WoA: An Infrastructural, Web-Based Approach to Digital Agriculture

Conference Paper · July 2023

DOI: 10.1007/978-3-031-43461-7_12

Stefano Chessa
University of Pisa

Giovanna Dimitri
University of Cambridge

Alexander Kocian
University of Pisa

WoA: an infrastructural, web-based approach to digital agriculture

Stefano Chessa[1][0000-0002-1248-9478], Giovanna Maria Dimitri[2][0000-0002-2728-4272], Marco Gori[2][0000-0001-6337-5430], and Alexander Kocian[1][0000-0001-8847-0768]

Department of Computer Science, University of Pisa, Italy
{stefano.chessa, alexander.kocian}@unipi.it
Department of Information Engineering and Mathematics (DIISM), University of Siena, Siena, Italy
{giovanna.dimitri,marco.gori}@unisi.it

Abstract. The Agritech project is a large Italian project funded with the National Recovery and Resilience Plan to contribute, among other objectives, to the digitization of Agriculture in Italy and to the implementation of an information technology platform and its web-based abstraction (called Web of Agriculture - WoA) that will support research, production and commercial activities in the agrifood sector in Italy, by leveraging on innovative artificial intelligence tools. We present here the design of the platform and of the WoA and its respective case studies.

Keywords: Digital Agriculture · Data space · Web of Agriculture

1 Introduction

The European Green Deal [2] is the response of the EU to the call for a sustainable development of the United Nations 2030 Agenda [11] and it sets, among others, specific goals for the agricultural sector to achieve climate neutrality by 2050. These goals address all sectors of agriculture and give a key role to the adoption of digital technologies. The objectives of the European Green Deal fall on the partner countries that are responsible for their implementation. In the case of Italy, this is included in the National Recovery and Resilience Plan (PNRR) and, in the specific case of the agrifood sector, implemented by the National Research Centre for Agricultural Technologies (AGRITECH) [6].

The AGRITECH center itself is structured into 9 Spokes, each covering a thematic area, and among these Spoke 9 addresses "new technologies and methodologies for traceability, quality, safety, measurements and certifications to enhance the value and protect the typical traits in agrifood chains". It is within this Spoke that rises the idea of building a technological platform, called METRIQA, to support the digitization of the agrifood sector in Italy.

The main purpose of METRIQA is tow-fold. On one hand it supports the entire chain of data and information production, storage and analysis typical of an intelligent environment. The data, mostly in form of experimental datasets (together with associated reports) produced by means IoT sensors and created by the on-site experimental activities of the researchers in the Spoke, are stored and processed by using AI tools offered by the platform to support the creation of innovative decision support systems. On the other hand, METRIQA implements the abstraction of the "Web of Agrifood" (WoA) which provides a web-like access to data (digital resources) and services to its users. According to this abstraction, WoA glues data lakes of different subjects by connecting their resources despite of their internal technology and structure, so that different providers may make available any kind of unstructured resources, either relevant for research or production and concerning the agrifood sector in Italy, and METRIQA user can rely on AI-based retrieval services to obtain relevant information about this sector. We expect that METRIQA and its abstraction of the WoA will contribute both to research and productive stakeholders of the agrifood sector. In particular, it will support researchers by the implementation of a "widespread laboratory" offering them the opportunity for collecting and sharing digital resources and state-of-the-art Artificial Intelligence (AI) tools, and a mean to make their results available to non-researchers stakeholders by leveraging on a web-style search engine that support integration and searches over unstructured data. At the same time, METRIQA will support private companies in the sharing of their digital resources by the creation of a decentralized data space supporting services like those related to traceability and certification. The integration in the same platform of resources coming from both the research and the productive worlds will also pave the ways for innovative agrifood services, for example those concerning the verification of the authenticity the genuineness of the products, as it is discussed in a use case Section 4.3.

2 Background and state of the art

In this section we will briefly give an overview of the main references for the state of the art for what concerns the components of the WoA.

