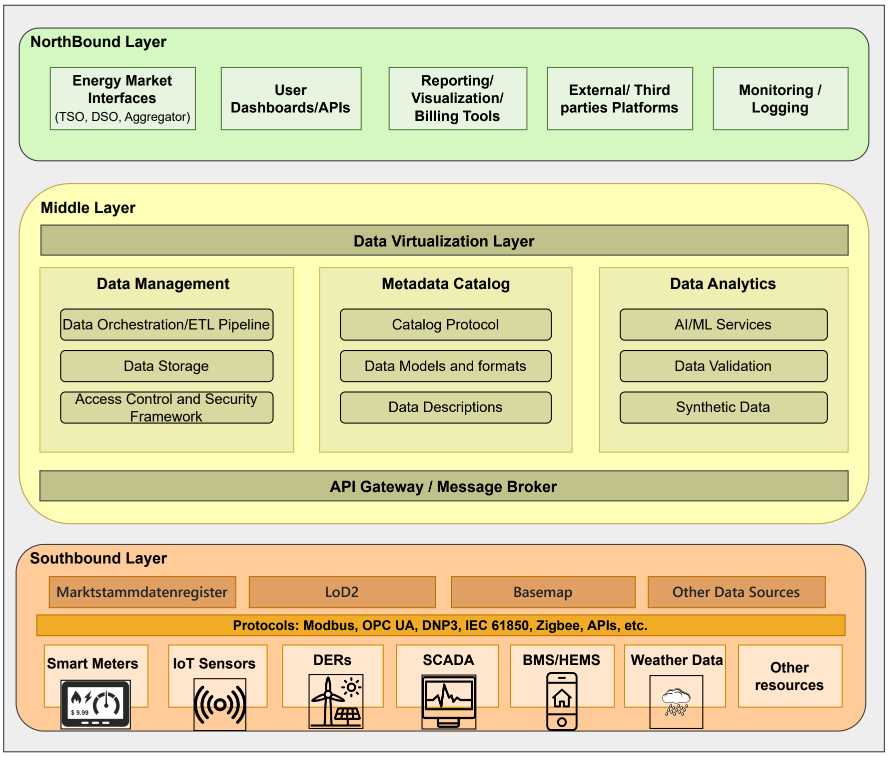

Figure 1: EER Components.

essential market data, such as pricing signals and transaction volumes from various trading platforms (e.g., day-ahead, intraday, and balancing markets). Meteorological data from weather services plays a key role in enabling accurate forecasting, particularly for variable renewable sources like wind and solar. Data from regulatory reporting systems and Virtual Power Plants (VPPs) also enrich the platform. This diverse and dynamic data landscape highlights the critical need for standardized data formats, secure communication protocols, and automated validation tools within the southbound interface of the EER, with adherence to OGC standards and the OGC API further enhancing interoperability across geospatial energy data sources.

3.2 Middle Layer

The Middle Layer acts as the processing core of the EER, responsible for tasks such as data connection and enrichment. It also manages metadata and ensures compliance with privacy and security standards.

3.2.1 Data Virtualization

Traditional data architectures, relying on centralized data lakes and warehouses, struggle to keep up with the complexity and distribution of modern data ecosystems. With data scattered across relational and NoSQL databases, graph stores, cloud filesystems, and SaaS platforms, centralizing all sources introduces redundancy, latency, and operational overhead. The EER addresses these challenges through a data virtualization approach. Rather than physically consolidating data, virtualization enables on-demand access to distributed sources through a unified interface [12]. This minimizes duplication, reduces latency, and ensures compliance. The platform architecture is structured around three key components: (1) Data Management, which handles orchestration, storage, and secure access, (2) Metadata Catalog, which ensures semantic alignment and discoverability of data, (3) Data Analytics, which supports insights, validation, and synthetic data generation. To support these components, the platform emphasizes interoperability and scalability. Interoperability is essential for seamless data integration across systems. This project focuses on APIs, standards, semantic technologies, and Service Based Architecture (SBA) as key pillars. In parallel, containerization technologies like Docker and Kubernetes are fundamental to ensuring scalability and service accessibility. In the EER, Kubernetes orchestrates containerized components, automating deployment, scaling, and life cycle management.

1. **Data Management**

   Data Orchestration and ELT Pipelines In modern data infrastructures, orchestration and transformation tools are essential for building efficient and scalable data workflows. In the EER project, we adopt an ELT (Extract, Load, Transform) approach using Dagster and dbt (Data Build Tool). Dagster orchestrates the extraction process, connecting to diverse data sources and extracting only the required subsets, avoiding full data downloads and reducing resource overhead. dbt performs transformations directly within the target system, ensuring consistency and performance. This separation of concerns enhances pipeline transparency and simplifies maintenance while enabling distributed data integration.

