Dissertation

Projet VDPA+ : Mieux valoriser les datas des pratiques agricoles

(November 14th, 2023)

Dafa YUMNA

Project tutors : Estelle MILLET[1], Jean-Eudes HOLLEBECQ[2]

MSc-Ag & Food Data Management program coordinator : Dr Jérôme DANTAN

22 place de l'Agronomie
CS 20040 / 91 123 Palaiseau Cedex
contact@alliance-harvest.com

1 Rue Vincent Ballu, 51200 Épernay, FRANCE / 44 Av. Paul Krugger, 69100 Villeurbanne, FRANCE+33 (0) 3 26 51 99 99

[1] Exxact Robotics
[2] Terres Inovia

Acknowledgements

I would like to express my deepest gratitude to all those who have contributed to the completion of this dissertation.

First and foremost, I am immensely thankful to my tutors, Estelle MILLET and Jean-Eudes HOLLEBECQ, for their unwavering support, invaluable guidance, and patience throughout this research journey. Also, to the VDPA+ project board, Hervé NOUVELLON, Vincent RACHET, Xavier PINOCHET, Sophie MARTIN, and Aurélie CORNUEJOLS for their expertise and encouragement in shaping the direction of this dissertation. Additionally, this endeavor would not have been possible without the generous support from the Alliance H@rvest and Exxact Robotics, who financed this research.

I also could not have undertaken this journey without Dr/Hab Salima TAIBI and Dr Jérôme DANTAN who generously provided knowledge and expertise. I extend my appreciation to UniLaSalle Mont-Saint-Aignan for providing the necessary resources and facilities for conducting this study. The Agri-Food Data Management has been a nurturing academic environment that has fostered my intellectual growth. Also, to all my classmates of GMSA-AGROFOOD DATA 2022 for all the supports during my learning process

My heartfelt gratitude goes to my family for their unwavering support and understanding during the challenging phases of this academic endeavor. Their love and encouragement provided the motivation to persevere.

Lastly, I want to express my gratitude to my Indonesian friends and all my colleagues in Exxact Robotics who stood by me during this journey, offering words of encouragement and a sense of normalcy when needed the most.

This dissertation would not have been possible without the support of these individuals, and for that, I am truly thankful.

Dafa YUMNA

Abstract

Nowadays, agricultural practices are transforming into massive utilization of IoT and big data technologies. As a result, more and more information technology (IT) solutions in agriculture are providing solutions to improve farming practices. As this digital revolution is still ongoing, data are being produced and consumed by different actors (government, industry, researcher, public, etc), and with different motivations. However, the current condition of valorizing data is still considered low. Therefore, Alliance H@rvest organized and supported this research, which aims to understand the current condition, identify the challenges, and address the actual need. To answer these questions, first, observation and analysis with a literature review of existing agriculture's digital tools/platforms, existing laws/regulations, actual development of data exchange, and reference from non-agricultural sectors (medical, aeronautics, industry) are carried out to understand the landscape and the requirement for better data valorization. Second, to address the actual challenges, interviews were carried out with Alliance H@rvest Members and relevant actors with a total of 11 respondents (5 from the Alliance, 6 from the external actors). To complete this study, a data fusion use case was conducted with data collected from Terres Inovia decision support tool and available open-source data. The results of this study show that agricultural data are diverse and unharmonized, and the application of FAIR principles is necessary to improve these conditions. This research also indicated that the main challenge today is how to obtain good-quality data and the study case shows that data fusion can easily leverage innovation.

Keywords: *Agriculture, data, digital, valorization, FAIR.*

Résumé

De nos jours, les pratiques agricoles se transforment par une utilisation massive des technologies IoT et Big Data. En conséquence, de plus en plus de solutions de technologies de l'information (TI) dans l'agriculture fournissent des solutions pour améliorer les pratiques agricoles. Alors que cette révolution numérique est toujours en cours, les données sont produites et consommées par différents acteurs (gouvernement, industrie, chercheurs, public, etc.), qui ont des motivations différentes. Cependant, la valorisation des données est encore considéré comme faible. Par conséquent, l'Alliance H@rvest a organisé et soutenu cette recherche, visant à comprendre la situation actuelle, à identifier les défis et à répondre aux besoins réels. Pour répondre à ces questions, une observation et une analyse avec une revue de la littérature sur les outils/plateformes numériques agricoles existants, les lois/réglementations existantes, le développement actuel de l'échange de données et l'analyse de secteurs hors agriculture (médicale, aéronautique) sont en premier lieu effectuées pour comprendre le paysage et la nécessité d'une meilleure valorisation des données. Deuxièmement, pour répondre aux défis réels, des entretiens ont été réalisés avec les membres de l'Alliance et certains acteurs des données agricoles avec un total de 11 personnes interrogées (5 de l'Alliance, 6 des acteurs externes). Pour compléter cette étude, un cas pratique de fusion de données a été réalisé à partir d'un outil d'aide à la décision Terres

Inovia et les données open source disponibles. Les résultats de cette étude montrent que les données agricoles sont variées et non harmonisées, et que l'application des principes FAIR est nécessaire pour améliorer ces conditions. Cette recherche a également indiqué que le principal défi aujourd'hui est de savoir comment obtenir des données de bonne qualité et le cas d'étude montre que la fusion de données peut facilement améliorer la qualité des outils et ainsi générer beaucoup d'innovation.

Mots-clés : *Agriculture, données, numérique, valorisation, FAIR.*

Table of contents

List of tables

List of figures

1. Introduction

1.1 General background and context

1.1.1 Pressure in the agricultural sector

Agriculture has a crucial role in the world, according to the United Nations (UN), The current world population of 7.6 billion is expected to reach 8.6 billion in 2030, and 9.8 billion in 2050. Placing significant pressure on the agricultural sector to produce more food. The next challenge to be addressed is sustainability. This matter has gained the attention of both the public and policymakers. Minimizing the long-term ecological effect by optimizing inputs and assets is the objective (Giray & Catal, 2021). That's why increasing the productivity of the agricultural sector, with minimum impact on the environment, has become a challenge for this modern day.

Based on Eurostat, the numbers of farms are steeply decreasing by 37% from 2005 to 2020. Where the majority (87%) of losing farms are small farms (less than 5ha). Also, the remaining farms are getting bigger. The phenomenon today is farm size getting bigger caused by integration, and numbers of farmers getting smaller. With farm sizes getting bigger, decision-making capabilities on operational methods are important to ensure not only improving crop performance but also reducing environmental footprint. Nowadays, investments in agriculture machinery with data acquisition sensors are being made to make real-time decisions on land or collect and store the data for future analysis. Furthermore, data produced from other sources such as weather, satellite images, and public data, were also exploited to enhance decision-making. This enriches the data for agriculture and creates a big potential for not only environmental but also economic benefits.

1.1.2 Value of data

Based on("Designing Data Spaces," 2022), there are four roles of data. First, as an enabler of operational excellence within companies. Effective and efficient data resource management is required for the integration and automation of business processes. Second, data as a commodity to be sold in the market. It's a common practice to sell and buy data from external parties to complete needed resources. Third, data as a source of innovation. Good quality data are important to the research and development of a product. And in some cases, to develop a product different types and sources of data are required. Moreover, good quality data plays an important role in a feasibility study before starting a project. Bad data can mislead the interpretation and harm the result. Fourth, data is considered as a strategic resource for the long-term sustainability of the economy. That means data can hold a significant value and has the potential to shape the direction of economic growth. Another important point is the value of data can only be discovered when the data is utilized (Otto, 2015). Data is a representative of an event. Well-exploited data could lead to an explanation of that event. Hence, to discover the true value of data, the data has to be circulated and utilized properly. In the modern day, data is not just a secondary product of digital interaction. But it is something that every actor should use as a tool to be more profitable. This is because data holds the potential to drive innovation, inform decision-making, improve efficiency, create new industries, enhance consumer experiences, and much more.