Digital information drives business innovation and transforms our daily life into prosperity with an incredible speed. Data generated at exponential pace, is one driving force to achieve this goal [10].

Data platforms form a market place that connects data providers and data users. To depend less large, quasi monopolistic players, the EU strategy for data is on the way to adopt multiple data spaces until 2027 Data spaces are a distributed data integration concept which is taken up by consortia. The open, transparent and secure digital ecosystem GAIA-X and International Data Spaces (IDSs) specify reference architectures for distributed data infrastructures and data spaces, respectively. International Data Spaces (IDS) enable the sovereign and self-determined exchange of data via a standardized connection across company boundaries. They address the many challenges such as interoperability, transparency, trust and security of data. The IDS develops a standardized open-source reference architecture model [9]. Similarly, the association GAIA-X with more than 334 members across 20 countries aims to increase the adoption of cloud services and accelerate data exchange by European businesses. To foster cybersecurity, the Information Security (NIS) directive and the European cybersecurity framework are the pillars on top of the requirements on technical and organizational measures related to personal data. Finally, we want to point out that the combined use of IDS technology and blockchain enables supply chain traceability.

Focusing on data space technologies [10], 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 to be shared in the data space. The IDS information model is based on a graph-based RDF ontology, to connect the conceptual functionalities, roles, and processes with the implementations in the Connector interfaces and endpoints. MY DATA Control Technologies can be sued to implement the ability of data sovereignty. To share data at the same time, there must be some sort of usage control enforcement. This concept builds on a certification process for components and environments based on a public key infrastructure. With data federation, integration takes place on the fly, Data is only accessed and integrated in the moment when the consumer ask for. The rise of trustworthy federated digital ecosystems will not only improve data availability but also handle the large volume of training data required to train machine learning and artificial intelligence, which are crucial for the competitiveness of entire business branches [10].

Over the past few years, several EU H2020 infrastructure projects aimed at digitize agriculture. Among them is the project "DESIRA" comprising an intervention planning platform along with other elements of the infrastructure (sensors, drones, geographic information system). The objectives are to i) reduce response times to citizens requesting interventions, ii) enhance the traceability and security aspects of the products, iii) foster communication, exchange of knowledge and joint activities, and iv) to provide training opportunities using gamification techniques [1]. "OPEN DEI", in contrast, proposes a 6C reference architecture framework for cross-domain digital transformation [4]. Finally, "i4Trust" lets share vineyard data with the wine value chain. Wine consumers, cellar personnel, viticulture advisors or, marketing companies can access relevant, secure, and voluntarily shared data from the vineyard [3]. To the best knowledge of the authors, this is the only solution on data spaces for agriculture at Technical Readiness Level (TRL) 9 at the time of writing.

3 The Agritech project

Agritech, the National Center for the development of New Technologies in Agriculture, is one of the 5 National Centers dedicated to frontier research financed through the MUR by the PNRR Plan as part of Mission 4 *"Education and research" - Component 2 "From research to business"*. The project objective is to promote climate change adaptation, reduction of the environmental impact in agrifood, marginal areas development, safety and traceability of the supply chains. Santa Chiara Lab (University of Siena) was selected by the Federico II University of Naples, National Center hubs leader, among the 9 Spokes, research nodes, to deliver the project objectives, jointly with the main agrifood companies sector in Italy.

Spoke 9 goal is to conduct research activities on *"Measurement, certification and enhancement of the quality, origin, typicality, safety and sustainability of products, processes, supply chains and agrifood companies"*, and was assigned an overall budget of around 40 million euro [1].

