Multi-Partner Project: Green.Dat.AI: A Data Spaces Architecture for enhancing Green AI Services

Ioannis Chrysakis[1,2,3], Evangelos Agorogiannis[1], Nikoleta Tsampanaki[1], Michalis Vourtzoumis[1], Eva Chondrodima[4], Yannis Theodoridis[4], Domen Mongus[5], Ben Capper[6], Martin Wagner[7], Aris Sotiropoulos[8], Fábio André Coelho[9,10], Cláudia Vanessa Brito[9], Panos Protopapas[11], Despina Brasinika[11], Ioanna Fergadiotou[11], Christos Doulkeridis[4]
[1]*Netcompany-Intrasoft, Research and Innovation Development Department*, Luxembourg,
[2]*Ghent University*, Belgium, [3]*KU Leuven University*, Belgium,
[4]*University of Piraeus*, Piraeus, Greece, [5]*University of Maribor*, Slovenia, [6]*Redhat, Ireland*,
[7]*Eviden, Spain*, [8]*AEGIS IT Research, Germany*, [9]*INESC TEC*, [10]*University of Minho, Portugal,*
[11]*Inlecom Innovation, Greece*
[1]{Ioannis.Chrysakis,Evangelos.Agorogiannis,Nikoleta.Tsampanaki,Michalis.Vourtzoumis}@netcompany.com,
[4]{evachon,ytheod,cdoulk}@unipi.gr, [5]domen.mongus@um.si,[6]bcapper@redhat.com,[7]martin.2.wagner@eviden.com,
[8]a.sotiropoulos@aegisresearch.eu, [9]{fabio.a.coelho,claudia.v.brito}@inesctec.pt,
[11]{panos.protopapas,despina.brasinika,ioanna.fergadiotou}@inlecomsystems.com

***Abstract*—The concept of data spaces has emerged as a structured, scalable solution to streamline and harmonize data sharing across established ecosystems. Simultaneously, the rise of AI services enhances the extraction of predictive insights, operational efficiency, and decision-making. Despite the potential of combining these two advancements, integration remains challenging: data spaces technology is still developing, and AI services require further refinement in areas like ML workflow orchestration and energy-efficient ML algorithms.**

In this paper, we introduce an integrated architectural framework, developed under the Green.Dat.AI project, that unifies the strengths of data spaces and AI to enable efficient, collaborative data sharing across sectors. A practical application is illustrated through a smart farming use case, showcasing how AI services within a data space can advance sustainable agricultural innovation. Integrating data spaces with AI services thus maximizes the value of decentralized data while enhancing efficiency through a powerful combination of data and AI capabilities.

***Index Terms*—data spaces, dataspaces, data ecosystems, data sharing, artificial intelligence, green transition**

I. Introduction

Artificial intelligence (AI) has seamlessly integrated into our daily lives, offering a diverse range of systems and services aimed at enhancing both productivity and lifestyle [1], [2]. AI technologies are widely employed, to streamline decision-making processes and deliver customised solutions tailored to specific client requirements [3]. The widespread adoption of AI services has resulted in a significant upsurge in data generation, presenting a dual opportunity: the potential for shared data among organisations to create value [4], [5], and simultaneously, an amplified computational burden on resources.

In response to these challenges, the introduction of Data Spaces (DS) emerges as a promising pathway for encouraging collaborative data sharing, aligning it with the advantages of AI services. Data spaces utilise a collaborative and efficient methodology in managing dispersed data, offering avenues to alleviate efficiency challenges linked with AI services.

The integration of data spaces and advanced AI services presents substantial opportunities for enhancing the functionality and accessibility of shared data infrastructures across established data ecosystems [6]. Technical foundations for data spaces have been established by initiatives such as IDS [7], GAIA-X [8], and Fiware [9], which provide frameworks and guidelines essential for secure, interoperable, and scalable data exchange. However, adopting AI in data spaces is still an evolving domain with many technical challenges [10], [11].

To effectively incorporate AI services into data spaces, several efficiency-focused considerations must be addressed: the orchestration of workflows across diverse Machine Learning (ML) models, the optimization of algorithms for energy efficiency to support sustainable practices, and the integration of current applications to ensure cohesive and functional data ecosystems. The combination of data spaces and AI requires a multi-layered approach that supports seamless collaboration while aligning with green transition goals and addressing the needs of end users, such as data consumers and domain experts.