1.1.3 Data utilization in agricultural practices

Agriculture sectors are considered to have a high uncertainty because of weather and environmental conditions, low efficiency in practical and cost, and volatility in balancing supply and demand (Osinga et al., 2022). Currently, agricultural practices have developed and adapted to the massive utilization of machines and IOT technologies. Also, data generated by food systems are believed to shape a transformation towards resilient, sustainable, and food-secure systems globally (Top et al., 2022). Because of that, many agriculture IT solutions are now competing to solve this problem by providing digital solutions. Digitalization in agriculture practices can help farmers manage current existing challenges more effectively, for example by providing insights into efficient water usage, optimal planting times, pest and disease management, and climate-smart practices. By leveraging data, farmers can optimize their operations, improve yields, and ensure sustainable food production to meet the growing demand. Also, big data in agriculture can reduce production costs by 10% (Cattaneo et al., 2020). But in reality, the application of big data solutions for optimizing agriculture practices is not as simple as that. Complexity is the nature of agriculture data, because of its heterogeneity of data types and great diversity of variables types. Hence, there are a lot of companies that propose various IT solutions to describe and explain a biological and physical phenomenon not only to humans (farmers) but also to machines that will run a specific action based on that set of information. Generally, the efforts of reaching sustainable and precision agriculture eventually generate the exponential growth of data being produced, by various sources and for different purposes. To understand better, a comprehensive analysis of the agricultural data landscape for better agricultural data valorization plays an important role.

1.1.4 Introduction to Alliance H@rvest and VDPA+ Project

Alliance H@rvest is a partnership chair supported by AgroParisTech Foundation to develop and accelerate the emergence of digital solutions and respond to the new needs in the agriculture world, by supporting its transition towards a more efficient and sustainable model. Also, this alliance has the ambition to make ‘digital’ a real lever of transformation for agriculture. To achieve its goals, Alliance H@rvest gathered expert actors from different specialties to support an ideal inter-discipline collaboration. Also, Alliance H@rvest organized and funded Project VDPA+ (Mieux valoriser les données des pratiques agricoles) to respond to the actual condition where agriculture data are considered not well-valuated caused of the data consolidating reference model absence. The projects in Alliance H@rvest are divided into 4 categories: data and artificial intelligence, soil and its microbiome, irrigation and common management, and bio aggressors. Also the academic objective by organize conferences and courses. Hence, project VDPA+ is a part of data and artificial intelligent topic. Therefore, the project was divided into three phases. The first phase of the project was to characterize and create a map of existing agriculture digital platforms, the second was to determine needs and creation of added value, and phase three was to create a possible schema of data valorization.

Table 1. Alliance H@rvest's members

Sector	Name
Private	EXCEL Industries La Société du Canal de Provence Sofiproteol
Academic	AgroParistech Télécom Paris UniLaSalle
Experimentation and application	Terres Inovia
Innovation	AgreenTech Valley

Source: https://alliance-harvest.com/

1.2 Problematic

This research was motivated by massive data collection from agricultural machines, open-source data movement, and current data utilization trends and regulations that generate digital agricultural practices. However, we consider that the data itself is undervaluated and not utilized optimally for agricultural practices. The main objective is to understand the current conditions and provide a real example from the study case. Therefore, these are problematics for this research:

a. What exists today in agriculture data? And where do they come from?
b. What are the limitations to valorizing them? And what are the possible actions/solutions to solve this?
c. What are the perspectives of the actors?
d. Can data fusion be performed? What are the benefits?

1.3 Objectives

In order to answer the questions from problematics, several objectives were defined as follows:

1. Evaluate and characterize the existing agricultural data platforms, databases, and sources to understand the types of data available, data quality, and accessibility.
2. Identify gaps, limitations, and challenges in the current agricultural data landscape, including issues related to data fragmentation, heterogeneity, reliability, and public regulation.
3. Apply data fusion methods to specific agricultural case studies to demonstrate their effectiveness in addressing real-world agricultural challenges.

2. Methodology

2.1 Methodology

The research focused on characterizing and analyzing the actual conditions of data utilization from existing platforms, projects, open-source data, previous research, and existing laws. Moreover, interviews are performed to know better the actual challenges and constraints of different stakeholders. After that, a requirements analysis was performed for better data valorization practices. In the end, data fusion study case from a platform's database are analyzed, and describe its difficulties and opportunities.

Table 2. Activities and methods used for the research.

Activities	Methods	Expected result
Identify and characterize agriculture data utilization and valorization	Review (publication, web, database, etc)	State of arts in agriculture data utilization and valorization
Identify actual challenges and constraints from stakeholders	Interview and discussion; questionnaire	Perspective and insight for future development of agricultural data valorization
Data fusion study case	Study case development	Reference architecture, analysis of study case

2.1.1 Review

Review and analysis were conducted to identify, characterize, and evaluate the actual condition of agriculture data utilization and valorization. This step covered actual, ongoing, or future projects/developments about agricultural data utilization and valorization, as well as recommendations to address these subjects. An overview of agricultural data is generated by evaluating and examining relevant topics from different sources of knowledge such as literature, publications, open-source databases, company/government websites, and conferences. Further, a review from non-agriculture sector was conducted as a reference, comparison, and to seek possible adaptation. This process was conducted to understand better and identify different approaches and efforts for data utilization. A literature review is essential for all research not only to seek findings but also as a tool to build and justify arguments (Snyder, 2019). For the research publication, the areas considered in this research are agriculture data, digitalization of agriculture sector, and the FAIR data concept (Wilkinson et al., 2016). Integrating findings and perspectives from various empirical knowledge is expected to answer research questions. Also, an assessment of actual agriculture product/service providers is important to describe the actual conditions of this topic, and that can be done by reviewing the business model or use case from the company website. Output knowledge from this stage is important as a foundation for describing the agriculture data landscape and its potential not only for economic but also for environmental factors. Documentation of available agricultural data in open source was conducted to measure the application of the FAIR concept by assessing the quality of the data. Furthermore, findings from this step are taken into account for determining question parameters for interview, and possible development for the study case.

2.1.2 Interview

The interview process was conducted to seek perspectives from different actors and to investigate the actual challenges, constraints, limitations, and opportunities based on real-world conditions. In total, there were 11 persons for this interview, 5 of them were alliance members and the rest (6 persons) were from external of the alliance. The first targets for the interview were Alliance H@rvest members which their opinions and experiences were considered important for the project. Also, inputs from other actors outside the alliance were considered as well as important to enrich the knowledge from different points of view. These respondents were selected because of their relevancy in this project's scope of work. The process of reaching the respondents is done by connection and recommendation. Interviews and discussions were carried out with direct question-and-answer sessions with relevant stakeholders from different backgrounds. The questionnaires used are designed to cover general issues and are not personalized between different respondents. The opinions expressed by them are considered and analyzed to provide an in-depth understanding and recommendations for better agricultural data valorization practices. As explained before, the questions posed come from the main findings in the review's part, to keep the subject relevant and representative of the actual condition.

2.1.3 Study Case Development

A study case was carried out to analyze and give a viewpoint of the applicable level of data fusion from actual data collected by public applications and open-source data. The interpretation of data fusion is a combination of different components that consist of available datasets, based on context and purpose, and are considered/assumed to improve the performance of subsequent data processing (Beddar-Wiesing & Bieshaar, 2020). In this research, a dataset collected from Terres Inovia's decision support tool was chosen as a base for the data fusion study case. This dataset was produced from a simple decision support tool for providing advice to farmers (users) about the level of risks and the necessity to or not to apply fungicide to avoid stem canker on rapeseed crops. The dataset has the potential to be connected by another relevant source and may be a good example of fusion data. Before performing data fusion, a study about the contextual background and mechanism of data collection was carried out to understand the characteristics and properties of data. The fusion was carried out with climate data, variety data, and registered parcel data with the assumption this fusion may improve the recommendation results by giving more inputs as a consideration for recommendations. Reference architecture was created, analyzed, and evaluated by connecting different open-source databases. Finally, difficulties, constraints, and limitations are explained for future development.

2.2 Workflow

The research was designed to answer the relevant questions with different holistic approaches. A combination of these approaches is expected to not only describe the current condition but also give guidance for further improvements. In order to reach that goal, a rigid workflow was constructed to manage the working progress. Below is the constructed workflow based on the chosen method in Figure 1, also the project's planning and execution in Figure 2. This project was carried out in the form of 6 months internship managed by Exxact Robotics. In the working process, weekly meetings and monthly meetings were conducted to keep the project on track. The weekly meetings were mostly done to solve technical and practical problems with the supervisors (MILLET Estelle; Exxact Robotics and HOLLEBECQ Jean-Eu) and the monthly meetings were designed as a project follow up and discussion for next month's objectives with project committee.