   Data Storage While data virtualization minimizes the need to copy or replicate datasets, temporary or persistent storage is still necessary for analytical processing, caching, and performance optimization. Our architecture supports a hybrid storage model, including: (a) Relational and NoSQL databases for structured and semi-structured data, (b) Object storage solutions like MinIO for large binary or unstructured data, (c) Software-defined storage systems such as Ceph to ensure scalability and high availability in containerized environments. This flexible approach allows the platform to scale horizontally, handle diverse data types, and optimize cost-performance trade-offs across workloads.

   Access Control and Security Framework Given the presence of sensitive data especially in the southbound layer, robust access control and security mechanisms are critical [13]. Each data source enforces its own access policies, but the platform also implements a middleware security layer. We use Keycloak for centralized authentication and authorization, enabling role-based access control (RBAC) across services. This ensures that only authorized users and systems can interact with protected resources, in compliance with data protection regulations like GDPR.

2. **Metadata Catalog**
   Catalog protocol is central to the semantic, ontology driven framework upon which the metadata catalog of the EER would be based. This approach emphasizes structured vocabularies and relationships that formally define the meaning of data entities.

   Data models such as the CIM (Common Information Model), SAREF (Smart Appliances REFerence ontology), and the European Union's Smart Data Models enable consistent representation of concepts like energy generation, consumption, grid assets, and market roles across national systems. Leveraging these models supports integration across heterogeneous energy datasets in Europe.

Data descriptions in the planned semantically enriched metadata catalog would use JSON-LD formatting and support SPARQL based querying and linked open data integration. This design allows datasets from different sources to be linked via shared concepts, facilitating automated data integration, validation, and compliance with regulatory frameworks.

3. **Data Analytics**

   Artificial intelligence and ML services can significantly enhance the utility of energy data. ML can go beyond traditional monitoring by learning complex patterns in consumption, generation, and grid conditions, enabling early detection of faults, more accurate renewable generation forecasts, and optimized balancing strategies. For example, ML-based predictive models can anticipate demand surges or wind power fluctuations hours ahead, allowing for proactive operational decisions that improve grid resilience.

   Data validation is necessary to ensure the accuracy and reliability of insights generated by ML models. Automated checks embedded in the data pipeline would assess completeness, consistency, and plausibility using a combination of rule-based logic, statistical profiling, and anomaly detection algorithms. In critical scenarios such as suspected sensor malfunctions a human-in-the-loop review can provide an additional safeguard, ensuring operational decisions are based on verified data.

   Synthetic data become important in scenarios where real-world data is scarce, incomplete, or sensitive. Techniques such as Generative Adversarial Networks (GANs) or agent-based simulations [4] can generate such data. Synthetic datasets also support cross-border experimentation by providing standardized test cases that reflect the statistical properties of actual energy datasets [14].

3.2.2 API Gateway/Message Broker

To support secure and efficient communication in a distributed, service-based architecture, the platform integrates both an API Gateway and a Message Broker. The API Gateway acts as a single entry point for external and internal API requests, handling authentication, routing, rate limiting, and protocol translation, ensuring secure and consistent access to microservices. Meanwhile, the Message Broker (e.g., Kafka or RabbitMQ) facilitates asynchronous, decoupled communication between components, enabling high-throughput data exchange, buffering, and event-driven processing.

3.3 Northbound Layer

The Northbound Layer serves as the interface between the energy data platform and end-users. It provides user-facing tools like dashboards, APIs, and visualization services, enabling seamless access to processed and validated data. Additionally, it facilitates integration with third-party platforms and supports monitoring and logging functionalities. This layer ensures that insights derived from underlying data systems are securely and efficiently delivered to relevant stakeholders for decision-making and operational use.

4 Discussion and Conclusion

The EER is an ambitious project that comes with a wide range of challenges. Designing and implementing a registry that can handle Europe's diverse energy data landscape is highly complex, from data integration to security and governance, to long-term scalability. Each energy actor generates data in its own format, often using different communication standards. Harmonizing such inputs requires sophisticated ETL pipelines and hybrid storage architectures that can evolve over time. Much of the data in question relates to critical infrastructure or sensitive operational behavior. The EER must enforce strict data protection through end-to-end encryption to ensure transparency and accountability. In addition, energy data is prone to inconsistencies, such as missing values, and errors stemming from device faults or communication failures. Hence, it is critical to implement robust ML integration for validation and data cleansing. Finally, the volume, velocity, and variety of energy data are expected to grow rapidly as more distributed energy resources come online. The EER must therefore adopt cloud-native architectures and scalable data processing pipelines to handle these future demands while ensuring low latency and high availability across a European infrastructure.