We will now describe the Spoke 9 objectives for the WP5. When developing enabling Information and Communication Technologies (ICT) to be applied for the AGRITECH world, it is important to understand how networking policies will impact production processes, system durability and resilience. Each solution will have to deal with the operating conditions, which could be typically described as extreme. Objectives of the projects will be minimizing the number of devices in the field, obtaining the absence of inline calculation, absence of local orchestration nodes, minimization of "intelligent" components allocated directly to the sensor nodes, off-grid and low-power energy production, use of low-cost field components, sustainable life cycle assessment and certification of operation in extreme conditions. Two main application areas are present in this field of research. The first one concerns the cultivation systems in a protected environment (horticultural and nursery plants in greenhouses). The second one relates to the open field cropping systems (horticulture and cash crops). The process phases, in fact, controllable by Decision Support System (DSS)-type reactive agents are characterized by certain processes to be checked (for instance irrigation at greenhouse level, single master mix fertigation, fertigation with mother mixes and a high content of single macroelements, etc) [7]. Variability in terms of measurements and open field environment should induce some considerations on the networking model to be used. For instance, the need of sending measurement data from medium/low frequency sensors (f.i. every hour). Further tasks might entail sending ending of request for state change by high frequency actuator nodes (minutes), medium/high data packet size (kbytes), local decision making processes only in case of network absence and for limited times (hours), very low network coverage distances (metres) and low operating powers (12/24 V). Same considerations could be done for DSS-type reactive agents, in the open field. Here, processes such as precision irrigation, fertigation and monitoring might be needed. Moreover problems related to energy supply and signal coverage might become critical since the sensor nodes typically operate on an island basis and should not be connected to power energy networks.

4 The METRIQA platform design and architecture

4.1 METRIQA and the Web of Agrifood abstraction

The initial objective of METRIQA is to provide a mean to handle all the data of the AGRITECH center by easing the storage and sharing of non-structured, het-

[1] https://santachiaralab.unisi.it/agroalimentare/progetto/agritech-spoke-9

erogeneous information (hereafter *d*igital resources) produced by all the project partners, but it is expected to grow, progressively enriched by other resources data and meta-data that can either created by experts or by the platform AI-engine, and by services built over those data and made available to all stakeholders (researchers, citizens, private companies, public bodies etc.), so to become a permanent asset in the Italian agrifood sector even after the end of the PNRR financial support. The key aspects of METRIQA are:

- It is neutral with respect to stakeholders: the platform is maintained by the center as a support to the stakeholders operating in the sector but no one of them controls it.
- It is technology neutral: operators in the Agrifood sector may connect their Information technology (IT) systems to the platform irrespective of their inner technologies.
- it provides a universal repository of unstructured, heterogeneous agrifood data (such as reports, documents or datasets), as those that are expected to be produced by the experimental activity of the center and by agrifood industries in Italy that enter into the platform.

To achieve all these objectives METRIQA adopts a model that closely reminds the spirit of the web: it is completely decentralised, and it makes use of most recent advances in AI to enrich with metadata the digital resources in the repository so to enable web-style search/anwser engines and the construction of application-specific AI-based decision support systems. In other terms, METRIQA provides the mean to implement the so-called "Web of Agrifood" (WoA) abstraction.

It should be observed that the need for handling unstructured digital resources comes from the need of supporting the activities of a large number of research laboratories and companies operating in the agrifood sector, each with its own objectives, tools, procedures and technologies, for which there is no standard and perhaps a common agreement on the structure of these resources and on protocols and procedures is not even possible, at least in the short/medium term. On the other hand, the idea of the project is to invest instead on artificial intelligence technologies to provide the needed structure and connection among these resources to overcome the problem by letting the machines to deal with that, as it currently happens on the web, especially considering the last developments in conversational agents such as Chat GPT [8] or Bart [5].