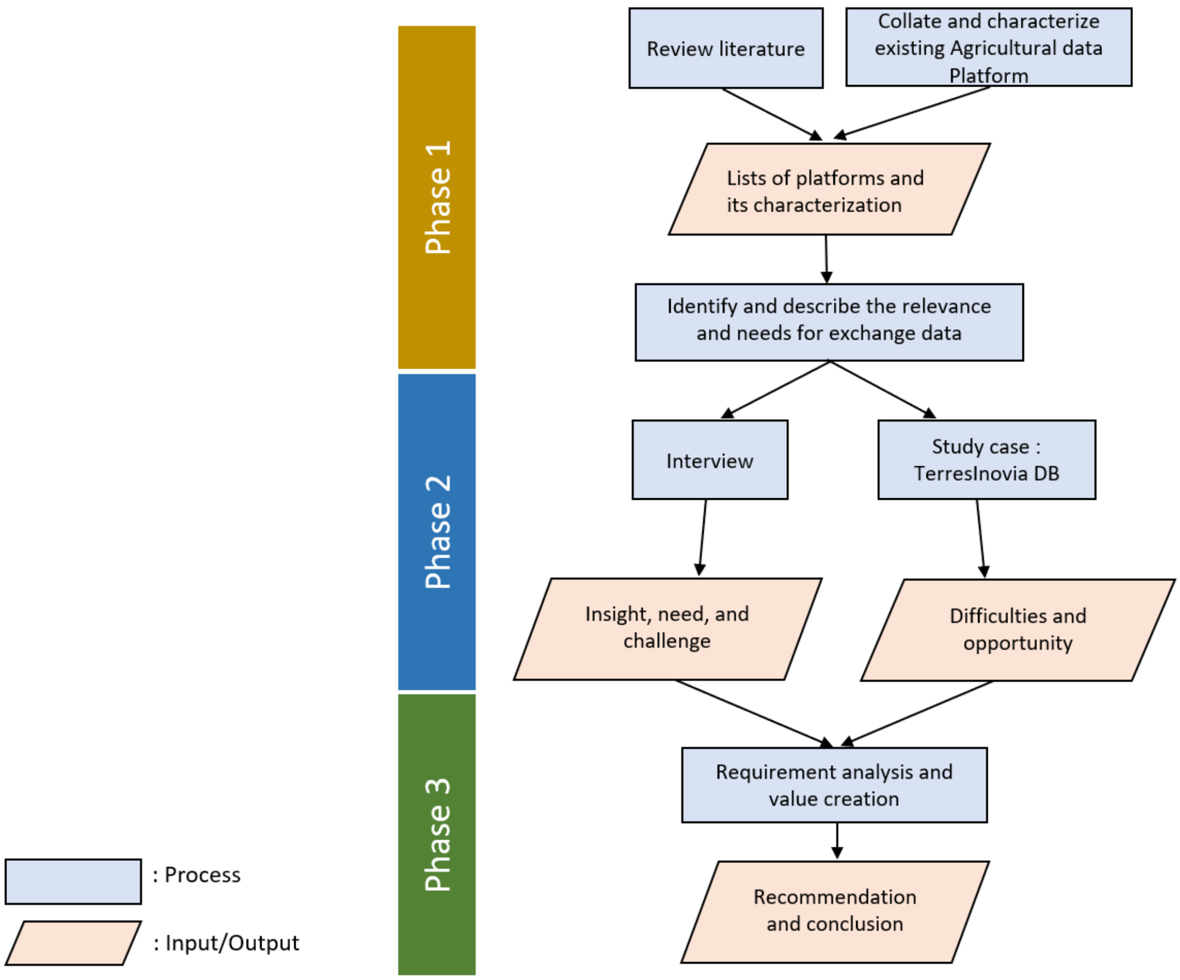


Figure 1. Workflow

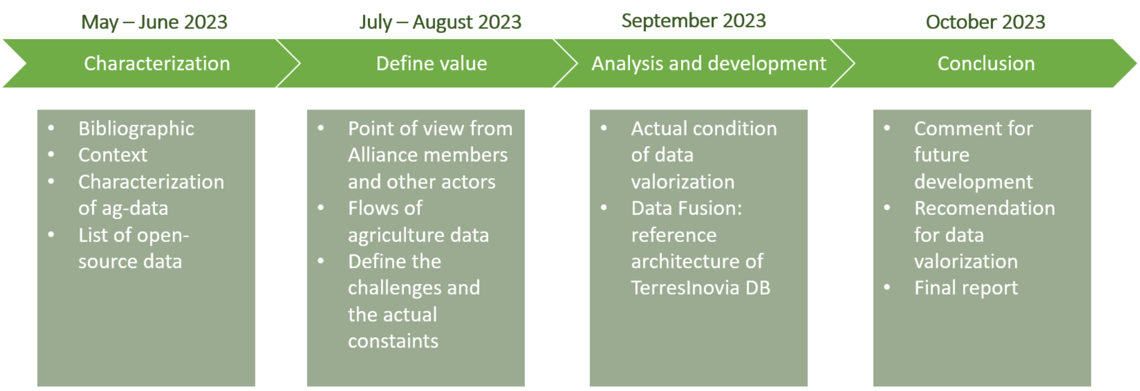


Figure 2. Project's planning and executions

3. Results

3.1 Data collected by agriculture practices.

3.1.1 Source of agricultural data

Agricultural data are collected in various ways with different methods and for different reasons. As a result, agriculture data are scattered and fragmented. According to (Janssen et al., 2017) there are three types of purposes for collecting data. First, governments collect data for monitoring, management, and insight for policy making. Second, research projects collect data for project interest. The data are often generated in different formats (depending on project needs) and collected on an irregular basis. Last, industries (including farmers) collect data for improving their own business whether it's for operational optimization or R&D. The data is often not harmonized, and they mostly don't share data for security and competitiveness reasons. This multi-motivation of collecting data creates the data mostly fragmented and hinders the true potential of the data. Data collected in agriculture can be categorized into three: *Machine-generated, process-generated, and human-sourced* (J & M, 2019). The difference between them can be seen in Table 3 below.

Table 3. Category of data collected in the agriculture sector

Category	Definition	Example
Machine generated	Data generated by agricultural machine or sensor	Satellite image, camera sensor, GPS, Humidity
Process generated	Data generated by certain processes such as monitoring processes or laboratory test	Soil fertility, pest monitoring
Human source	Data generated by information from human (farmers)	Crop rotation, direct observation

Today, data are generated mostly by machines and human sources because of the digitalization movement in agriculture. Data generated by process is relatively high in cost and used only for specific cases. On the other hand, Advanced agriculture machinery produced more data and helped manufacturers or agronomics consultants in creating value for decision-making. There are a lot of platforms and tools that offer various features for helping farmers by giving recommendations. These tools and platforms use data collected from various devices, manually or automatically as the base for giving recommendations. Moreover, generating precise data is important to avoid redundancy, and in the agriculture sector, this is a big challenge caused by the uncertainty of nature. Sometimes, to create a meaningful insight, more than one data is required. But before using them, it's important to understand the source, how frequently they are produced, and for what objectives. An example of diverse data collected related to agriculture practices can be seen in Table 4 below.

Table 4. Example of data collected related to agricultural practices

Data	Collecting method	Frequency	Output	Objective
Yield	Harvest weighing Yield monitor from the harvesting machine	Seasonal	Yield data Yield map	Productivity analysis Storage & supply chain management
Soil	Soil type measurement Fertility measurement Electric conductivity measurement Soil humidity	Once (soil type) Seasonal (fertility, EC) Real-time (soil humidity)	Soil type map Soil fertility distribution map Electric conductivity map Soil humidity sensor	Fertilizer optimization Irrigation management
Machinery, equipment	Performance sensor GPS track	Real-time	Machine performance data Machine work plan	Cost efficiency Fleet management Maintenance scheduling
Irrigation, fertilizer, phytosanitary	Spraying sensor	Real-time	Water use maps Fertilizer use maps Phytosanitary use maps	Cost efficiency Reduce soil contamination Yield optimization
Crop monitoring	Visual inspection Remote sensing UAV	Daily Weekly Each crop-growing stage	Field information Vegetation index map	Diseases & risk monitor Yield prediction
Crop rotation	Manual input	Seasonal	Crop type	Historical report
Farm Spatial data	Georeferencing	Once	Farm boundaries	Farm data properties
Climate	Weather sensor (thermometer, anemometer, barometer, etc)	Daily	Weather measurement/ prediction	Farm recommendation

Furthermore, several organizations provide freely accessible data platforms to support the public liberation of insight. For example, FAO has a FAOSTAT platform where users can download a database about agriculture from all countries (a list of open data platforms can be seen in the Annex). The open data platform provides various types of data and doesn't include sensitive data. This data can be freely used for many applications. Another example, open-meteo.com provides weather data based on location. This data can be utilized to give early warning to farmers for a weather condition change. This can help farmers decide when the perfect time to apply fertilizer or fungicide, when to sow or plant, etc.

