Chapter 17
Agricultural Data Space

Ralf Kalmar, Bernd Rauch, Jörg Dörr, and Peter Liggesmeyer

Abstract The digital transformation strongly affects the agricultural domain. Still, there is a lot of potential for optimization in many work and business processes. In the current agricultural digital ecosystem, numerous isolated, often non-interoperable solutions exist. In this chapter, we motivate the need and added value of an "Agricultural Data Space" (ADS for short). We outline an ADS concept, which resulted mainly from the Fraunhofer lighthouse project "Cognitive Agriculture" (COGNAC) and describe the necessary prerequisites and technical solution approaches. Complemented by the possibilities of a transparent and open marketplace for data, digital products, and software services, such a data space would address many of the existing obstacles to widespread acceptance and take-up of digital technologies. Overall, an ADS as part of an extended digital ecosystem will significantly advance digitalization in agriculture. In the end, we provide application scenarios for which an agricultural data space can add value.

R. Kalmar (✉) · B. Rauch
Fraunhofer Institute for Experimental Software Engineering IESE, Kaiserslautern, Germany
e-mail: ralf.kalmar@iese.fraunhofer.de; bernd.rauch@iese.fraunhofer.de

J. Dörr
Fraunhofer Institute for Experimental Software Engineering IESE, Kaiserslautern, Germany

Chair of Digital Farming, TU Kaiserslautern, Kaiserslautern, Germany
e-mail: joerg.doerr@iese.fraunhofer.de

P. Liggesmeyer
Fraunhofer Institute for Experimental Software Engineering IESE, Kaiserslautern, Germany

Chair of Software Engineering: Dependability, TU Kaiserslautern, Kaiserslautern, Germany
e-mail: peter.liggesmeyer@iese.fraunhofer.de

B. Otto et al. (eds.), *Designing Data Spaces*,
https://doi.org/10.1007/978-3-030-93975-5_17

17.1 Digital Transformation in Agriculture

17.1.1 The Agricultural Domain

Agriculture is the oldest domain of mankind, lasting back more than 10,000 years. With the industrial revolution in the last century, both crop harvesting and livestock breeding have been highly optimized. In 1960 a German farmer fed 17 people—in 2017 this number had risen to 140 people—an enormous increase. Feeding the world's population is a great challenge, and zero hunger is one of the UN sustainable development goals. For 8.9% of the world's population, food is a scare resource and so is farmland. Most of the landmass is already cultivated, and erosion and desertification are putting farmable land at risk. Optimizing yield output is therefore one of the prime goals in agriculture. In the past years, a second goal has gained attention by society and also policy makers: sustainability. This addresses in particular environmental sustainability but also includes economic and social sustainability.

There is still a lot of potential for optimization in many processes in the agricultural value chain, especially when looking at the "big picture." Working with plants and animals faces agriculture with the complexity of the biosphere of our planet and many physical and biochemical processes. This complexity and unpredictable weather impact make it much more difficult to control and optimize than, for example, a production environment in a factory. For digitalization, this results in the challenge building many complex models, with lots of parameters and a high amount of required data.

Also, on the dimension of the horizontal value chain, agriculture combines many different stakeholders in the production chain, including farmers, contractors, manufacturers of agricultural machinery, resource providers, public authorities, traders, food processors, stores, and finally consumers.

Among these stakeholders, there is a large variety in different aspects like size and operating style, from global operating companies down to small farms operated by one family.

17.1.2 Agricultural Digital Ecosystem

Improvement potential in agriculture mainly exists in the work processes and the higher-level planning and decision processes. This requires comprehensive contextual information from the past, the present, and the predicted future. Collection, processing, and interpretation of this requires automation (i.e., through software) in order to be recorded and provided in the necessary quantity and quality.

To enable optimal operational management supported by software services, all data required for decision-making and process automation must therefore be available in digital form.

As of today, many processes in agriculture cannot be interconnected and automatically provided with the necessary information because of lacking interoperability regarding data, interfaces, and protocols. There is a fragmented landscape of islands of smaller domain ecosystems. One larger group (of yet not fully interoperable systems) is formed by machine manufacturers that comprise processes executed by agricultural machinery and the corresponding data, such as field data, machine data, or crop care documentation. Another group of ecosystems are grouped around business processes for farm management, including planning processes, resource management, sales, certification, or taxes. Bringing these and others together, so that, for example, a service provider for crop care consultancy could offer the same service on many platforms using the same implementation and being provided with the necessary data, would greatly improve the infrastructure and give a boost for digital transformation.