4.2 METRIQA architecture

From an architectural point of view, the METRIQA platform comprises one *authority* node and several, decentralized nodes one of which is the *public* node developed and maintained by the Center itself). It also defines four roles for the participants, namely service (and data) providers, end users, WoA nodes managers, and a governance body (WoA authority). A high-level view of the platform is represented in Figure 1. The nodes may be to servers (or clusters

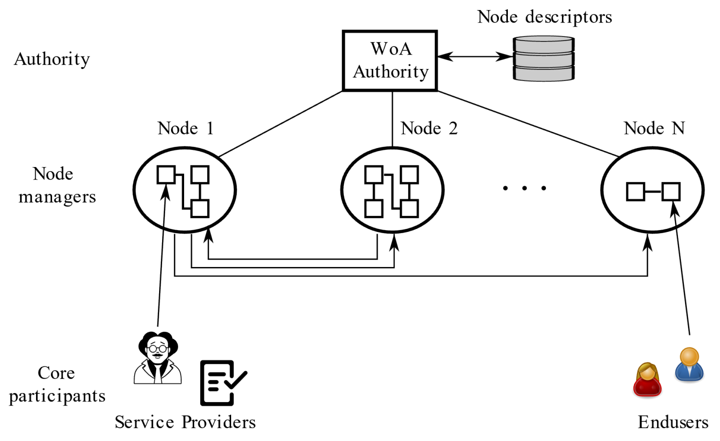

Fig. 1. A high-level view of the METRIQA platform.

of servers) on the cloud or even at the edge of the network that are run and managed by any public or private entities. Each node provides hosting for data of the participants that belong to that node and possibly a set of services of the WoA. There is not a standard for the nodes, as they may host any arbitrary set of services and they may be implemented by using different technologies. In general, the services hosted by a node respond to the needs of its participants. In the case of the public node managed by the center, it offers hosting of the research data produced by the center, with services suitable to let the researchers to upload, share and publish their data, and private workspaces to implement AI-based analysis of the data. At the same time, the public node also hosts the search engine (which is publicly available) that operates over all public data of the entire platform. Private nodes may implement additional services specific for companies or production. For example, companies along a production and distribution chain may share their traceability data, even if each of them operates with its own systems and technologies, so to implement on top of this food traceability services, "from grower to consumer". The implementation of such services will require the integration of data from different sources and exported by means of different services (in Fig. 1 such services are represented by square boxes within the nodes).

Even if there are no strict requirements on the nodes, they should be registered at the centralized authority node that plays a role similar to the Domain Name Server on the Internet, but at a higher level. The authority node keeps a database of descriptors of the nodes in the platform that, when relevant, contains also a specification of the services that are present in the the nodes (e.g. a publishing services for their digital resources). The information about how to reach the resources published by each node, which is available in the authority node, is used by the search engine to retrieve the public resources of the platform and to respond to the relative queries. Concerning the description of the services of each node, we consider three cases. The first is that of publishing digital resources service, which happens in a web-style way, and for which the descriptor is just the conventional one for the web (address of the web server and port, while the protocol is the usual HTTP). There may be however proprietary services for which the description is not available/public, and in this case the authority node just keeps an informal, textual description of those services, along with a reference to the address and port of the server. The third case is that of services that are meant as public, but that require specific ways of interaction among clients and servers. For these we adopt the same approach of the International Data Spaces initiative (IDS) [9], so that, in parallel to the web-style access to digital resources METRIQA also offers an interface compliant with that of IDS.

Concerning the roles, they reflect different classes of participants to the platform, where a participant may take more than one role at a time.

The service and resources providers are those that populate the platform with resources or services. For example, they can be researchers creating new knowledge in the agrifood sector that they publish in the WoA in form of reports and datasets, or private companies in the agrifood sector that offer public resources (e.g. food traceability data), or services to other companies, but they can also be independent professionals (e.g agronomists) or even public bodies.

The end users are any potential stakeholder of the WoA. They may be just citizens, researchers, companies operating at any level in the agrifood sector, professionals, public bodies, etc.. They are consumers of data, information and services offered by the WoA, and they interact with it either by using the search/answer engine of WoA, or by connecting directly to the interfaces of the services offered by the platform.

The node managers are entities running nodes. In the case of the public node the manager is the center itself, but managers of other nodes may be IT companies with an interest in the agrifood business that run their WoA node to host their own services or even resources or services provided by third parties (e.g. independent service providers). In some cases node managers may even be individual researchers hosting (possibly a limited number of) services and resources in their own servers.