The characteristic of open data is accessible to everyone, users can download or disseminate the data, and right to reuse the data. Moreover, there are a lot of agriculture-related data that can be obtained easily and at no cost from all over the world. Each country mostly publishes its statistical agriculture production on an open public platform. However, these statistical data are not too detailed, and they were published only during certain periods of the year. The purpose of this data is to describe the general condition of agriculture production in a country. On the other hand, some open-source data is updated every day like weather data and satellite images. Besides open public data, some farmers trust agronomy consultants to give them advice. Nowadays, there are a lot of platforms and tools that offer various features for helping farmers by giving recommendations. These tools and platforms use data collected from various devices, manually or automatically as the base for giving recommendations.

3.1.2 Analysis of crop yield data flow

As explained before, the data are collected in various ways and motivation. To give a detailed description, an analysis of yield data is conducted by characterizing the scales, motivation, and actors, as explained in Figure 3. This analysis was carried out to understand better the flow of data. The analysis started by defining the scale of data. Scale specified the occupation measure/influence in terms of size. First, the smallest scale of yield data is field level, mostly generated by combined harvester machines to give information about the distribution of production, and in this level, the common format of the data is a distribution yield map. The actors concerned in this type of data are farmers to evaluate and monitor their production, agriculture counselor/consultant may need this type of data to analyze their client's fields, and agtech businesses as input for their services/products. Yield data on the field level are considered important because they can be used as a parameter to describe the fertility of the soil, effectiveness of fertilizer, production capacity, and many more. In general, yield data at the field level are mostly used for precision agriculture practices. In the next scale, there are farm levels that consist of multiple fields. It's a common practice now farmers have more than one field to manage. At this level, the motivation is mostly for farm management purposes, which helps farmers maintain their farms by using clear information about their farm's yield. After that, yield data are scaled to the cooperative level, which is mostly gathered by the cooperative agent for documentation, or selling farmer's product. Furthermore, there are regional scales that give information about yield data regionally, and most of this data is aggregated. The data are collected by the government, by farmer's declaration, surveys, and information from cooperation. These data are published in public, and often used by academic stakeholders for research, and also as a decision-making consideration at the government level. Lastly, the biggest scale is the national/global level where the data are the aggregation of national production, and these data are published in FAO database. To gather this data, FAO sent a survey to each country's stakeholders. Similar to the regional yield data level, this data represents the national production and often used by researchers to develop a model to predict the national/world production.

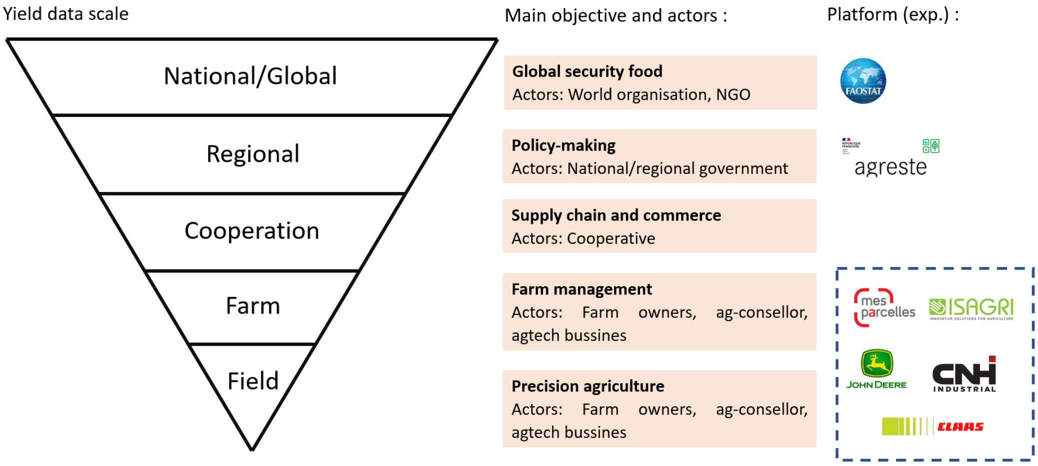


Figure 3. Yield data scale, actors, and objectives for the exploitation

Today, yield data are utilized in many agriculture's digital solutions. A lot of innovation now depends on yield data. As explained, the process of generating a yield map is started with a harvester machine/tractor, and the data is transferred to the product's cloud platform. The data may be used by other decision support services as an input parameter or to enrich the existing model. For example, for creating a map of future potential yield, historical data about previous production are needed. Also, this model may need other supporting data such as historical weather conditions, soil fertility conditions, variable rates applications, satellite imageries, crop rotation, and other possible parameters. On the other hand, yield data also flow to different actors in different scales. It's a normal practice that cooperatives collect their member's data as strategies for marketing the farmer's product or to find a prospective buyer. Then, the government also collects data from surveys to cooperatives, and when farmers declare information for Common Agriculture Policy. In the end, these data are aggregated at regional and national levels. Figure 4 shows data flow's example of yield.

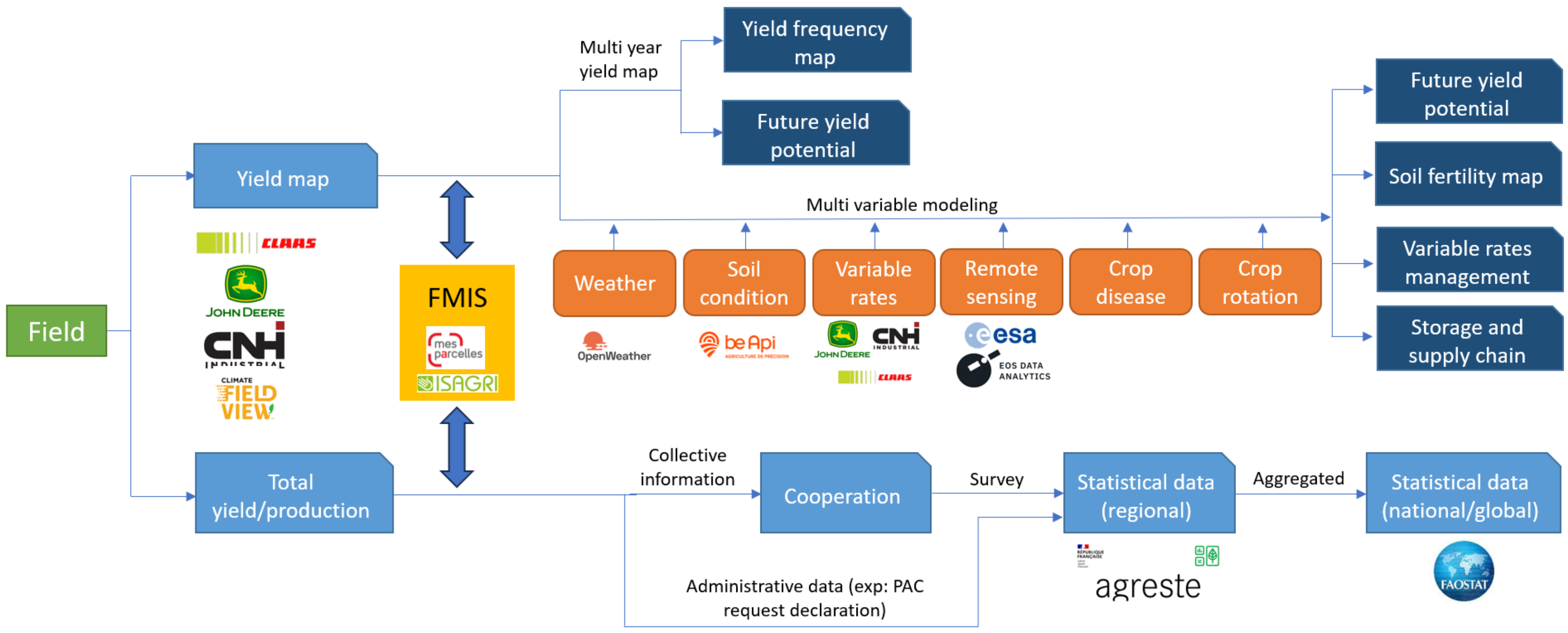


Figure 4. Example of yield data flows

3.1.3 Existing regulations about data

Rapid adaptation of information communication and technology (ICT) in all sectors increases the amount of data and leads to an automation process. Like any other sector, agriculture IT solutions are growing rapidly and consume data from many sources. The idea of protecting personal information and data regulation already existed, especially in the European Union. Massive technological applications and exponential growth of personal data usage affected the requirements of effective legal infrastructure implementation to face the privacy and data security implications (Denitza, 2017). In other words, with all the complexity and voluminous of data, a transparent yet clear regulation is needed to drive data optimization and utilization to release its great potential. Acknowledge the existing or future implementation of regulations about data is important to understand how to respectively treat data and prepare for future adaptation of new regulations. Below are the data regulations in the EU and their impact on agricultural IT solutions.