As a preliminary step in integration, a Reference Architecture becomes indispensable to address the aforementioned challenges, serving as a blueprint for establishing sustainable data spaces. Towards this goal, the Green.Dat.AI Reference Architecture (RA), launched within the Horizon Europe project *Green.Dat.AI* [12], is now being adopted by six industry partners across energy, agriculture, mobility, and banking sectors [13]. The RA features three components: Data Providers, who integrate applications; Service Providers, who supply AI tools (AI & Service Participants); and Data Space Common Services, which handle data discovery and access. This architecture models a comprehensive data-sharing solution and demonstrates the practical combination of AI and data spaces as a driver for sustainability.

978-3-9826741-0-0/DATE25/© 2025 EDAA

II. State of the Art

Data Spaces. Data spaces were first introduced in [14] as a data integration framework designed to support co-existence and foundational functionality across diverse data sources. Over time, this concept has evolved into a robust approach for data collaboration, promoting shared goals through streamlined data sharing [15]. Data spaces now include essential services such as secure authentication, semantic data definition storage, data discovery, contractual data usage frameworks, and secure data exchange facilitated through data connectors. Collectively, these components form a trusted platform for data sharing across organizational boundaries.

Aligned with the European Strategy for Data [16], the EU supports the development of European Data Spaces to drive economic growth and optimize data utilization for societal benefit [17]. However, data spaces represent dynamic ecosystems that face challenges such as integrating existing systems and adopting advanced technologies, such as AI, to enhance data enrichment, quality, and collaborative decision-making [10]. To address these complexities, standardized reference architectures are crucial for establishing common technological foundations, methodologies, and standards that further data collaboration, supporting the sustainability and viability of data ecosystems over the long term.

Reference Architectures for Data Spaces. Several European initiatives are advancing reference architectures for data spaces, with notable examples including IDS [18] and GAIA-X [19]. Additionally, the Data Spaces Support Center [20] is defining a blueprint containing foundational building blocks, such as data connectors, to streamline implementation. Domain-specific reference architectures are also emerging; for instance, MobiSpaces [21], [22] addresses data governance needs in mobility, while initiatives like [23] focus on privacy and security within the health sector.

The development of data spaces remains an evolving field, requiring architectures that adapt to unique sectoral demands. This paper proposes a reference architecture designed to incorporate AI services, facilitating enhanced decision-making and meeting the diverse requirements of data space participants.

III. The Green.Dat.AI Reference Architecture (RA)

The Green.Dat.AI. Reference Architecture consists of three core elements: (i) Data Space Common Services, (ii) Data Providers, and (iii) AI and Services Participants (depicted in Fig. 1). The *Data Space Common Services* provide the core infrastructure services for deploying a data space. On the other hand, *Data Providers* are data owners who wish to provide controlled access to their datasets to third parties in a secure and trusted way, possibly for a fee. *AI and Services Participants* offer technical solutions for AI services, predictive analytics, visualisations and for calculating the energy footprint of ML algorithms. In the following, we elaborate on the functionality of each element and its role in the Reference Architecture.

A. Data Space Common Services

The Data Space Common Services element provides all essential services for creating a functional data space, which

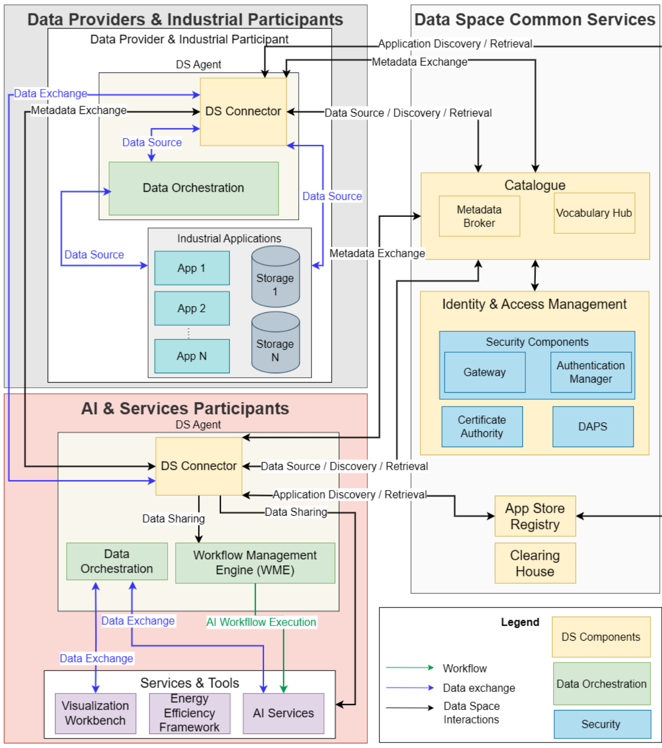

Fig. 1. The Green.Dat.AI Reference Architecture (RA)

external application users or internal users (e.g., industrial users) can employ in conjunction with existing applications and databases. These services include *Catalogue* components for data discovery, *Identity and Access Management* components for ensuring data governance and security, and an *App Store Registry* for registering apps or other artifacts acting as a marketplace for both developers and users of the data space. Also, a *Clearing House* Component acts as a certification component to ensure alignment with state-of-the-art data spaces initiatives (e.g. IDS, GAIA-X).