Finally, the WoA authority is the Agritech center and runs the Authority node. It receives request of accreditation of new nodes of the platform and updates accordingly the database of node descriptors.

As discussed above, beyond their ability of publishing in web-style their resources there are no other technical requirements on the nodes. On the other hand, it is possible instead to provide a functional description of the public node managed by the Agritech center. Specifically, this node plays two roles: on one hand it implements the web-style search engine over the entire resources available in the platform. On the other hand, it provides services to research in terms of an individual workspace of researchers to host their digital resources, to share them with other researchers to create the so-called "widespread laboratory", to process them by using AI engines available in the platform and finally to publish

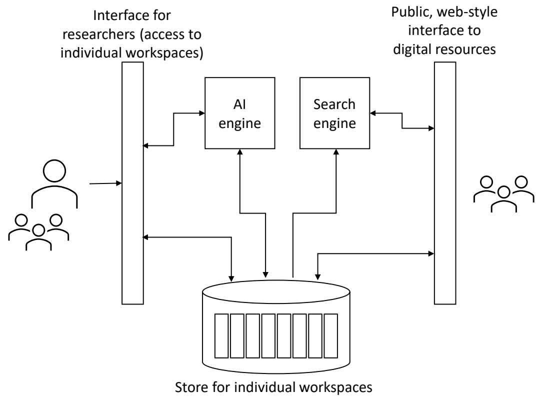

Fig. 2. A sketch of the functional components of the public node.

in web style their resources so to make them publicly available in the WoA. A sketch of the functional components of the public node is shown in Fig. 2.

4.3 State of the project and use cases

The first part of the project has been devoted to the definition of requirements and use cases related to both research and production, based on interviews with researchers and companies operating in the sector. So far, we had 31 meetings including 7 with individual researchers and 5 with companies, to filter out the requirements.. This activity is still undergoing and the first tangible result of the project is expected in September 2023 with the delivery of the public node of the platform with an initial set of services. Most of the interviews with researchers aimed at finding research-related use cases put emphasis on the fact that researchers have, in general, very well established methodologies in terms of acquisition of data, in their analysis even based on complex machinery's as it is the case for the chemical and biological analysis of samples, and for the reporting and the publication of results. At the same time, being non-IT experts, they look with interest at the exploration of AI technologies provided this results in a complementary with respect to the research activity that they normally carry out. Beyond confirming the need for a platform that deals with unstructured and heterogeneous data, these interviews highlighted the need of services to upload data in spots or continuously, even in streaming from Internet of Things sensors, and cooperatively among researchers from different side. Interviews with companies instead highlighted a strong interest in food traceability services and for the certification of chains. Under this respect many companies expressed an interest in blockchain technologies.

To illustrate one of the expected behaviour of the platform we briefly report here a research-oriented use case. Some teams of researchers scattered in different locations in Italy (e.g. the Chianti area of Tuscany or area of Squinzano in Apulia) perform sampling of the soil and biological material (fruits, leaves) to perform chemical analyzes in the laboratory to be associated with the wine supply chains of those locations and identify specific characteristics of each grape production site. These characteristics can be used to certify production and/or to its quality and for the traceability of the products. Each team performs several samples over a long sampling campaign. Each sample is associated to the location, altitude and other metadata, and it is analyzed by different machines (e.g. spectrometers) along a series of steps along which different types of intermediate data (e.g. Electron Paramagnetic Resonance spectroscopy, HyperSpectral Imaging - HSI) are produced. Each team then produces a final report at the end of the campaign that reports in natural language the details of the campaigns, the procedure used for the analysis, and the results in terms of tables, graphs etc. Although all these data are treated in a digital format (digital resources) they have many different formats, that may even differ between different teams, and some of these information are available as unstructured information in the reports. Returning to our Chianti example, the HSIs are plugged into an machine learning algorithm that assesses grape variety, water and nutritional statuses, disease-causing viruses, and the grape maturity. All these data, determining the figure-of-merit for the quality of this particular wine, is available in the accompanying report. The researchers use the platform to upload their material, each in his own workspace, and decide which resources are to be shared with the other teams, and which documents or reports to make available publicly on their website. The intermediate data, which have been processed by using established methodologies, are however available in the platform to experiment other methods of analysis by using the AI tools available in the platform, even cooperating in this with other IT teams. Some months later, a wine center in another country has to test the authenticity of a wine delivery, which is supposed to be produced with grapes of the Chianti area. They access to the WoA search engine to look for methodologies to check the authenticity of wines of that area, and they discover both the methodology (type of chemical tests) and the parameters that the tested wines should meet. When the wine center wishes a report on quality of the particular delivery, Generative Pre-Trained Transformer (GPT) technology forms a reply.