> An agricultural digital ecosystem should provide an infrastructure for efficiently supporting all agricultural business and work processes with services and information, provide means for flexible adaptation of needs from different stakeholders, and enable new business models.

Each digital ecosystem is formed by two or more stakeholder groups. In agriculture, these stakeholder groups comprise farmers, contractors, consultants, public authorities, research organizations, machine manufacturers, operating resource producers (seeds, fertilizer, forage, etc.), traders, logistics, services, food processors, commerce, and customers—just to name the big ones. It becomes clear that bringing these together, even in smaller steps, is a great endeavor.

17.1.3 Domain-Specific Challenges and Requirements

Apart from the great variety of processes, existing solutions, and resulting lack of data interoperability, there are other domain-specific challenges:

- Connectivity and network infrastructure: many rural areas have no high-speed Internet access or even none at all. Despite political promises, improvement is very slow.
- Offline-data collection (as a result of the above) and the need of synchronization and integration of data from different sources.
- Agriculture gets more and more attention from the general public, especially with the stronger emphasis on sustainability. It operates in public space—cultural landscape is also a public space, which impacts the environment, and therefore the interests of people.

- SME-structured farmers need to be educated in digital technologies. This is important to create acceptance and qualify farmers to actively participate in creating added value with the data.

Especially for the acceptance of systems by farmers, data sovereignty is a key factor [1]. Farming industry has defined a "Code of Conduct on Agricultural Data Sharing",[1] which addresses basic rules for using shared data. The initiative "Ag Data Transparent"[2] goes one step further and tries to certify products regarding 11 key questions of data usage. However, in most cases today, as soon as farmers provide their own data for evaluations in digital services, they currently feel to lose their sovereignty over this data. The platform providers are called upon to create solutions with which farmers can control and monitor data sovereignty easily and in a self-determined manner. Interoperability for universal data usage should not end with the farmer, however; rather, it should be enabled along the entire value chain from processing operations to the consumer. This is the only way that all stakeholders can benefit from data analyses and decision support based upon them and the only way in which comprehensive transparency can be achieved.

17.2 Agricultural Data Space (ADS)

In the agricultural domain, there are many different, partially isolated platforms and systems with redundant data, services, and software solutions. Even as integration of platforms is moving on with bilateral connectivity and the emergence of data routers, there is still no thorough connectivity for components across the digital domain. Furthermore, specific challenges of the agricultural domain, like offline capabilities or a lack of IT infrastructure at farms, have to be considered for any integration.

An important aspect to keep in mind is that even though there is a lack of interoperability and data sovereignty, the design of an ADS needs to consider the already emerged digital ecosystems. New designs cannot start from scratch but have to be adapted and integrated in the already existing environment by accepting certain boundaries and specifications. This is also where we see the main contribution of a domain-specific adaptation of IDS concepts. The implementation of IDS functionality is possible as well as promising, but it needs to be adapted to the status quo and possibly extended. Furthermore, the agricultural domain is strongly influenced by the business interests of single, big market players, which could hamper a fully interoperable data space when they strive to concentrate data within their respective realms.

---

[1] EU CODE OF CONDUCT ON AGRICULTURAL DATA SHARING BY CONTRACTUAL AGREEMENT, 05/2020. https://copa-cogeca.eu/Publications

[2] Ag Data Transparency Evaluator Inc., https://www.agdatatransparent.com/

The Fraunhofer lighthouse project Cognitive Agriculture[3] (COGNAC) researches and develops concepts for such an ADS while placing challenges like data sovereignty and interoperability in the middle of a thriving, digital domain ecosystem for agriculture. In the following sections, we describe the domain architecture as we perceive it, explain possible levels how IDS concepts can be used for an ADS, and briefly explain the benefits of an interoperable ADS.

17.2.1 Domain Architecture

Following the concepts of the International Data Space (IDS), the ADS comprises all components of digital ecosystems that generate, store, manage, or consume data and are interconnected. Just as in typical digital ecosystems, the ADS needs one or more digital platforms as key part of its infrastructure. One of the key goals is the best and most holistic possible integration of the various components of the digital agricultural domain ecosystem. To this end, we do not envision the ADS as being enabled by a single or specific digital platform as the sole player in the ecosystem. The reason for this is not only to reflect the current situation in an already emerged domain ecosystem but also the vast diversity of agricultural business and work processes and, consequently, the broad variety of specific existing (sub-)ecosystems in the domain. In order to integrate actors, services, and data, we propose a framework or reference architecture for ADS-ready components that fulfill basic requirements like interoperability and data sovereignty.