a). *General Data Protection Regulation (GDPR)*

General Data Protection Regulation (GDPR) is an evolution in EU data protection legislation and has already been applicable since May 25th, 2018. It refers to the businesses, organizations, and other entities established in EU, (or not) but collecting or processing data of EU subjects. GDPR is not a sector-specific regulation, which means it works for all sectors, including agriculture. The main scope of GDPR is protection of personal data, data are considered 'personal' when they can directly or indirectly allow identification of a natural person (Mildebrath, 2023). For the agricultural sector, this may impact IT solutions (FMIS, precision agriculture solutions), researchers who may process data containing personal information, and also for the cooperative, syndicate, or federation in managing their client's data. For example, GDPR mandates require the controller/processor of the personal data to demand consent from the users in an easily accessible form, clear context, and plain language. They also have to notify the authorities within 72 hours if the data breach accident happens and notify the affected individuals with undue delay if the data breach is likely to pose a high risk to their rights and freedoms (GDPR, 2016). In general, the security and privacy of farmer's or user's personal data has become priorities for all actors (private and public sectors, scientists, cooperatives, federations, and others) who handle data from natural person in EU.

b). *Data Governance Act (DGA)*

Data Governance Act was proposed by the EU Commission in 2020, then it was approved by the European Parliament on May 16th, 2022. This regulation will be applicable from September 2023. DGA was formed as one of EU's data strategies to facilitate the data economy and improve EU's digital single market by strengthening the fundamentals of data sharing with compliance of GPDR. The idea is to increase data sharing and reuse of public data, business data, and personal data with the help of trusted intermediaries and with data altruism (data voluntarily made available by individuals or companies for the common good) concept (Data Governance Act Proposal, 2020). Furthermore, to make sure that data is properly protected, public sector bodies must perform anonymization on personal data, and modification of commercially confidential data. Modification could be in the form of aggregation or otherwise disclosure controls. The regulation occupied many actors such as small and medium-sized enterprises (SMEs) to reuse data and encourage innovations. Also, encourages scientific research in healthcare, climate, industry, and other sectors (Ruohonen & Mickelsson, 2023). For example, in the agricultural sector, SMEs may re-use farmer's (user's) data using trusted intermediary services, with the consent of the farmers. Under DGA, the data intermediation service must notify the competent authority before can legally work as a trusted intermediate of data exchange between individuals or enterprises and data consumers.

c). *Data Act (DA)*

Data Act proposal was adopted on 23 February 2022, and proposed regulation to harmonize rules on fair access to and use of the data. The idea is to optimize the potential of data by making it more available for public interests under certain conditions. The Data Act clarifies who can create value from data and under which conditions. In the agriculture sector where IT service-connected products are used for optimization, it might give more opportunities to users (farmers, cooperatives, related service providers) to use more of the data from private sectors. However, only non-personal data will be available to be used and processed by the public interest. The Database Directive direction also gives legal uncertainties protection to substantial investment databases that contain data generated by related service products such as sensors, machines-generated data, etc (Data Act Proposal, 2022). On 28 June 2023, the European Parliament and the Council of the EU on the Data Act agreed on a Political Agreement, which means the Data Act will be applicable after 20 months of Official Journal publication.

3.2 Map of agriculture data platform

3.2.1 Digitalization of agriculture sector

The term digitalization refers to the adoption of information communication technology (ICT) to improve the utilization of digital content and the development of services and applications (Mcfadden, 2022). For the past decade, the agriculture revolution, also known as 'Green Revolution' has driven industries to the massive development of new digital technologies and services in the agriculture domain. Supported by the idea to reduce input and produce more, many studies have been published towards 'greener' crop farming practices, and many of them have been adopted into campaigns or integrated into new services. The main goal of the utilization of digital agriculture technology is to help users to increase their productivity by providing or improving crop or livestock management practices. Digitalization activity clearly requires data as an input and may produce new data as well. Thus, the relationship between digital services and data is remarkably close. Thereby, (Mcfadden, 2022) identified that technology developers are one of the key actors, besides farmers. Numerous technology developers with different levels of technology, types of services, scopes, pricing, and many more have created a diversification and differentiation of products and services. Therefore, the identification of existing digital services and products in the agriculture sector is important, as an effort to understand before valorizing the agriculture data. There are several perspectives to understand the nature of digital tools in agriculture.

Sunding and Zilberman (2007) distinct the digital tools into *embodied* and *disembodied* innovations.

Embodied: Sensors are integrated with the machine. Embodied digital tools may be interpreted as precision farming or precision animal husbandry practices, for example, integrated sensors in machinery equipment to detect weeds, or tracking harvest production, in livestock to detect stress levels or diseases (Birner et al., 2021).

Disembodied: Refers to unintegrated to farm-specific pieces of equipment's digital tools. Mostly digital software apps such as advisory apps, farm management systems, and online recommendation systems.

Based on the technological level, Daum T et al. (2022) explained two types of digital tools, simple and smart digital tools.

Simple tool: Give access to generic information such as infographics on when to sow or harvest, tutorials on how mix a feed for livestock's nutrient, documentation of insects and disease, etc. Although this tool has a limited capability, sometimes it's still used because of its generality and low cost.

Smart tool: Allow data collection from farms and livestock manually or automatically.

On the other hand, connected smart digital tools allows data sharing with other tools with different directions and targets (Porter & Heppelmann, 2014).

Moreover, data can be shared on a one-to-one, one-to-many, many-to-one, or many-to-many (Daum et al., 2022).

one-to-one: An example of a one-to-one basis application, when a product or tool is connected to another product, user, or manufacturer, is when a farmer's machine generates data and connected to an agronomy consultant or another partner.

one-to-many: For one-to-many, it's a normal practice when a digital services provider gives information/updates to all users about the tools.

many-to-one: On a many-to-one basis, this activity is related to the collection of data from users to tools developers, and normally it's written on the leasing contract when the users first buy the tools, they may agree/not agree to transfer their data.

many-to-many: data transfer occurs from one large entity to another. An example of this is when customers/buyers are connected with farmers or cooperatives and mostly can be found in digital marketing platforms or traceability product innovation.

All types the connections are designed to improve agricultural practices, and simultaneously this big data-related activity creates an opportunity for the data processor/controller to develop an innovation. The following characterization of digital tools based on their functionality describes their heterogeneity and the data as well. The evolution of the digitalization system can be seen in Figure 5 below. The figure describes the connectivity of several smart and connected products in one product system, which often produces data and its connectivity to other digital tools in FMIS. In the end, the data intermediary becomes a 'bridge' for data exchange.

