Proceedings of the 19th Conference on Computer Science and Intelligence Systems (FedCSIS) pp. 543–548 DOI: 10.15439/2024F5291 ISSN 2300-5963 ACSIS, Vol. 39

Agricultural Data Space: the METRIQA Platform and a Case Study in the CODECS project

Manlio Bacco
Inst. of Information Science and Technologies (ISTI)
National Research Council (CNR), Pisa, Italy
0000-0001-6733-1873

Giovanna Maria Dimitri
Dept. of Information Engineering and Mathematics (DIISM)
University of Siena, Italy
0000-0002-2728-4272

Alexander Kocian
Dept. Computer Science
University of Pisa, Italy
0000-0001-8847-0768

Paolo Barsocchi
Inst. of Information Science and Technologies (ISTI)
National Research Council (CNR), Pisa, Italy
0000-0002-6862-7593

Antonino Crivello
Inst. of Information Science and Technologies (ISTI)
National Research Council (CNR), Pisa, Italy
0000-0001-7238-2181

Gianluca Brunori
Dept. of Agriculture, Food and Environment
University of Pisa, Italy
0000-0003-2905-9738

Marco Gori
Dept. of Information Engineering and Mathematics (DIISM)
University of Siena, Italy
0000-0001-6337-5430

Stefano Chessa
Dept. Computer Science, University of Pisa, Italy
ISTI-CNR, Italy
0000-0002-1248-9478

***Abstract*—This work describes the ongoing design and development of the METRIQA platform, hosting the Italian agrifood data space. Both are key components that the Italian National Research Centre for Agricultural Technologies is putting forward in its activities. We present a high-level description of the platform, which is designed to provide web-like access to digital resources and services following an approach called Web of Agri-Food, to support the digital transformation of the sector in Italy. To show its potential, we also present a real case study demonstrating both the benefits and impacts of the proposed architecture, connecting stakeholders and authorities at different levels.**

***Index Terms*—agri-food, data space, agriculture, traceability, data sharing**

I. Introduction

THE European Green Deal is part of the European Union (EU)'s response to the Sustainable Development Goals (SDGs), and it sets specific targets for the agricultural sector, among others, emphasizing the potential for the adoption of digital technologies. Member States, like Italy, are responsible for implementing these objectives, funded by the National Recovery and Resilience Plan (PNRR) and being implemented through the National Research Centre for Agricultural Technologies (AGRITECH). The latter is organized into nine thematic areas called "Spokes," with Spoke 9 focusing on "new technologies and methodologies for traceability, quality, safety, measurements, and certifications to enhance the value and protect the typical traits in agri-food chains." In the activities of spoke 9, the METRIQA technological platform [1] emerges as a solution to support the digital transformation of Italy's agri-food sector. The primary objective of METRIQA is to implement the "Web of Agri-Food" (WoA), which will provide web-like access to digital resources and services. The WoA integrates several data sources, connecting their resources regardless of internal technology and structure. This allows various providers to make unstructured and heterogeneous data available to users, such as the actors in the agri-food sector and other stakeholders. METRIQA users will rely on novel AI-based retrieval services to obtain relevant information, benefiting both research and sectoral stakeholders. METRIQA will also assist private companies in sharing digital resources. A core component of the METRIQA platform is a *data space* [2], which will be used at the national level for traceability and certification services in the test phase of the system.

The EU data strategy [3] aims at fostering the creation and use of data spaces for data integration. Data spaces, which can be described as decentralised data ecosystems focusing on data relationships rather than data management, have the potential to play a crucial role in facilitating a cooperative environment.

This work has been funded within the EU Horizon Europe research and innovation programme under GA no. 101060179 and the project AGRITECH Spoke 9 - Codice progetto MUR: AGRITECH "National Research Centre for Agricultural Technologies" - CUP CN00000022, of the National Recovery and Resilience Plan (PNRR) financed by the European Union "Next Generation EU".