Figure 17.1 depicts an exemplary and conceptual ADS with multiple digital (sub-)ecosystems like digital platforms of machine manufactures, a routing platform, and service-specific platforms like Farm Management Information Systems. The single digital platforms would not need to be interconnected to all other existing components; data transfer or service interoperability can also be achieved by utilizing routing platforms. Connectivity would comprise syntactical and semantical interoperability for APIs and data.

For the ADS, we assume that there is no need for one big ecosystem that completely connects each and every system and player. There are many diverse facets; some players don't need to interact. Rather, the ADS framework would enable the fundamental connectivity so that all participants could engage in collaborations, but we moreover expect the ADS to develop industry-based factions like arable farming, livestock breeding, produce refinement, and so on that will keep a certain autonomy. While those factions are loosely coupled and share small portions of the value network, a tighter coupling can be expected within those factions which will develop more deeply integrated sub-ecosystems.

In this context, due to the existing (sub-)ecosystems, we strongly advocate flexibility when it comes to variants of communication channels. This means, e.g.,

[3] www.cognitive-agriculture.de

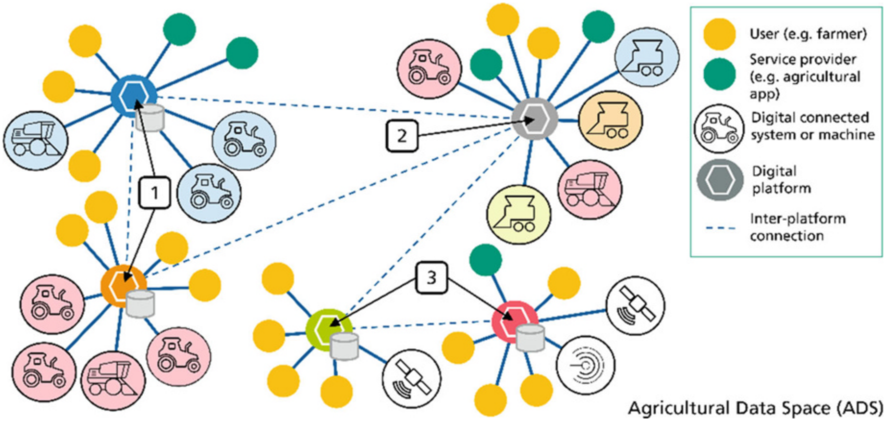

Fig. 17.1 The agricultural data space as a domain ecosystem with interconnected digital platforms and specific digital (sub-)ecosystems (machine manufacturers platforms (1), routing platform (2), and service specific platforms (3)). (illustration ©2021, Fraunhofer IESE)

that in an envisioned ADS, it should not be mandatory to utilize IDS connectors as the sole means for data exchange and protection, but there could be different mechanisms bound to the needed level of data sovereignty. If some entity is in the need of high level of data protection, IDS connectors should be used. If not, data could just be transferred via existing access controlled interfaces.

Given the diversity of the agricultural domain and the status quo with already developed ecosystems, the ADS needs to provide a high level of flexibility in both concepts and technologies to be accepted in the domain. On the other hand, the IDS provides functionalities and concepts that would fulfill yet unmet demands of the domain and thus contribute to a successful digital transformation in agriculture.

17.2.2 Possible Levels of IDS Integration

Given a multitude of perspectives in the agricultural domain, we see various entry points for an integration of IDS concepts. Currently, most activities working on connectivity and interoperability are based in a context where farmers' data is worked on and where farmers increasingly demand data sovereignty. On the other side, industry players have also an interest in data sovereignty as they seek to protect sensitive machine data like exact fuel consumption in certain situations.

If one would like to segment the implementation of an ADS in phases, we would recommend to start with the farmers' context, which can be further divided in arable and livestock farming. Farmers increasingly use software solutions and digitized machinery when handling their work processes. Farm management information systems (FMIS) collect, organize, and process data that is produced as well as consumed by machinery in the field. Figure 17.2 gives an exemplary illustration of