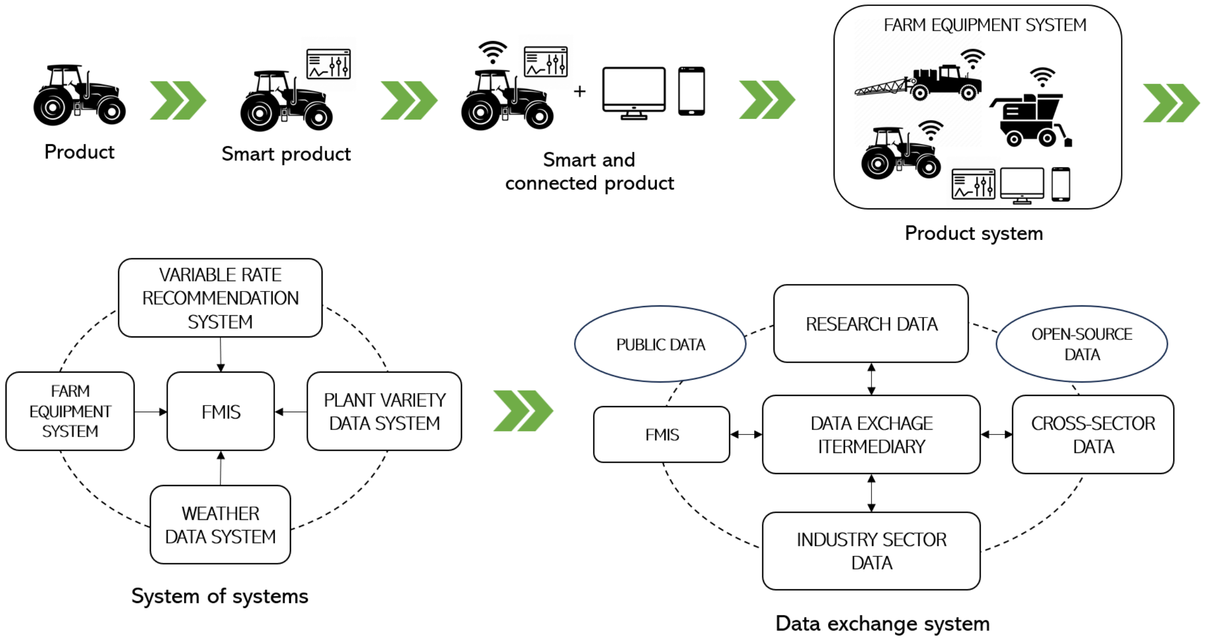


Figure 5. Evolution of digitalization's system from a product into data exchange system
Source: Author, referred to Porter and Heppelmann (2014)

3.2.2 Characterization of agriculture digital tool

In total, 36 digital services and 27 open databases have been listed and analyzed (see Appendices). Based on exploration, the majority of the available digital tools in France are proposed services dedicated to farmers to help them manage their farms/livestock or can be considered as FMIS (Farm Management Information System). The features proposed are mostly related to the visualization of farm conditions, yield monitoring, and variable rate application scheduling. Also, some digital tools are specialized to solve more specific problems such as giving recommendation maps on specific needs of fertilizer, soil fertility mapping, and recommendations on fungicide application. On the other hand, digital tools or platforms that provide services as trusted third parties are relatively new and still outnumbered compared to other digital tools. For the open database, the data type provide are varies. The data type can be categorized into three types; structured, semi-structured, and unstructured (Cravero et al., 2022). Structured data can be found in FAOSTAT where the data is organized in logical structure. Semi-structured data can be found in open-meteo (weather database) where the data is formatted in JSON and changeable when there's new data inputted. Lastly, sci-hub database provides satellite images are categorized as un-structured data. These data may be used by digital provider to improves their decision-making. See Table 5 for a detailed distribution of analyzed digital tools. Illustration of digital tools can be seen in Figure 6 below.

Table 5. List of digital tools analyzed

Type	Count
Farm equipment	4
FMIS; field and herd operation management	9
FMIS; Best practice (precision farming)	8
Agriculture digital marketplace/sales	4
Decision support tools platform	3
Data exchange and consent management tool	2
Management and administrative	3
Cloud insfrastructure services	3

(data on October 30th, 2023)

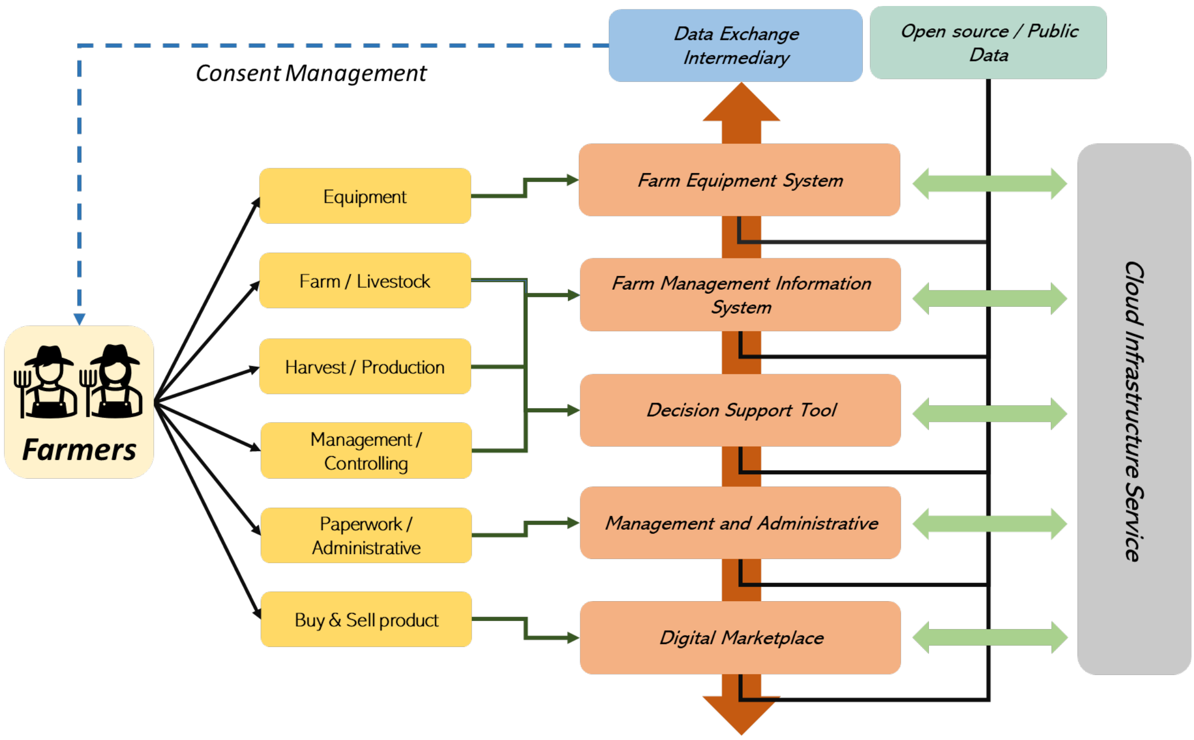


Figure 6. Illustration of agriculture's digital tool landscapce

Farm equipment system generate data from machine's integrated sensors connected to product cloud services and help farmers to organize their resources. Farmers may improve their farming practices by managing, recording, and monitoring their machine's activities and performance. Also, farmers can monitor their crop production. Farm equipment digital tools are mostly distributed as a buying package when farmers buy equipment for the first time. This tool mostly needs a farm border as an initial input, and some parameters need to be stored manually such as crop rotation, fertilizer specification, and other details of work (sow, harvest, spray, etc). The output data from these tools are machine performance details (fuel consumption, speed, working duration), real-time GPS location, also some agronomic data such as yield maps and variable rates maps. The manufacturer collects the machine data for product/service improvement from the users. This tool also mostly provides plug-in features to integrate into other digital services. Nowadays, farm equipment systems can connect to not only their product but also other products from different manufacturers. Examples of this tool in the market are MyJhonDeere by Jhon Deere and MyPLM Connect by New Holland.

Farm Management Information System first was introduced for simple farm record keeping and operations planning, now evolved into sophisticated and complex decision-support software to support production management (Fountas et al., 2015). Farmers can optimize their farming and practices by reducing the application rates, well resources management, detecting crop diseases, finance (sales and accounting) management, product traceability, and equipment management. In general, FMIS enhances farmer's decision-making by providing actionable information and helps farmers manage their resources. To do that, FMIS requires input data to be processed and the data may come from various sources such as sensors, manual input, open source, another FMIS/tool, etc. Normally, FMIS allows a connection from other digital tools to promote decision making, and for record keeping. Also, FMIS often consumes data from open-source data such as satellite images and weather data. With the challenge of Volatility, Uncertainty, Complexity, and Ambigiousity (VUCA) of agriculture domain, it exists numerous FMIS with different service specifications in the market which propose different approaches and methods to solve these challenges. For example, BeApi proposes a soil mapping solutions approach to increase farmers' productivity and optimize the application rates. Farmstar gives fertilizer recommendations based on satellite images, chlorophyll, and biomass measurements. There are also dedicated services to farm management to record activities, and monitor yield, fertilization, and phytosanitary scheduling such as GeoFolia by ISAGRI and MesParcelles by Chambre d'Agriculture. The data generated by FMIS might be useful and important to other stakeholders. Therefore, it must be handled properly since most of them are considered personal data.