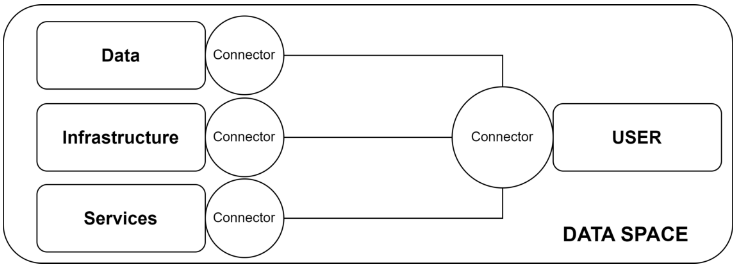

Fig. 1: Conceptual description of Data Spaces: data, infrastructure, and services shared for use according to predefined policies in a common data market.

Such ecosystems rely on mutually agreed software building blocks. The vision is that data remains at the source, and using *connectors* (specialized software components) [4] they can be accessed by interested parties (data consumers). Figure 1 shows a conceptual representation of the main functionalities offered by data spaces. Rules for data access and use are established by data owners in the form of policies. A central authority is in charge of the identity verification of connected entities and other functionalities, which will be further discussed in the following of this work.

Notable initiatives on the road to data spaces are worth mentioning, such as the IDS (International Data Spaces), the IDSA (International Data Spaces Association), and DSBA (Data Space Business Alliance) that puts together the GAIA-X European Association for Data and Cloud AISBL, BDVA, FIWARE, and IDSA. There are initiatives also outside the EU, such as the Data Society Alliance (DSA) in Japan. We remark the the DSSC (Data Spaces Support Centre) initiative, and the development of a smart platform at the EU level, namely SIMPL [5] to further support the development, testing, and use of data spaces through the development of a common middleware solution. The idea is to provide a reference implementation to facilitate the exchange of data across company boundaries and contemporarily address challenges such as interoperability, transparency, trust, and data security. Coupled with the use of W3C semantic standards, data spaces aim at formal and machine-interpretable specifications of concepts for unified data understanding. The use of e.g., MyData Control Technologies [6], which provides a technical implementation of data sovereignty, the aforementioned control on data by owners can be exercised.

In this work, we explore the current state of play and ongoing initiatives when it comes to data spaces, especially in the agricultural sector, in Section II, and provide details on the ongoing efforts for the development of the aforementioned METRIQA platform in Section III. Additionally, we explore a real case study, located in Tuscany, Italy, in Section IV and discuss how data spaces can provide benefits in such a context. We conclude the work in Section V.

II. State of the art

Digital information is a key driver of business innovation, social innovation and prosperity, thus it plays an important role in transforming our lives. One of the driving forces behind this transformation is the exponential growth of data in recent years, considering the wide spread of the Internet of Things devices and, in general, low-cost and commercial hardware widely spread and daily used. Digitalization has also arrived in the agriculture sector [7]. As an example, high-tech greenhouses have become versatile and multifunctional environments equipped with sensors that acquire data related to the microclimatic environment as well as the status of the plant. A specially designed ”computer brain” learns from the sensed data to infer the health status of the crops statistically [8], and to perform (optimum) control automatically [9]. More advanced closed-loop high-tech greenhouses have moved information processing to a dedicated cloud but the infrastructure is scattered, unstructured and vertical [10]. To improve efficiency, current high-tech solutions need to be capable of inter-operating and communicating with other systems and clouds making an infrastructure-based approach essential. As a consequence, data platforms are creating a marketplace able to connect data providers and data users.