The *Catalogue* component enables efficient data discovery within the data space through search, filter, and browse functions, making datasets accessible for diverse applications. It also ensures data governance and access control, aligning data availability with participant needs. A critical subcomponent of the Catalogue, the *Metadata Broker*, facilitates the registration, publication, and maintenance of metadata through Self-Descriptions. Self-Descriptions summarise information about data space participants and their *DS connectors*, supporting various data exchange operations within the data space. These Connectors serve as a communication gateway between the Data Space Common Services and the other Reference Architecture elements, offered by Industrial Participants and AI and Services Participants.

The *Vocabulary Hub* component, part of the *Catalogue*, provides a vocabulary storage service that promotes collaborative governance and ensures interoperability within the data space by fostering mutual understanding. It supports semantic interoperability with tools to publish, manage, and enhance vocabulary usability, advancing FAIR principles [24] through semantics [25].

To facilitate *Identity & Access Management*, *Security Components*, are provided: a *Gateway* for managing API traffic, load balancing, and rate limiting, and an *Authentication Manager*, supporting various authentication mechanisms, from password-based to token-based systems. Together, these ensure that only verified users access the data space, enhancing security and performance. In addition, a *Certificate Authority (CA)* component issues digital identities to data space participants, establishing transparency and trust among them. Complementing this, a *Dynamic Attribute Provisioning Service (DAPS)* enriches participant identities and their DS connectors.

The *App Store Registry* functions as a digital platform for distributing software applications and ML algorithms utilized by AI services within the AI & Services Participants component. Specifying the types of ML models or applications could provide a clearer idea of the use cases. It catalogues each app's details—such as ownership, functionality, and usage requirements—enabling data space participants to browse, evaluate, and download applications and ML models, thereby facilitating streamlined data sharing.

B. Data Providers & Industrial Participants

Industrial applications, managed by various participants with specific storage needs, function as data providers within the data space ecosystem, making their data accessible across the Green.Dat.AI. For each data provider, the Reference Architecture recommends adopting a DS Agent module, which integrates a Data Spaces Connector and a Data Orchestration module to enable seamless interaction with Data Space Common Services.

The *Data Orchestration* module facilitates communication among components and supports real-time data exchange, utilizing platforms for high-performance data streaming. Additionally, a Storage Service should handle multiple data types, including historical data, which is essential for control, monitoring, and predictive insights via AI-driven services provided by the AI & Services Participants. This architecture ensures robust management of both historical and real-time data, enhancing data availability and functionality across the ecosystem.

C. AI & Services Participants

With the data space enabled and equipped with data and relevant tools, the final element of the Green.Dat.AI Reference Architecture integrates AI capabilities. This includes a suite of *AI services* designed to support specified AI workflows, a *Workflow Management Engine* to define, schedule, and execute these workflows, and a *Visualization Workbench* to display results and provide visual data analytics for data space participants. Additionally, an *Energy Efficiency Framework* is incorporated to assess and optimize the performance of ML algorithms used within AI services, promoting further enhancements in efficiency.

AI Services. The Reference Architecture intends to support AI services. To integrate such AI services with the architecture, a set of interfaces is defined. First, an AI service needs to be able to consume data from a DS Connector, thus enabling the integration between data spaces and AI algorithms. Second, AI services can be part of complex workflows defined, managed and executed by the Workflow Management Engine.

AI Services need to interface with the Data Orchestration module, for supporting data exchange with other components of the Reference Architecture. Notice that this data exchange path is orthogonal to the data sharing mechanism provided by the data space, thus providing an alternative way of data exchange that increases the applicability and flexibility of the reference architecture in real-life use-cases. The Reference Architecture emphasizes energy efficiency, particularly for AI services that are computationally intensive during training. To address this, it incorporates an Energy Efficiency Framework, which monitors ML models' energy consumption and provides insights for achieving models that balance high accuracy with low energy use.

Workflow Management Engine. To meet the AI demands of data space participants—such as leveraging advanced analytics—domain experts design complex AI workflows. Automating these workflows requires a Workflow Management Engine (WME), which enables the definition, management, storage, and execution of these tasks. This automation minimizes human error, boosts efficiency, and ensures workflow consistency and reliability, particularly when handling multiple concurrent tasks. The *Green.Dat.AI WME* provides this structured automation to streamline processes and maintain robust, dependable operations across AI-driven purposes.