Digital marketplace for agriculture provides a platform where a distributor can sell their product/services such as fertilizer, seeds, agriculture equipment, and even digital tools to farmers or cooperatives. An example of this digital marketplace is Aladin.farm (https://www.aladin.farm/).

Public decision support tool is mostly dedicated to the specific agricultural problem. These tools help farmers by producing a recommendation based on required parameters. The difference with FMIS is to use these tools, users mostly can have an access freely or by account creation. For example, Terres Inovia developed free tools in their platform, and one of them is a decision-support tool for rapeseed fungicides to give the level of risk and recommendation to/not to apply the fungicide based on some data filled by farmers.

Management and administrative dedicated tools are related to help users manage their paperwork from farming businesses. Sometimes farmers have to deal with contracts for the use of machines from agriculture contractors services or archiving their documents. Digital documentation may help farmers to securely keep their documents.

Data exchange intermediary is a service developed to solve data exchange problems such as privacy, security, and interoperability by providing a trusted platform for data consumers and data providers. Under EU vision and regulation of data, the data exchange intermediary became a solution for stakeholders that willing to exchange their data and may help them to enrich or gain new knowledge, develop a product/service, and for innovation.

Cloud infrastructure service is mostly used by software developers to develop infrastructure for digital tools. Big companies who provide these services such as Microsoft and Amazon now looking to agriculture as a potential growing sector.

Public and open-source provide non-personal data and can be used for providing general insight. Farmers may benefit from these free access tools to gain knowledge and as a consideration for decision-making. Therefore, the simple features proposed are dedicated to general/public needs, then the outputs mostly are not detailed knowledge. Some online platforms provide public information for total production in some areas, number of farms, total farm area, etc. These platforms are mostly managed by public sector. Satellite images and climate data are often used to solve agricultural problems. Satellite images can be accessed through Copernicus Sci-hub and climate data can be accessed freely on not commercial reasons on the Open-Climate platform. For commercial reasons, the developers have to pay a certain amount of money for the services. Public and open-source data are often used for research and may hold the potential for better data valorization.

Nowadays, product innovation in agriculture has evolved. Many manufacturers develop their connected systems not only to give the best services to the users but also to add a competitive level of product. In the modern era, exclusive personification services approach is a common practice in developing businesses, and in agriculture sectors as well where services are adjusted based on client's needs. Moreover, with the digitalization trends the data produced is enormous. The challenge of data valorization should be analyzed from the very beginning process before the data is transferred or reused. Analyzing the root problem can accelerate the data valorization process by increasing the data-driven solution and data economic value.

3.2.3 Findable, Accessible, Interoperable, Reusable (FAIR) principle on agricultural data

Before we can generate some knowledge, the resources have to be findable and accessible. Then, with the implementation of standards we can have interoperability and permit the reuse the data (David et al., 2020). This FAIR principle (Findable, Accessible, Interoperable, Reusable) was first introduced on 2016 by Wilkinson et al. Important to notice that FAIR principle is not a standard, but a guideline to encounter big data issues by making the data itself more reachable and useable. The application of the FAIR concept leads to better valorization of the data, as a door to open the potential knowledge.

a). Findable

The ability to easily discover and locate a dataset is the main interpretation of the findable. As we know, humans and machines face a different problem when searching a data on the Web. Humans often can find the right dataset with the help of sense to identify context through visualization, word interpretation, basic knowledge, and other aspects. Machines mostly have difficulty in identifying relevant data because of insufficient data information or metadata. The difference is machines have a bigger scope and higher speed to search and identify various formats of datasets (Wilkinson et al., 2016). Agricultural data is often scattered across various sources, including research institutions, government agencies, agricultural companies, and individual farmers. This fragmentation can make it difficult to discover and access relevant data as there is no centralized repository or standardization of data sources. Agricultural datasets often lack comprehensive and standardized metadata. Incomplete or inadequate metadata makes it challenging for users to understand the content, context, and applicability of the data. However, there exists a database platform/repository for searching agriculture data. This data catalog platform can help users find a desired dataset by redirecting to a registered organization's website or directly providing the dataset. In addition, there is a website that provides explanations and guidance for FAIR data (https://www.go-fair.org/). In general, agricultural data especially weather data, spatial images, country statistics, and research publications are mostly findable. Because they have a unique identifier and are registered/indexed in searchable resources. Despite that, various agriculture data are unavailable or difficult to find such as very specific data and very old records. Those data possibly aren't transformed into a findable format, or even do not exist in searchable resources.

b). Accessible

Accessible means that the data are retrievable by their identifier using an open, free, and universally implemented communication protocol. The efforts to establish accessible data in agriculture are still in development. One of the accessible realizations is an open data portal. Various platforms provide accessible data that can be retrieved freely. However, some platforms require an authentication and authorization procedure to access the data. This scheme is also part of the accessible aspects because sometimes the data are sensitive and require some action for security and privacy. In general, open public agriculture data are mostly accessible but some platforms give a limitation of access.

c). Interoperable

The ability to exchange and interpret data between humans, machines, or human and machine are necessarily important in terms of interoperability. The idea of interoperable in the FAIR concept is to increase the possible interpretation not only for humans but also for machines. Moreover, the efforts to reach data interoperability are to provide different levels of data or metadata, using a standardized format. By providing different levels, it gives a higher chance that the (meta)data can be used based on preferred levels or formats. For example, by providing more than one format in a (meta)data, such as JSON, CSV, and PDF to extend the scope. This can leverage the readiness levels of data, and also accommodate humans and machines. Today, we can say that lack of interoperability has become one of the problems, especially in agriculture databases. Based on the exploration of open-source databases, incomplete metadata, and single data format still appears in some database platforms. Another aspect related to interoperability is data standardization, which plays an important role as a data exchange fundamental. Today, standards do exist and are applied to many specific sectors. Top et al. (2022) explained that the challenge today is to harmonize the white spot or gap between existing standards. For example, the EU opensource database published a new consolidated plant varieties catalog, and France has a GNIS code for the plant variety but there are no links between these two standards. So, standardization doesn't always mean creating a new standard or method, this action can be done by formalizing and linking existing standards, and standardization should be done only on necessarily related subjects. For example, there's no urgency to standardize plant genotypes and machine spare parts. As we know, agriculture data is diverse and will always grow in line with increasing world needs for food. Also, notice that data standardization is not only the solution for interoperability. A digital service that can read, interpret, and translate data could be a solution aside from data standardization.

d). Reusable

Maximizing the value and impact of data could be established by performing data reusability in promoting, sharing, integrating, and repurposing data. Reusable refers to the ability to use data or resources for multiple purposes and by different stakeholders beyond their original context. To achieve that, the (meta)data should be enriched by relevant attributes that describe accurately the context of the data. Also, the quality of the data is important in this matter, good quality data is more reusable. Reusable data can save time and cost by avoiding repetitive work on data, especially in agriculture. A huge variety of data is the main challenge today in agriculture. In applying reusable data, one data may be used for different purposes. The efforts to increase reusable agriculture data were initiated by creating open public data and data-sharing practices. Therefore, there's also a boundary to reuse data, normally when it's concerned about privacy and security. So, traceability for sensitive data is also considered important in this aspect. Nowadays, open public data are mostly reusable for general purposes, however, there is also data published in open platforms that are not compatible with reuse. Moreover, the details of FAIR guidelines are described as follows (Wilkinson et al., 2016):

<u>*Findable:*</u>	F1. (meta)data are assigned a globally unique and persistent identifier
	F2. data are described with rich metadata (defined by R1 below)
	F3. metadata clearly and explicitly include the identifier of the data it describes
	F4. (meta)data are registered or indexed in a searchable resource
<u>*Accessible:*</u>	A1. (meta)data are retrievable by their identifier using a standardized communications protocol
	A1.1 the protocol is open, free, and universally implementable

A1.2 the protocol allows for an authentication and authorization procedure, where necessary

A2. metadata are accessible, even when the data are no longer available

Interoperable: I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.