Focusing on data space technologies, semantic World Wide Web Consortium (W3C) standards, such as the resource description framework (RDF) and Web Ontology Language (OWL) have been developed over two decades, to ensure a common understanding of the data shared in the data space. The IDS information model is based on a graph-based RDF ontology, to link the conceptual functionalities, roles, and processes with the implementations in the Connector interfaces and endpoints. MyData Control Technologies can be applied to implement the capability of data sovereignty. To share data simultaneously, there must be some form of usage control enforcement. This concept relies on a certification process for components and environments based on public key infrastructure. The emergence of federated digital ecosystems has the potential to efficiently improve data availability and also offer new strategies in dealing with large volumes of data, as required by model training for AI-based approaches. Anyway, a key hurdle on the road of such an ecosystem is the availability of trustworthy data sharing and management [11]. As argued in [12], trust should be defined as a function of time, encompassing a lengthy, intricate, and continuously adapted chain of trust between services, providers, and users to adapt to the high dynamicity of the data market.

Several studies (e.g., [13]–[16]) and EU-funded projects have studied and demonstrated new approaches and solutions for digitalising agriculture in recent years, and we cite a few of them in the following. For instance, the H2020 DESIRA project (GA 818194) has explored the role of digital technologies in agriculture, forestry, and rural areas, highlighting both the potential and the risks linked with their increasing use [17]. The HE Data4Food2030 (GA 101059473) project focuses on the data economy and the HE CODECS project (GA 101060179) is carrying out a cost-benefit analysis of technologies in the agricultural sector, assessing changes in the socio-economic process by comparing the situation before and after the introduction of digital technologies in 20 case studies all over Europe [18]. The "OPEN DEI" initiative presents a 6C (connection, cyber, computing, content/context, community, customization) approach in the form of a reference architecture framework designed to facilitate cross-domain digital transformation [19]. Such a framework builds on the extensive use of sensing units in most scenarios nowadays, and the need to have data-driven pipelines and workflow management to fully exploit the huge amount of available data for decision support, especially in the industrial sector. With specific reference to agriculture, it is worth mentioning the ongoing activities of the HE AgriDataSpace action (GA 101083401) and of the HE AgriDataValue project (GA 101086461), which consider data spaces as a key infrastructural component for secure and trusted data exchanges in the EU. At the national level, we point interested readers to valuable initiatives, such as SIEX[1] in Spain, a set of interconnected databases and administrative registers, with information about agrarian holdings, and to the GC4SHEEP[2] initiative when it comes to livestock.

III. Italian Agrifood Data Space (IADS)

The National Research Centre for Agricultural Technologies (AGRITECH) in Italy aims at developing "new technologies and methodologies for traceability, quality, safety, measurements and certifications to enhance the value and protect the typical traits in agri-food chains". The centre aims to "create an information platform integrating all work packages". Following this approach, the technological platform METRIQA (MEasurements, TRaceabilIty and Quality in Agrifood chains) is designed to offer a wide spectrum of services for research, agri-food companies, citizens and public bodies.

In particular, from the point of view of services to research, METRIQA provides cloud storage to keep the data and information produced by the research activity of the Agritech Centre and makes use of Artificial Intelligence (AI) to enrich such digital resources with metadata [20], to discover knowledge, and to construct application-specific decision support systems, sharing similarities with the web but tailored for the agrifood sector. Furthermore, it also offers personalized notebooks allowing for the advanced analysis of research data to build novel AI models and tools. At the same time, METRIQA implements the abstraction we named WoA to enable other stakeholders to access the information using search engines and conversational engines based on GPT (Generative Pretraining Transformer) technologies. As an example of the services provided to research, consider the scenario of a research centre that is operating a study for the certification of the origin of a specific product, by leveraging a data collection campaign on fields to collect samples of soil and biological material from the place of production, and on the analysis of these samples to build a hyper-spectral profile of the products of the area. The results of this research and the methodology to prove the origin are published in a technical report for other experts. METRIQA can support this research by providing a place where to store all this information (reports and datasets), and provide the tools to process such data with advanced AI tools. At the same time however, the ability of METRIQA to index the report and to embed its knowledge into a conversational chatbot allows even non-researchers to access the information about the certification of origin of that product. In turn, this information may be used by companies to provide guarantees about the origin of their products to their consumers.