Visualisation Workbench A visualization component is essential in the proposed architecture to effectively display AI services' outputs. The *Green.Dat.AI Visualisation Workbench (VW)* provides interactive visual data analysis within the established data space. This component retrieves data from the *Data Orchestrator* module. The *Visualisation Workbench* dynamically crafts visualisations catering to the unique requirements of each industrial pilot. These visualisations encompass interactive timeline displays for time-dependent data, graph-based visualisations, chord diagrams, geospatial representations, and more.

Energy Efficiency Framework. The *Energy Efficiency Framework (EEF)* enables standardized, measurable evaluation of AI algorithms based on key metrics: energy consumption, model accuracy, data volume, and execution time. This joint assessment framework ensures comparability without modifying algorithm code (black-box testing). Each algorithm must implement a standard interface to manage execution and data inputs, running in pre-set configurations for consistency. Metrics collected provide developers with insights into code efficiency and performance impacts. Baseline benchmarks for algorithms, such as Federated Learning and Time-series forecasting, are set on uniform hardware, while participants can also run benchmarks on their infrastructure, preserving data privacy. This framework is essential for gathering comprehensive metrics beyond energy profiling, supporting AI model evaluation and continuous improvement.

IV. GREEN.DAT.AI IN ACTION

In this section, we present how the proposed Reference Architecture is applied on a real-life pilot for smart farming optimisation. First, we present the pilot scenario (Sect. IV-A), and then we present validation results (Sect. IV-B).

A. The Pilot Scenario: Smart Farming Optimisation

The pilot scenario concerns *smart farming* and is applied to the agricultural sector. While existing industrial applications rely on data sources from remote sensing technologies such as satellites, drones, and in-situ sensors, the idea is to exploit the Data Space Common Services of the Green.Dat.AI Reference Architecture to facilitate data exchange and use of AI-services to reduce use of pesticides and fertilisers. This is achieved through: (i) real-time generation of fertilisation maps tailored to soil and crop requirements for optimising fertilisation; (ii) early detection of plant pests and diseases through (near) real-time object assessment, achieved by fusing satellite imagery and drone inspections; (iii) prognostication of soil health status and determining optimal harvesting times. The implementation of this pilot includes the development of digital twins that addresses the needs of farming advisors to capture the behaviour of agricultural lands and optimise their management, reducing the volume of necessary training data and shortening model training times. Therefore, this pilot: a) enhances the learning of various plant pests and disease detection models by drawing insights from different users with diverse crops and pest/disease challenges, b) generates a comprehensive soil health status map by integrating data from various sources and, c) optimises fertilisation strategies based on different crops and soil conditions.

B. Pilot Validation

We detail the practical implementation of data spaces and AI services in the smart farming optimization pilot using the proposed Reference Architecture. Specific design choices and implementation details for each element of the Green.Dat.AI Architecture are presented, as summarized in Fig. 2.

Data Space Implementation. To establish the data space following the proposed Reference Architecture, we initially deployed the Catalogue component using Pivaue [26], an open-source, scalable solution designed to manage the entire metadata's lifecycle. This metadata conforms to established standards such as DCAT-AP [27] and RDF, ensuring semantic interoperability and seamless integration with external data platforms. In the smart farming pilot, this catalogue enhances data discoverability for resources like satellite images, IoT sensor data, and historical crop yields, providing farm managers and stakeholders with a structured view to support informed decisions on actions such as fertilisation and pest control.

Simultaneously, DS Connectors are deployed at both the Data Providers & Industrial Participants layer (pilot side) and the AI & Services Participants layer (AI service provider side). This setup used a customised EDC connector [28], enhanced with Sovity's open-source extensions [29], providing flexibility through features like "Connector-as-a-Service" for efficient data space management.

Keycloak [1] and KrakenD[2] were chosen for Identity and Access Management due to their advanced security, encryption, and access control capabilities. These were integrated with Sovity's Dynamic Attribute Provisioning Service (DAPS) [30], ensuring role-based access for pilot participants like farm managers and data scientists, enabling real-time access validation and compliance.

The Clearing House functionality was adapted using the GAIA-X Clearing House framework [31] to ensure GAIA-X compliance and compatibility with IDS standards. Additionally, an App Store Registry was introduced to offer certified applications tailored to the smart farming pilot, including data analysis, mapping, and monitoring tools, which align with data space standards and streamline pilot participant access to compliant and interoperable tools.