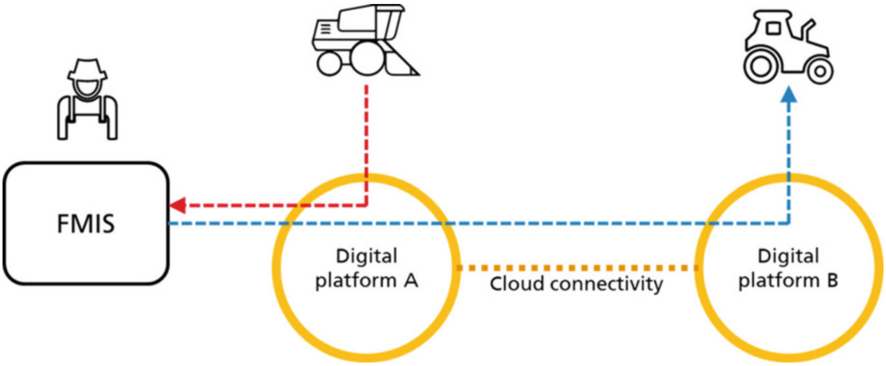

Fig. 17.2 Data flow between machinery and systems crossing platforms. (illustration ©2021, Fraunhofer IESE)

a typical data flow in arable farming, where data flows from different machinery to a software system (FMIS) and vice versa while being exchanged across different platforms. Crossing platforms is here a necessity. Machinery manufacturers often make data and API access to their machines exclusively available via own digital platforms.

In such a scenario, there is a demand for interoperability and data sovereignty between the farmers' software, digital platforms, and systems like agricultural machines. There are many activities in research and industry working on connectivity and interoperability, but data sovereignty so far is most times just implemented as access control, which is not sufficient to assure full sovereignty (Chap. 8).

As stated initially, IDS concepts could be applied on multiple levels. One general aspect to consider is that there are typically three actors involved in a scenario that focuses on farmers:

- The farmers as the data owners.
- The Platform providers (often machine manufacturers).
- Third-party system or service providers (FMIS, digital and farming services, etc.)

Farmers as well as third-party providers are participants in an ecosystem that is enabled by a digital platform, but they still are independent entities. Consequently, they strive to have data sovereignty regarding data from their respective assets against platform providers. In such a context, one could think of different levels to integrate IDS concepts. We start with a discussion of independent IDS integration possibilities, discuss possible problems in adoption, and conclude with a proposal for a hybrid approach.

- Inter-platform connectivity

  This approach would be the analogy to having IDS protection between companies, but in our scenario, it is about the data of the farmers and not the platform providers. IDS connectors can be used to connect the different platforms. A data broker could be aware of all existing services offered on the various platforms. The viewpoint of the end-user is less integrated compared to other options.

- System and IoT connectivity

  In this approach, IDS connectors can be used to connect IoT devices and the various systems. Data could be protected end-to-end between system or IoT and software like FMIS. While this would be a very thorough protection, it is unlikely to succeed on short term as it would contradict the current machine manufacturers' goals and would be very demanding to implement.
- User context connectivity

  Here, IDS protection would be implemented at a user's perimeter, meaning that farmers protect their respective data with own infrastructure while system and service providers do the same on their side, again, a highly demanding approach where farmers would need to build up or outsource extensive infrastructure, which is limiting the applicability of this solution.

As a result, a hybrid approach could be a good way to go. In this, IDS concepts could be applied on platform level, while typical systems and users within the platform context build up trust and/or legal frameworks towards the platform owners. For more sensitive data, one still can power up single systems and contexts with IDS protection at their respective perimeters as needed. In order to realize such a hybrid approach, further requirements and concepts have to be considered and can help to show the value of an interoperable data space in agriculture.

Federated Services

Another requirement to look at are federated services for the domain. As initially stated, we do envision the ADS as a digital ecosystem consisting of decentralized digital platforms, systems, and users. Those need functions to navigate the ADS like marketplaces for services and data in the ecosystem. In a day-to-day-example, a service operator in ecosystem A would like to provide a service to a farmer that has his or her data stored in ecosystem B. Via a data marketplace, the service provider can find the needed data for the service and fulfill its provision.

Digital Twins

In order to exploit the full benefit from making data across platforms and systems accessible, the domain could build up an infrastructure for digital twins. For arable farming those would be digital field twins, while for livestock farming, single animals could be twinned. In such a context, central functions like clearinghouse logging can be done in the twin directly to capsule data with access information. In addition, data usage policies would also be integrated in those twin objects. Such a twin concept supports the idea of a decentralized environment, where twins can exist at any arbitrary platform in the ADS as long as they fulfill the requirements for interoperability and can be used across multiple platforms. This also helps in organizing data of physical assets, as it is no longer distributed across different systems.