The proposed European Energy Registry represents an important step towards creating a more integrated and transparent energy data ecosystem across the EU based on a detailed reliable metadata catalog. As we have seen, the path to implementing the EER involves a high degree of complexity. Ultimately, the EER should be viewed as a long-term initiative that must remain flexible and open to future innovations in data management, artificial intelligence, and grid technologies. As the European energy landscape continues to evolve, so the infrastructure that supports it. The framework outlined in this paper provides a foundation for such an effort, but much work remains to turn this vision into a practical and widely adopted system.

Acknowledgements

We gratefully acknowledge financial support through the project executing agency Jülich (PTJ) with funds provided by the Federal Ministry for Economic Affairs and Climate Action (BMWK) due to an enactment of the German Bundestag under Grant No. 03EN3077%.

References

[1] R. Verzijlbergh, L. Vries, G. Dijkema, and P. Herder, "Institutional challenges caused by the integration of renewable energy sources in the european electricity sector", *Renewable and Sustainable Energy Reviews*, vol. 75, Nov. 2016. DOI: 10.1016/j.rser.2016.11.039

[2] G. Rostirolla et al., "A survey of challenges and solutions for the integration of renewable energy in datacenters", *Renewable and Sustainable Energy Reviews*, vol. 155, no. C, 2022. DOI: 10.1016/j.rser.2021.111787 [Online]. Available: https://ideas.repec.org/a/eee/rensus/v155y2022ics136403212101056x.html

[3] S. Onwusinkwue et al., "Artificial intelligence (ai) in renewable energy: A review of predictive maintenance and energy optimization", *World Journal of Advanced Research and Reviews*, vol. 21, pp. 2487–2799, Jan. 2024. DOI: 10.30574/wjarr.2024.21.1.0347

[4] M. Duchon et al., "A platform ecosystem providing new data for the energy transition", *SIGENERGY Energy Inform. Rev.*, vol. 4, no. 4, pp. 226–237, Feb. 2025. DOI: 10.1145/3717413.3717435 [Online]. Available: https://doi.org/10.1145/3717413.3717435

[5] B. Jørgensen and Z. Ma, "Regulating ai in the energy sector: A scoping review of eu laws, challenges, and global perspectives", *Energies*, vol. 18, p. 2359, May 2025. DOI: 10.3390/en18092359

[6] A. Braud, G. Fromentoux, B. Radier, and O. Le Grand, "The road to european digital sovereignty with gaia-x and idsa", *IEEE network*, vol. 35, no. 2, pp. 4–5, 2021.

[7] S. Bader et al., "The international data spaces information model–an ontology for sovereign exchange of digital content", in *International Semantic Web Conference*, Springer, 2020, pp. 176–192.

[8] J. Pampus, B.-F. Jahnke, and R. Quensel, "Evolving data space technologies: Lessons learned from an ids connector reference implementation", in *International Symposium on Leveraging Applications of Formal Methods*, Springer, 2022, pp. 366–381.

[9] M. Faraji Shoyari and M. Duchon, "Scalable and interoperable data infrastructure in energy domain", in *2. NFDI4Energy Conference 2025*, 2025. [Online]. Available: https://doi.org/10.5281/zenodo.14961100

[10] J. Theissen-Lipp et al., "Semantics in dataspaces: Origin and future directions", in *Companion Proceedings of the ACM Web Conference 2023*, 2023, pp. 1504–1507.

[11] D. M. Dave and B. Mittapally, "Data integration and interoperability in iot: Challenges, strategies and future direction", *INTERNATIONAL JOURNAL OF COMPUTER ENGINEERING & TECHNOLOGY*, vol. 15, pp. 45–60, Jan. 2024.

[12] O. Oyeniran, O. Modupe, A. Otitoola, O. Abiona, A. Adewusi, and O. Oladapo, "A comprehensive review of leveraging cloud-native technologies for scalability and resilience in software development", *International Journal of Science and Research Archive*, vol. 11, pp. 330–337, Mar. 2024. DOI: 10.30574/ijsra.2024.11.2.0432

[13] C. Vigurs, C. Maidment, M. Fell, and D. Shipworth, "Customer privacy concerns as a barrier to sharing data about energy use in smart local energy systems: A rapid realist review", *Energies*, vol. 14, p. 1285, Feb. 2021. DOI: 10.3390/en14051285

[14] J. Matar, M. Koubaa, and M. Duchon, "Exploring the potential of freely available satellite data for energy applications", in *2. NFDI4Energy Conference 2025*, 2025. [Online]. Available: https://doi.org/10.5281/zenodo.14843602