Data Management and Industrial Applications Integration. We validate the proposed Reference Architecture using the environmental intelligence data fusion stack from [32]. Three existing industrial applications contribute to data collection and support decision-making for end users, such as farmers and domain experts, as follows:

- *My Farm (Farm Manager Module)*: A smart farming application, widely used by Slovenian advisory services, integrates data from multiple sources (e.g., satellite, UAV, in-situ, third-party) to support data-driven decision-making. It employs a digital twin to simulate ecosystem behaviors—such as soil, vegetation, and field conditions—enabling optimized farming practices. This application delivers accurate, high-quality data, allowing consultants to provide informed recommendations to farmers for efficient production planning.
- *GFS Farm Maps*: It is dedicated to mapping data production, ensuring that the generated data products meet stringent quality criteria, including accuracy, spatio-temporal resolution, and well-structured encapsulation of information. It plays a crucial role in supporting decision-making processes for farming consultants (Fig. 3).
- *GFS Platform*: It enables the processing and delivery of intermediate and final data products identified for the pilot's needs and collaborates with the Data Orchestration module to store or exchange these products.

The Data Orchestration module was implemented via the Streamhandler platform [33] which facilitates communication, data transferring, data exchange, and results sharing among these existing applications and other Green.Dat.AI components. It is based on Apache Kafka [3], providing the necessary infrastructure to handle high-throughput and low-latency data streams, enabling the system to process vast amounts of data efficiently. This ensures that analytic processes and ML models receive timely and accurate data, which is critical for maintaining AI services' overall performance and effectiveness. Moreover, additional data storage mechanisms have been tested

[1] https://www.keycloak.org/
[2] https://www.krakend.io/open-source/
[3] https://kafka.apache.org/

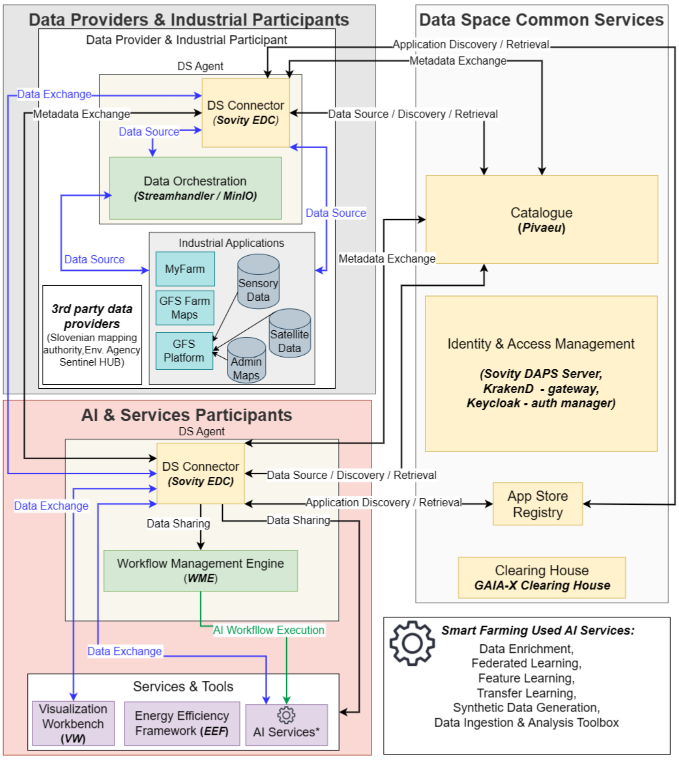

Fig. 2. Pilot implementation environment adapted on Green.Dat.AI RA

in this module, such as enabling MinIO [4] for applications that rely on S3 for storage or for data used for historical retrieval.

During the execution of the pilot, open data are gathered from three distinct sources (the Slovenian Mapping Authority, Slovenian Environmental Agency, and Sentinel Hub); These data are managed by the Streamhandler platform. The same data is essential to power and operate the AI services offered in the context of Green.Dat.AI project.

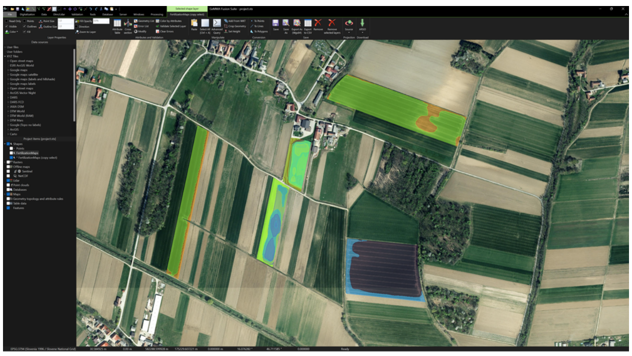

Fig. 3. GFS Farm Maps: view of the area

AI Services and New Tools Integration. The AI Services consume data from the established DS connectors to address the objectives of the pilots as follows:

1) Harvest Optimization based on crop development prediction: This involves analyzing past and predicted weather data along with Sentinel 2 images for a designated cadastral unit (see Fig. 4).