From the point of view of services to agri-food companies, METRIQA offers two main services: the support to traceability of products/goods leveraging on blockchain technologies and the design and implementation of the IADS according to the specifications of IDS mentioned above. The objective is to provide a scalable data space that can grow over time with the contributions of individual project partners as well as service providers. To this purpose, METRIQA encompasses several nodes, all run by the Agritech centre, such as an authority node (that serves also as an identity provider), a broker, a transaction node, and several others interconnected through suitable connectors. Among those, a *public node* will offer services to research, citizens and public bodies as discussed above. Four classes of stakeholders have been identified, namely service and data producers, consumers, node managers, and the governance body (the Agritech Centre). Figure 2 shows a high-level view of METRIQA, highlighting that the data space is a key component of the platform. Each participant must agree with the Data Space policy at registration. The registration process is controlled by the (trusted) authority node.

The other nodes can be servers or clusters of servers located in the cloud or at the network edge, managed by other public or private entities. Each of these nodes hosts data and services specific to the needs of the respective participants. In particular, each of these nodes may implement additional services tailored to companies or specific production processes. For example, through their membership to the data space, companies along a production and distribution chain can share traceability data, enabling food traceability services from grower to consumer. Integration of data from different sources and export through various services is required for such services. Service and resource providers populate the platform with data, knowledge, and services, including re-

[1] Agrarian Holding Information System: www.fega.gob.es/en/content/siex

[2] Federated Data Cloud Platform with Artificial Intelligence Layer for the Genetic and Reproductive Improvement of the National Dairy Sheep at https://gc4sheep.com

Fig. 2: A high-level visualisation of the METRIQA platform and its nodes.

searchers, companies, professionals, and public bodies. End users consume data, information, and services offered by the platform. Node managers are entities running nodes, which can include IT companies, individual researchers, and the Agritech Centre itself.

At the current stage of the project, we have designed the reference architecture of METRIQA, specifically the three layers of storage, components, and applications. Starting with the Storage Layer, its main components are the message broker and the stored data. The storage contains the data lake of the (research) data that Spoke 9 is generating. At the Components Layer, METRIQA implements a decision support system offering web-style retrieval services (such as a natural language question-answering engine), modelling notebooks using interpreted languages (e.g. Python) for the exploratory data analysis, APIs for the data flow among data providers and data consumers, metadata enrichment, Software as a Service (SaaS) and Platform as a Service (Paas) models, and containers of sectorial decision support systems internally developed. Presently, the development is focusing on the public node, using the cloud services of the Italian provider ”Consorzio Interuniversitario dell’Italia Nord Est per il Calcolo Automatico” (CINECA). The development of the public node includes the implementation of a (limited) number of decision support systems aimed at supporting the research activities of the centre, and the deployment of the main elements that support the creation of the data space.

IV. Real case study in Tuscany, Italy

The case study described in what follows comes from the ongoing activities in one of the living labs[3] of the EU project CODECS cited above. The project puts forward a vision of sustainable digitalisation in the field of agriculture to carefully consider the existing barriers and the potential negative impacts [17] which the push of digital technologies may trigger and builds on the results of the DESIRA project, such as co-design activities to reduce risks and maximise benefits for farmers and other stakeholders [21].

A representation of the case study is provided in Fig. 3 from the viewpoint of data spaces. From the left, actors in the field (farmers, farmers’ associations, factories, and so on) are shown: they represent an ecosystem of milk producers and cheese makers in southern Tuscany. Farming requires lots of paperwork for compliance with regulations, and the provision of data to control entities is a crucial activity to have access to incentives, such as those foreseen in the EU CAP (Common Agricultural Policy). Most of the data and information are not natively digital yet, meaning that several paper documents are compiled and exchanged among actors and public bodies. An added complexity is because similar data are often requested by more than one control entity, which translates into being forced to provide the same information in different formats to different recipients through different procedures. Such a documentation burden represents a significant hurdle, especially

[3]Living Labs are networks of farmers, knowledge intermediaries, stakeholders, and policymakers to address agricultural challenges with a system-level approach to provide insights to policymakers and support sustainable farming practices.