5 Conclusions

The Agritech project is in its early stage, and consequently also the activity of design of the METRIQA platform and of its WoA abstraction is still in a preliminary phase, in which we collected and analysed interviews with various research and production stakeholders. The platform itself is designed as a scalable data space that can grow over time with the contribution of independent resources and service providers, following a data space model similar to that of IDS. At the same time, however, the platform provides the WoA abstraction that frees from the need of rigid, predefined structures of the digital resources by leveraging on advanced AI-based metadata enrichment tools and search engines. Future work will be the delivery of a public call for the implementation of the first ”public node” to be made available to the project community on the short term, so to focus the work of the rest of the activity on the population of the platform with digital resources, on the development digital services for the agrifood sector based on of metadata enrichment tools and decision support systems.

Acknowledgements This work is funded within 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”.

References

1. Bacco, F.M., Ferrari, A., Brunori, G.: 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 (2022)
2. European Commission and Directorate-General for Communication: European green deal : delivering on our targets. Publications Office of the European Union (2021). `https://doi.org/10.2775/373022`
3. Goncharuk, A.: Wine value chains: Challenges and prospects. Journal of Applied Management and Investments **6**(1), 11–27 (2017)
4. Kung, A., Gusmeroli, S., Monteleone, G., Dognini, A., Nicolas, L., Polcaro, C.: Reference architectures and interoperability in digital platforms. techreport, H2020 Open Dei (Apr 2022)
5. Lewis, M., Liu, Y., Goyal, N., Ghazvininejad, M., Mohamed, A., Levy, O., Stoyanov, V., Zettlemoyer, L.: BART: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension. In: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. pp. 7871–7880. Association for Computational Linguistics, Online (Jul 2020). `https://doi.org/10.18653/v1/2020.acl-main.703`, `https://aclanthology.org/2020.acl-main.703`
6. Ministero dell'Università e della Ricerca (MIUR): Agritech - national research centre for agricultural technolgies (2022), `https://www.mur.gov.it/sites/default/files/2022-10/Schedadiprogetto-CN2.pdf`, accessed: 2023-03-16
7. Naud, O., Taylor, J., Colizzi, L., Giroudeau, R., Guillaume, S., Bourreau, E., Tisseyre, B.: Agricultural Internet of Things and Decision Support for Precision Smart Farming, chap. Support to decision-making, p. 183–224. Academic Press (2020). `https://doi.org/10.1016/B978-0-12-818373-1.00004-4`
8. OpenAI: GPT-4 is OpenAI's most advanced system, producing safer and more useful responses, `https://openai.com/product/gpt-4`, accessed: 2023-03-16
9. Otto, B., Steinbuss, S., Teuscher, A., Lohmann, S.: IDS reference architecture model (version 3.0). International Data Spaces Association (2019). `https://doi.org/10.5281/zenodo.5105529`, `https://internationaldataspaces.org/download/16630/`
10. Otto, B., ten Hompel, M., Wrobel, S. (eds.): Designing Data Spaces. Springer Cham (2022). `https://doi.org/10.1007/978-3-030-93975-5`
11. United Nations: The 2030 Agenda and the Sustainable Development Goals: An Opportunity for Latin America and the Caribbean. Santiago (2018)