- The task utilizes the *Data Enrichment Service* to associate weather data with IoT data. This service relies on RDF-Gen [34], a flexible and scalable tool for data transformation to RDF. The input datasets are transformed into a semantic representation (using RDF), which allows data interlinking [35], [36].
- The *Data Ingestion Service and Analysis Toolbox* is used to monitor weather data, produce basic weather forecasts and insights on the soil quality. This service efficiently gathers data from diverse sources like APIs, FTPs, and various Data Storage systems such as SQL and MongoDB. It uses an ontology-based GraphQL server as a central hub for metadata management, with Neo4j [5] for storage.
- Two additional AI services are used: the *Feature Learning Service* and the *Transfer Learning Service*. Particularly, these services combine a series of prediction models for forecasting crop development, acting on raster data and including predictions of different vegetation, soil, and moisture indices.

2) Pest and disease detection: Probability maps are generated based on anomaly detection, considering anticipated and actual crop development.

- The *Feature Learning Service* extends the GeMMA Fusion Suite[6] for digital twins construction, including suitable Feature Learning algorithms and map generation tools. The service also relates to eXplainable AI [37], due to the evaluation of features for the learning task at hand. As such, the explainable feature learning approach described herein extends on the method proposed by Vlahek and Mongus [38] by considering additional feature construction operators for dealing with the categorical variables. Essentially, an efficient iterative learning approach for explainable feature learning is adopted, aiming to exploit non-linear interdependencies between features and thus improves the classification efficiency of any classifier and at the same time helps in knowledge discovery.
- For improved anomaly detection of plant pests and diseases from multiple sensory and image data, the *Federated Learning Service* is used. It provides privacy-preserving FL methods for time series forecasting at the fog/edge. Besides privacy-preservation, learning is performed in a distributed way on fog/edge devices over subsets of the complete dataset. This results in significant energy savings during model training, which is processing-intensive.

3) Fertilization and spraying maps generation: Here, the objective is to estimate the amount of specific fertilizer or pesticide for a particular geospatial region.

- The *Feature Learning Service* is used again for generating features that that the prediction algorithm will use to achieve improved accuracy.
- The *Synthetic Data Generation Service* offers tools for generating synthetic data [39] for model training, which

[4]https://min.io.

[5]https://neo4j.com/

[6]https://apps.gemma.feri.um.si/gemmafusion/

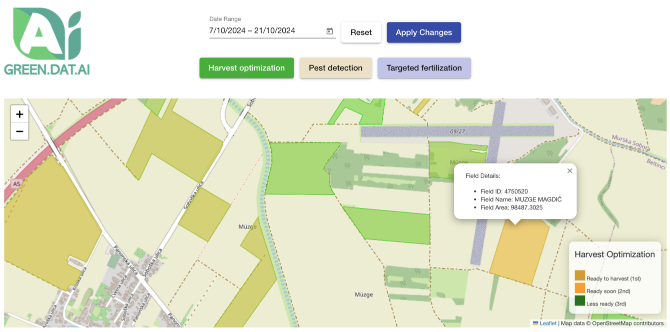

Fig. 4. Visualisation Workbench: Harvest Optimisation

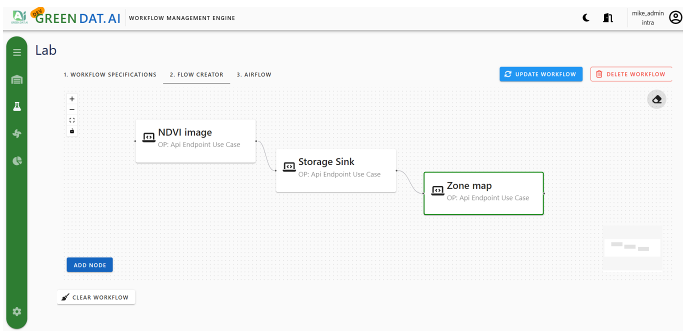

Fig. 5. Workflow Management Engine: A Pilot ML Workflow

is particularly important in the case of insufficient real-world data. As such, it fulfills the need for extensive testing/validation data and complements the transfer learning service, for particularly hard setups where the transferred learned model does not perform sufficiently well.