Semantic Interoperability

For semantic interoperability, we do not propose another broad, common standard but moreover a common, basic meta-model for principals of data exchange along with flexible mechanisms that supports semantic interoperability between datasets. Incorporating vocabularies and ontologies, like in the IDS, one can integrate conversion functionality directly in the digital twins.

17.2.3 General Benefits of an Interoperable Agricultural Data Space

The discussed interoperable data space, the ADS, supports the digital transformation of agriculture by enabling thorough data sovereignty and interoperability. A common framework for digital twins could enable multiple actors and systems to work collaboratively on specific sets of data, which supports the robustness of processes and enables the ecosystem to develop a reliable data economy. By ensuring data sovereignty in various levels as needed, an ADS can enable the willingness to share data, which further supports the digital transformation of agriculture. More data will become available for new business models and new services that add value to the domain. Still, the data providers can keep control over their data. In the next section, we will describe concrete usage scenarios how an ADS infrastructure can offer benefit to the domain and its stakeholders.

17.3 Application Scenarios

In the following, we describe three different application scenarios where an ADS can provide value.

17.3.1 Sustainable Management of Nutrient Cycle

One application example for the ADS and corresponding services is the evaluation of ecological and economic sustainability via the nutrient cycle. In agriculture, a balanced and appropriate nutrient cycle forms the core of the efficient, productive, and sustainable production of plant as well as animal products. In addition to documentation, e.g., for monitoring by public authorities, the focus is increasingly on the optimization of the nutrient cycle. In our example, we consider a dairy farm with crop cultivation and grassland farming. Optimization can only be achieved by linking different specialist areas. On the one hand, there is agricultural machinery data, which records both the quantities of fertilizer applied and, indirectly via yield mapping, the nutrients applied. Various suppliers of sensor systems record soil, plant, and weather data. In the dairy farm sector, conclusions can be drawn about the nutrients entering and leaving the cycle by looking at data about the feeding together with the milk yield. Fig. 17.3 depicts the general steps in which nutrients are

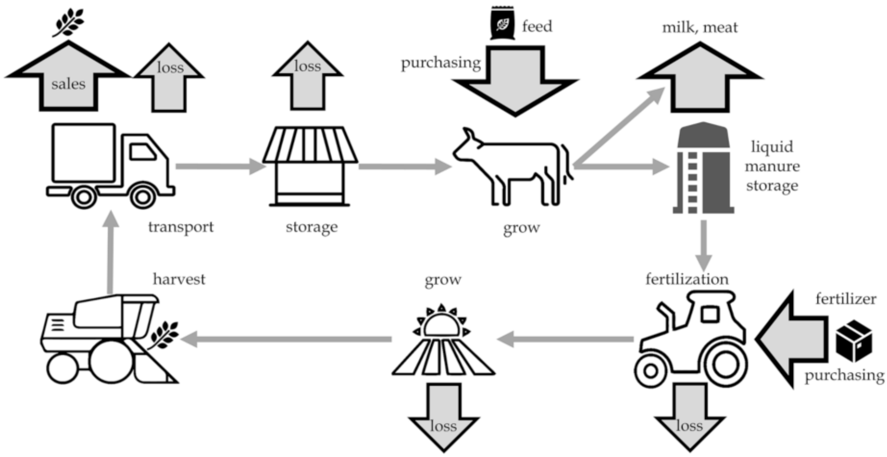

Fig. 17.3 Nutrients are brought in and out at many stages in the nutrient cycle and therefore require multiple measurements and data processing

generated. However, in almost all places (air, soil, groundwater, etc.), nutrients are leaked, and these losses can be reduced to a certain level (although practical implementability must be given and feasible).

In this example, many different stakeholder groups are represented, which either elicit, collect, or evaluate data, or record and store data for legal documentation. In this case, the ADS approach can offer supporting services and at the same time protect the farmer's data, as not all data necessarily needs to be and should be accessed by everyone. In this example, data sovereignty plays an important role. Data is generated from different sources and stakeholders. In order to optimize the nutrient cycle, the relevant data must be complete and must be available in sufficient quality. Important components here are interoperability, uniform ontologies, and cognitive processing of the data. Missing data, for example, must be interpolated or modeled accordingly. By representing the nutrient cycle in the form of a digital twin, the farmer can get information about their current nutrient balance and thus identify possible problem areas. Based on the digital twin, services offering the farmer appropriate decision-making aids can be purchased on the service marketplace for a fee. Since the nutrient cycle is a highly complex representation of various parameters, there are several sub-areas in which services can provide support. Examples include optimal feeding aimed at reducing the amount of nitrogen in the liquid manure, subplot-specific fertilization, or improved utilization of the nutrients in the liquid manure.