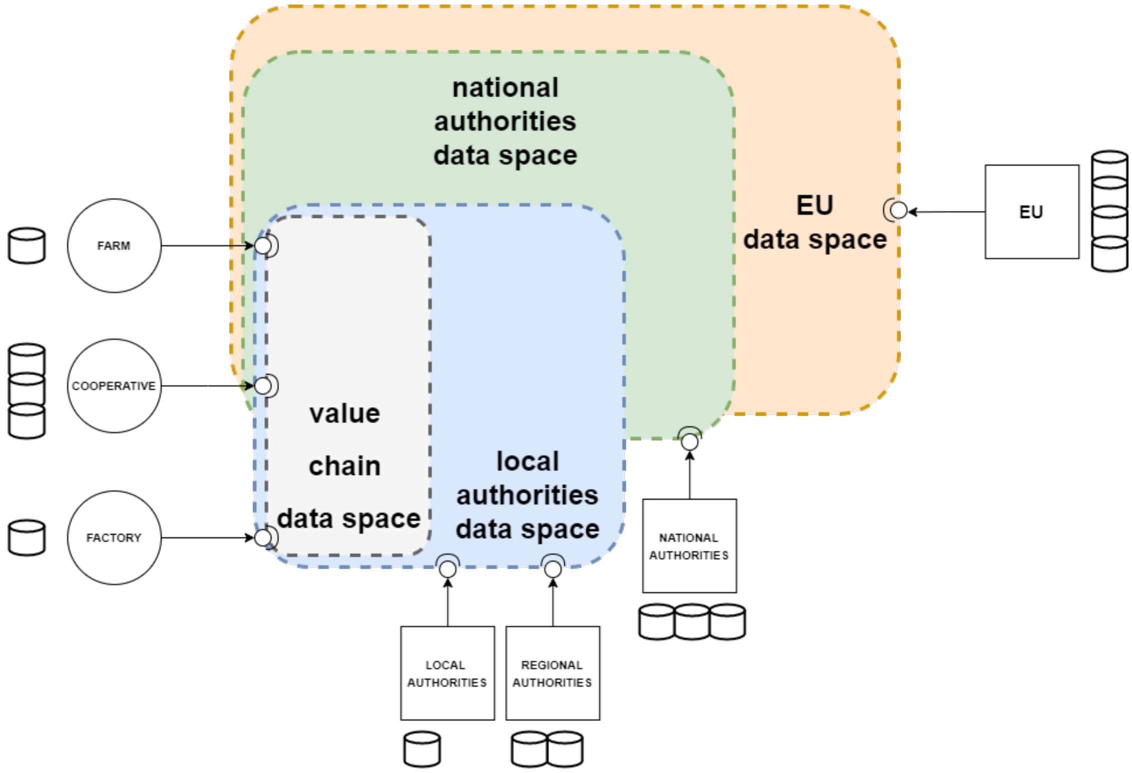

Fig. 3: A schematisation of the case study under consideration, highlighting several data spaces at different levels, nested and overlapped to meet different needs.

for small farmers. Farm Management and Information Systems (FMISs) can be of strong support in this regard, but not all portals and platforms for declarations and certifications can be programmatically queried nor automatic upload procedures can be carried out. Again, this means that most procedures must be carried out manually. Looking again at Fig. 3, what just presented translates into authorities (at the bottom) not being *connected* to the farms or to the advisory services farms rely upon. In other words, the lack of a programming interface (in the form of a standard API) may have severe impacts on how burdening the work can be. An additional *connection* must be foreseen between national and EU authorities (i.e., between the national agency for agricultural subsidies and the EU CAP system), although on a different scale.