- Finally, the *Transfer Learning Service* is used to estimate the amount of fertiliser required during the current fertilisation phase. It offers distributed learning solutions, federated and transfer learning methods [40] for complex application scenarios. Both the federated aspect and transferability across domains are key to energy efficiency, reducing energy costs during training.

To enhance energy efficiency, we deployed these AI services close to the data sources and used the *Energy Efficiency Framework* (see Fig. 6) to evaluate AI models' performance and optimize service utilization. In practice, pilot requirements were initially translated into user stories focusing on data management aspects, which were then transformed into ML workflows to provide AI-driven decision support for agricultural optimization. We defined, managed, and executed these workflows in the Workflow Management Engine (WME; see Fig. 5), while the resulting outputs were visually represented through the Visualization Workbench (VW; see Fig. 4).

V. Conclusion and Future Work

In this paper, we introduced the Green.Dat.AI Reference Architecture, offering a novel approach to integrating data spaces with AI services to create added value through seamless symbiosis and integration for diverse participants within an

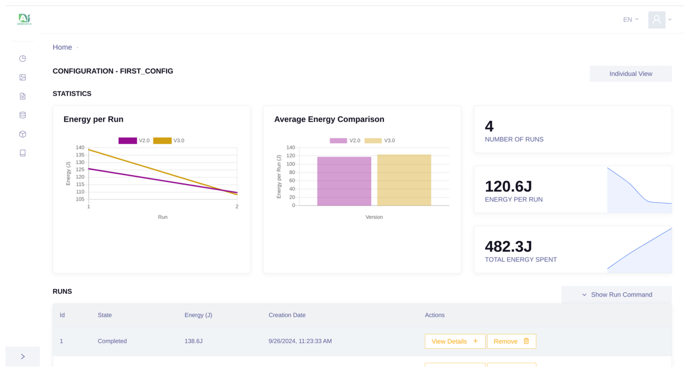

Fig. 6. Energy Efficiency Framework: Calculations on AI models

established data space. We validated this approach by adapting the Reference Architecture in a smart farming pilot scenario focused on optimizing harvests based on crop development predictions, pest and disease detection, and targeted fertilization. This pilot utilizes innovative AI services, existing applications, and tools like a workflow management engine for ML workflows, an energy efficiency framework for evaluating AI algorithms, and a visualization workbench that presents AI outputs and aids expert decision-making.

We plan to expand and validate our approach in other sectors, including energy, mobility, and banking. This will involve reusing the same technology for data spaces and data management aspects, as well as leveraging some of the proposed AI services (e.g., for data ingestion and aggregation or feature and transfer learning). Additionally, we aim to connect with other established data spaces to maximize available data and service offerings, fostering larger data ecosystems that benefit society and businesses.

Acknowledgement

The research leading to the results presented in this paper has received funding from the European Union's funded project Green.Dat.AI under Grant 101070416.

References

[1] S. J. Russell and P. Norvig, *Artificial intelligence a modern approach*. Prentice Hall, 2010.
[2] T. H. Davenport, R. Ronanki *et al.*, "Artificial intelligence for the real world," *Harvard business review*, vol. 96, no. 1, pp. 108–116, 2018.
[3] K. L. Siau and Y. Yang, "Impact of artificial intelligence, robotics, and machine learning on sales and marketing," in *Proc. of MWAIS*, 2017.
[4] V. Mayer-Schönberger and K. Cukier, *Big Data: A Revolution That Will Transform How We Live, Work, and Think*. Houghton Mifflin Harcourt, 2013.
[5] A. Zwitter and J. Hazenberg, "Decentralized network data: The new paradigm of data sharing," *Frontiers in Big Data*, vol. 3, 2020.
[6] E. Curry, "The elements of big data value: Foundations of the research and innovation ecosystem," 2021.
[7] "International data spaces association." [Online]. Available: https://internationaldataspaces.org
[8] "Gaia-x." [Online]. Available: https://gaia-x.eu
[9] "Fiware." [Online]. Available: https://www.fiware.org
[10] E. Curry, S. Scerri, and T. Tuikka, "Data spaces: Design, deployment and future directions," 2022.
[11] S. Kumar, S. Datta, V. Singh, S. K. Singh, and R. Sharma, "Opportunities and challenges in data-centric ai," *IEEE Access*, vol. 12, pp. 33 173–33 189, 2024.