17.3.2 New Business Models and Fulfilling Legal Obligations with Data in the ADS

Agriculture is a diverse sector, and farms are part of a complex network with various interest groups. A large amount of different data is generated on a farm, which in the future will be increasingly collected, stored, evaluated, and documented by many different systems. For this reason, the approaches of the ADS with its possibilities for integrated data access and data usage control are of essential importance. The collected data can now be used in various ways bringing benefits for the farmer or the whole agricultural ecosystem. On the one hand, the documented data of the farmer can be made accessible for service providers who have an interest in this data (e.g., to compare yield, evaluate the effectiveness and efficiency of equipment, get data on soil quality, or provide transparency in the food chain). The farmer can participate in the profit made with these data-intensive applications. On the other hand, the documented data can be used by the farmer in order to fulfill legal obligations. Here, the IDS mechanisms of usage control can be used to its full potential: in general, the farmers are very sensitive towards which data from their farming activities will be given to which public authority. On the other hand, the farmers experience a strong burden by many obligations to document activities. Therefore, the farmer can specify usage control policies and give specific public authorities usage rights to obtain the data. It is important to mention that in this scenario the farmer needs to be in full control of this data usage.

17.3.3 Governmental Platforms

Besides the farmers themselves and the companies in the agricultural domain, the public authorities are important stakeholders in the domain. Recently, a study was published on the feasibility of a governmental data platform in the agricultural ecosystem [1]. In this study, various concepts are outlined how the public authorities could build up a (set of) platforms in order to handle data from public authorities and also specific data from the farmers for various purposes. Such purposes are providing information from governmental institutions, providing information on regulations, but also obtaining data from farmers that would like to get subsidies or for fulfilling regulations (see also previous section). This governmental platform can conceptually be seen as an own (sub)ecosystem in an agricultural data space. One result of this study was that the data should not be made available for the consumption by the farmers via portals, but also in a machine-readable, interoperable fashion. In order to exploit the full benefit of interoperability and data exchange in the domain, such governmental platforms should therefore be connected to the nongovernmental IT systems and digital platforms. If this is achieved, farmers as well as companies in the agricultural sector can benefit from the governmental data that is made available.

17.4 Summary and Outlook

In this chapter, we have motivated and presented the concept of an Agricultural Data Space, which can greatly advance digitalization in agriculture. To do so, the Agricultural Data Space takes up the concepts of the International Data Spaces and adapts and extends them with solutions for the agricultural sector.

This data space integrates data and services from different plat- forms without restricting them. Enabling technology is required for this, into which further data and services can be integrated successively, provided that other platforms implement a corresponding connector and describe data access via a service directory.

Many of the elements outlined above address current challenges, but even greater investments are required on the part of providers and users to realize the vision of a common data space for the agricultural sector.

The current activities around the GAIA-X initiative (Chap. 4) which are supported by the IDS can be a strong facilitator to address these challenges. In GAIA-X, agriculture is an own domain, and first use cases are realized, e.g., by the nationally funded projects Agri-Gaia [2] and NaLamKI [3]. Those projects aim to make use of AI components to leverage the potential of agricultural data. For this, the ADS and its concepts can be a foundation to get access to the different data sources. As the challenges in agriculture are typically not solved on national levels, but on a broader scale, and GAIA-X is a European initiative supported by many countries, it has the potential to support the initialization and realization of concepts for a European ADS.

References

1. Bartels, N., Doerr, J., & Fehrmann, J., et al., (2020). *Machbarkeitsstudie zu staatlichen digitalen Datenplattformen für die Landwirtschaft*. https://www.bmel.de/DE/themen/digitalisierung/datenplattform-machbarkeitsstudie.html; zuletzt besucht April 11, 2021.
2. Webpage of the Project Agri-Gaia. https://www.bmwi.de/Redaktion/EN/Artikel/Digital-World/GAIA-X-Use-Cases/agri-gaia.html. Last visited April 14th, 2021.
3. Webpage of the Project NaLamKI. https://www.digitale-technologien.de/DT/Redaktion/DE/Standardartikel/KuenstlicheIntelligenzProjekte/KuenstlicheIntelligenzProjekte_ZweiterFoerderaufruf/ki-projekt_NaLamKI.html. Last visited April 14th, 2021.