In such a scenario, the use of one (or multiple, as in Fig. 3) data space may represent a viable solution to *connect* the aforementioned systems and actors by using standard and interoperable software components (connectors) that support data sharing and data exchange functionalities. Going back to the proposed case study, it means that FMISs at the farm level have connectors towards e.g. a value chain (or local) data space enabling data sharing among milk producers, the cheese factory, the farmers' cooperative, and other relevant actors (e.g., an external laboratory for milk testing). A local space can ensure that data are willingly shared by each actor according to predefined rules, thus feeding and being fed by the different management systems in use. A part of those data is of interest to local authorities (for instance, the Tuscany region in this case) in charge of e.g., control and certification procedures, thus an additional data space could be set up for data sharing and exchange, with different rules set by data owners to limit the access by the regional authorities to data as required by existing regulations in the area. Data spaces can be nested and overlapped, as depicted in Fig. 3 and discussed in [22], two features that should be strongly exploited in our opinion, as we hinted above. In such a way, data can be shared across multiple groups of actors according to different requirements, rules, and needs by setting up a single connector e.g. at the farm level instead of setting up multiple connectors to connect to different data spaces or having a single data space spanning from the local to the EU level. In Fig. 3, the overlapping and nesting of data spaces are represented by different colours and partial overlapping of rounded areas, each representing a different space.

V. Conclusions

The METRIQA platform is under active development within the activities of PNRR in Italy. The potential of data is being explored in both the literature and the market, and we reported a real case study in which data spaces are supposed to provide benefits. In the foreseeable future, we plan to connect the information systems at the case study level to the METRIQA platform, thus integrating data and putting *connectors* in place.

We expect two broad families of benefits. On the one hand, such convergence will allow the actors on the ground (e.g., in the case study) to interconnect with other companies nationwide, with the perspective of connecting in a (semi-)automatic fashion to additional supply chains. On the other hand, said convergence will open to deeper integration with the centre research services, as the data produced by national companies will be made available to research and industry according to rules the data owners will set, in order to experiment with innovative, AI-based data analysis methodologies.

REFERENCES

[1] S. Chessa, G. M. Dimitri, M. Gori, and A. Kocian, "WoA: an Infrastructural, Web-based Approach to Digital Agriculture," in *Proc. 14th Int. Symposium on Ambient Intelligence*. Guimarães, Portugal: Springer, 2023, p. 10.

[2] M. Atzori, A. Ciaramella, C. Diamantini, B. Martino, S. Distefano, T. Facchinetti, F. Montecchiani, A. Nocera, G. Ruffo, R. Trasarti *et al.*, "Dataspaces: Concepts, Architectures and Initiatives," in *CEUR WORKSHOP PROCEEDINGS*, vol. 3606. CEUR-WS, 2024.

[3] "European Data Strategy," 2019. [Online]. Available: https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/europe-fit-digital-age/european-data-strategy_en

[4] J. Pampus, B.-F. Jahnke, and R. Quensel, "Evolving Data Space Technologies: Lessons Learned from an IDS Connector Reference Implementation," in *Leveraging Applications of Formal Methods, Verification and Validation. Practice: 11th International Symposium, ISoLA 2022, Rhodes, Greece, October 22–30, 2022, Proceedings, Part IV*. Springer, 2022, pp. 366–381.

[5] EU, "Simpl: Cloud-to-Edge Federations Empowering EU Data Spaces," Tech. Rep., 2024. [Online]. Available: https://digital-strategy.ec.europa.eu/en/policies/simpl

[6] L. Nagel and D. Lycklama, "How to Build, Run, and Govern Data Spaces," in *Designing Data Spaces: The Ecosystem Approach to Competitive Advantage*. Springer International Publishing Cham, 2022, pp. 17–28.

[7] M. Bacco, P. Barsocchi, E. Ferro, A. Gotta, and M. Ruggeri, "The Digitisation of Agriculture: a Survey of Research Activities on Smart Farming," *Array*, vol. 3, p. 100009, 2019.

[8] A. Kocian and L. Incrocci, "Learning from Data to Optimize Control in Precision Farming," *Stats*, vol. 3, pp. 239–245, 2020, editorial.