[12] "Green.dat.ai project," 2023. [Online]. Available: https://greendatai.eu/
[13] "Green.dat.ai: Enabling energy-efficient ai services." [Online]. Available: https://www.innovationnewsnetwork.com/green-dat-ai-enabling-energy-efficient-ai-services/
[14] M. Franklin, A. Halevy, and D. Maier, "From databases to dataspaces: a new abstraction for information management," *ACM Sigmod Record*, vol. 34, no. 4, pp. 27–33, 2005.
[15] B. Otto, M. ten Hompel, and S. Wrobel, Eds., *Designing Data Spaces: The Ecosystem Approach to Competitive Advantage*. Springer, 2022.
[16] "A European strategy for data," 2020. [Online]. Available: https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52020DC0066&from=EN
[17] "Eu staff working document on data spaces," 2022. [Online]. Available: https://digital-strategy.ec.europa.eu/en/library/staff-working-document-data-spaces
[18] "International data spaces reference architecture." [Online]. Available: https://internationaldataspaces.org/offers/reference-architecture/
[19] "Gaia-x framework." [Online]. Available: https://gaia-x.eu/gaia-x-framework/
[20] "Data spaces support center." [Online]. Available: https://dssc.eu/
[21] C. Doulkeridis *et al.*, "MobiSpaces: An architecture for energy-efficient data spaces for mobility data," in *Proc. of IEEE Big Data*, 2023.
[22] C. Doulkeridis, I. Chrysakis, S. Karagiorgou, P. Kranas, G. Makridis, and Y. Theodoridis, "The MobiSpaces manifesto on mobility data spaces," in *Proceedings of the 4th Eclipse Security, AI, Architecture and Modelling Conference on Data Space*, 2024, pp. 66–75.
[23] A. S. Ahmadian, S. Franke, C. Gnoguem, and J. Juerjens, "Privacy-friendly sharing of health data using a reference architecture for health data spaces," in *Proceedings of the 4th Eclipse Security, AI, Architecture and Modelling Conference on Data Space*, 2024, pp. 103–112.
[24] M. D. Wilkinson, M. Dumontier, I. J. Aalbersberg, G. Appleton, M. Axton, A. Baak, N. Blomberg, J.-W. Boiten, L. B. da Silva Santos, P. E. Bourne *et al.*, "The fair guiding principles for scientific data management and stewardship," *Scientific data*, vol. 3, no. 1, pp. 1–9, 2016.
[25] M. Hauff, L. M. Comet, P. Moosmann, C. Lange, I. Chrysakis, and J. Theissen-Lipp, "Fairness in dataspaces: The role of semantics for data management," in *The Second International Workshop on Semantics in Dataspaces, co-located with the Extended Semantic Web Conference*, 2024.
[26] "Piveau." [Online]. Available: https://doc.piveau.io/general/introduction/
[27] "Dcat-ap." [Online]. Available: https://semiceu.github.io/DCAT-AP/releases/3.0.0/
[28] "Edc connector." [Online]. Available: https://github.com/eclipse-edc/Connector
[29] "Edc extensions." [Online]. Available: https://github.com/sovity/edc-extensions
[30] "Sovity community edition edc." [Online]. Available: https://github.com/sovity/edc-ce
[31] "Gaia-x clearing house framework." [Online]. Available: https://gaia-x.eu/gxdch/
[32] D. Mongus, M. Brumen, D. Žlaus, Š. Kohek, R. Tomažič, U. Kerin, and S. Kolmanič, "A complete environmental intelligence system for LiDAR-based vegetation management in power-line corridors," *Remote Sensing*, vol. 13, no. 24, p. 5159, 2021.
[33] "The streamhandler platform." [Online]. Available: https://ocean-twin.eu/marketplace/product/streamhandler
[34] G. M. Santipantakis *et al.*, "RDF-Gen: generating RDF triples from big data sources," *Knowl. Inf. Syst.*, vol. 64, no. 11, pp. 2985–3015, 2022.
[35] G. M. Santipantakis, A. Glenis, C. Doulkeridis, A. Vlachou, and G. A. Vouros, "stld: towards a spatio-temporal link discovery framework," in *Proceedings of the International Workshop on Semantic Big Data, SBD@SIGMOD 2019, Amsterdam, The Netherlands, July 5, 2019*, 2019, pp. 4:1–4:6.
[36] F. Scharffe *et al.*, "Enabling linked data publication with the Datalift platform," in *Proc. of AAAI Workshops*, 2012.
[37] C. Molnar, *Interpretable Machine Learning: A Guide for Making Black Box Models Explainable*. Leanpub, 2019.
[38] D. Vlahek and D. Mongus, "An efficient iterative approach to explainable feature learning," *IEEE Transactions on Neural Networks and Learning Systems*, 2023.
[39] Blanco-Justicia *et al.*, "Generation of synthetic trajectory microdata from language models," in *Proc. of PSD*, 2022, pp. 172–187.
[40] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Transactions on knowledge and data engineering*, vol. 22, no. 10, pp. 1345–1359, 2009.