[9] G. Burchi, S. Chessa, F. Gambineri, A. Kocian, D. Massa, P. Milano, P. Milazzo, L. Rimediotti, and A. Ruggeri, "Information Technology Controlled Greenhouse: A System Architecture," in *Proc. IoT Vertical and Topical Summit for Agriculture*. Tuscany, Italy: IEEE, May 2018.

[10] A. Kocian, G. Carmassi, F. Cela, S. Chessa, P. Milazzo, and L. Incrocci, "IoT-based Dynamic Bayesian Prediction of Crop Evapotranspiration in Soilless Cultivations," *Computer and Electronics in Agriculture*, vol. 205, Feb. 2023.

[11] M. Koch, S. Kober, S. Straburzynski, B. Gaunitz, and B. Franczyk, "Federated Learning for Data Trust in Logistics," in *FedCSIS (Position Papers)*, 2023, pp. 51–58.

[12] C. A. Ardagna, N. Bena, N. Bennani, C. Ghedira-Guegan, N. Grecchi, and G. Vargas-Solar, "Revisiting Trust Management in the Data Economy: A Roadmap," *IEEE Internet Computing*, no. 01, pp. 1–8, 2024.

[13] J. Doerr, R. Kalmar, B. Rauch, and S. Stiene, "Data Spaces in Agriculture - Status Quo and Perspectives," in *LAND.TECHNIK 2022*. VDI Verlag, 2022, pp. 511–520.

[14] R. Kalmar, B. Rauch, J. Dörr, and P. Liggesmeyer, *Agricultural Data Space*. Springer, 2022, ch. Designing Data Spaces.

[15] R. Falcão, R. Matar, B. Rauch, F. Elberzhager, and M. Koch, "A Reference Architecture for Enabling Interoperability and Data Sovereignty in the Agricultural Data Space," *Information*, vol. 14, no. 3, p. 197, mar 2023.

[16] M. Šestak and D. Copot, "Towards Trusted Data Sharing and Exchange in Agro-Food Supply Chains: Design Principles for Agricultural Data Spaces," *Sustainability*, vol. 15, no. 18, p. 13746, 2023.

[17] A. Ferrari, M. Bacco, K. Gaber, A. Jedlitschka, S. Hess, J. Kaipainen, P. Koltsida, E. Toli, and G. Brunori, "Drivers, Barriers and Impacts of Digitalisation in Rural Areas from the Viewpoint of Experts," *Information and Software Technology*, vol. 145, p. 106816, 2022.

[18] C. Mannari, M. Bacco, A. Ferrari, L. Ortolani, M. B. Lai, C. Mignani, A. Silvi, A. Malizia, and G. Brunori, "A Methodology for Process Modelling in Living Labs to Foster Agricultural Digitalisation," in *2023 IEEE International Workshop on Metrology for Agriculture and Forestry (MetroAgriFor)*. IEEE, 2023, pp. 19–24.

[19] A. Kung, S. Gusmeroli, G. Monteleone, A. Dognini, L. Nicolas, and C. Polcaro, "Reference Architectures and Interoperability in Digital Platforms," H2020 Open Dei, techreport, Apr. 2022.

[20] M. Bauer and C. Augenstein, "Can Unlabelled Data Improve AI Applications? A Comparative Study on Self-Supervised Learning in Computer Vision," in *2023 18th Conference on Computer Science and Intelligence Systems (FedCSIS)*. IEEE, 2023, pp. 93–101.

[21] M. Bacco, A. Ferrari, and G. Brunori, "Co-design of Technological Solutions for Agriculture and Rural areas: Methodology and Cases for Responsible Innovation," in *IEEE 8th World Forum on Internet of Things - Vertical Track: Agriculture, Yokohama, Japan*, Nov 2022.

[22] L. Nagel, D. Lycklama, and U. Ahle, "Design Principles for Data Spaces: Position Paper," *International Data Spaces Association*, 2021.