I2. (meta)data use vocabularies that follow FAIR principles

I3. (meta)data include qualified references to other (meta)data

Reusable: R1. meta(data) are richly described with a plurality of accurate and relevant attributes

R1.1. (meta)data are released with a clear and accessible data usage license

R1.2. (meta)data are associated with detailed provenance

R1.3. (meta)data meet domain-relevant community standards

The effort of FAIR implementation can be found in the European Union (EU) Data Portal (https://data.europa.eu/en), a platform to share numerous datasets collected from its members. In total, there are more than 1.5 million datasets stored on the platform and it's freely accessible by everyone. Finding the desired dataset can be done by filtering by keyword, year published, format, category, also quality of metadata. The identifier used for the dataset such as URI or URL can be found in metadata based on which publishers. There are some datasets also available in different formats such as CSV, JSON, or XML. This supports the concept of interoperability since the published format is compatible or in machine-readable format. The metadata also contained the license for data use, giving clarity about the right to reuse the data. However, there's a possibility that users obtained data with unmatched encoding. It occurs because of the difference in encoding defaults on individual personal machines. In agriculture, the implementation of FAIR is necessary to accelerate the digitalization of agriculture, since agricultural data are remarkably heterogeneous and complex. Nowadays, technology such as computer vision, artificial intelligence, and IOT are being implemented for agriculture and these technologies rely on data. Nevertheless, the collective implementation of the FAIR agriculture data concept will stimulate innovation and simultaneously create a sustainable data ecosystem. Besides, FAIR guidelines are applicable from a basic level of data management process. Currently, many data publishers in France started to apply data FAIRness in their platforms and it's indicating a positive trend for the upcoming period. Thus, the FAIR concept may be adopted especially in the agriculture sector not only as practical guidance for good data management but also as a stepstone towards digitalization and better valorization of agriculture data.

3.2.4 Data exchange initiatives.

The idea of exchanging data with exchange facilitators such as third parties or intermediaries has already existed before. Schweihoff et al. (2023), revealed that there are many types of data intermediaries, such as data marketplaces or data trusts, and also different functions and frameworks. But in general, data intermediaries work as data enablers between data providers and data users. Another important modality of a data intermediary is to provide governance service between two parties to give authority on data protection and security. Governance is an important aspect of data sharing, to regulate the legal access of the data. Besides governance, the services have to be compliant with existing laws and regulations about data exchange (GDPR, DGA, DA). It's important to provide a trustful platform because the fundamental reason to use an intermediary is to resolve conflict by using neural, and legal platforms. Even if the platform is secured, another question about the data quality is raised. Data intermediaries mostly provide exchange at the B2B level, to smoothen the process of exchange, data quality must be ensured by the data provider before it can be reused by data consumers. Also data intermediaries have to ensure that their platform are supportable in providing correct metadata . FAIR concept can be adapted as a guideline, as explained earlier. An illustration of data exchange schema referred to AgDataHub model can be seen in Figure. 7.

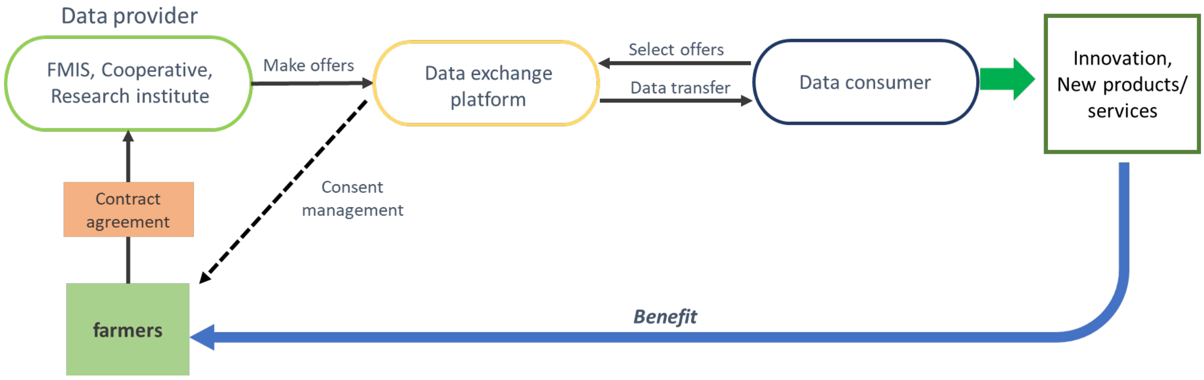


Figure 7. Data exchange schema with intermediary

Application's users are the actors that produce the data. Then, their data is stored in the product's or agriculture digital solution's cloud (digitalization process). This process happened when farmers leased a smart connected machine and then agreed to give data access to the manufacturer/dealer, or when farmers subscribed to an FMIS service. Data providers may put an offer in a data exchange platform, but before that, there are several aspects to notice. First, data providers must ensure that the data offered is compliant with the regulation. Pseudonymization, anonymization, aggregation, or censorship can be done to prevent data misuse. Second, the data or metadata are clear enough, and following the FAIR concept would be better. As soon as the offer has been published and the data consumer is interested, the data intermediary can be the agent to ask for consent agreement from the users and consolidate a contract between the data provider and data consumer to apply government and sovereignty. Then, the data transfer process can be done. Data consumers may use this data as input for their operational process, develop a product/service, feed their algorithms, and many more (the regulation about who can create value and in which conditions are defined in Data Act). In the end, not only data consumers will gain the benefits, but the users will also benefit from a better products/service, higher accuracy of decision support model, etc.

However, the conditions in reality are very complex. Nowadays, B2B data transfer is done mostly without using thirds parties. Different digital products and services from various digital solutions and their technologies' maturity impacted the way they transferred the data. For example, in a farm equipment management system, there exist multiple brands with different levels of software infrastructure, and some of them could be integrated into a specific system. Then, data integration from one farm equipment system to another digital solution is interoperable with plug-ins and some are providing additional installable equipment, or even still exchange with hard disk. On another level, FMIS may exchange with other FMIS directly by creating sharing clouds, and some FMISs use trusted third-party services. The illustration of the complexity of data exchange/sharing can be seen in figure 8.

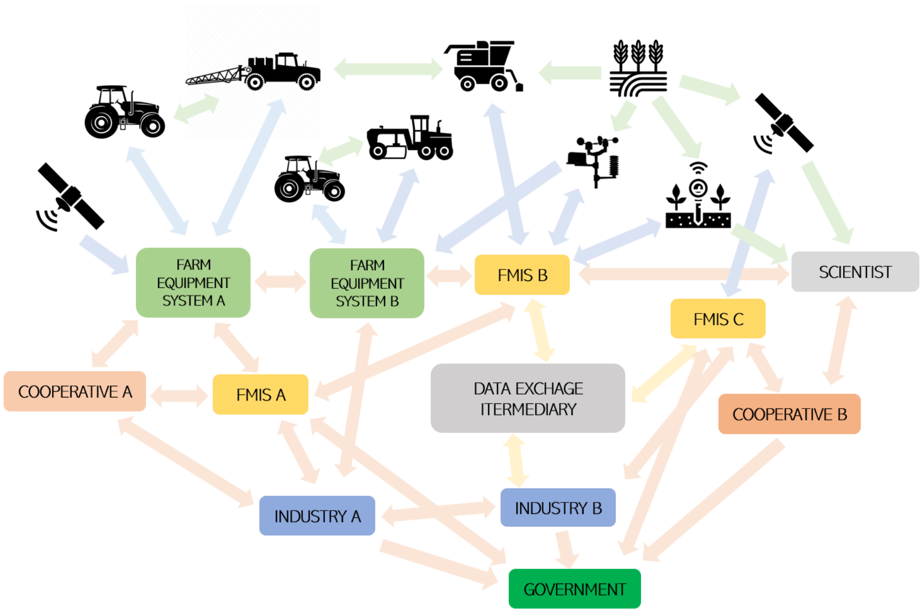


Figure 8. Illustration agriculture data exchange complexity

A concept was designed based on the existing solution to improve the performance and to better valorize agriculture data. The data exchange conceptual system proposed below might be a radical solution for leveraging data valorization in the agriculture sector. The concept covers all actors and opens the opportunity for innovation while respecting the rights of each stakeholder. This concept also allows connectivity with other sectors, which can lead to cross-domain collaboration. It's important to note that the foundation of this concept was constructed on FAIR data application. The process of data exchange must be done in accordance with existing regulations. The regulations about data and data exchanges are expressed in GDPR, DGA, and DA. Processes that are concerned with personal data must be treated properly. Then, data must have the ability to be read easily by humans and machines. To achieve that, metadata must contain relevant attributes to support the exchange process. Another important point is data intermediaries should not have a right to access and modify the data. Figure 9 below illustrates that the FAIR data concept should be applied as soon as possible when digital tools collect data from users, and before the data becomes available in the intermediary platform. It's important to remember that the FAIR data concept should be applied even if the data are not shared because it would make data management and data stewardship much easier (David et al., 2020).

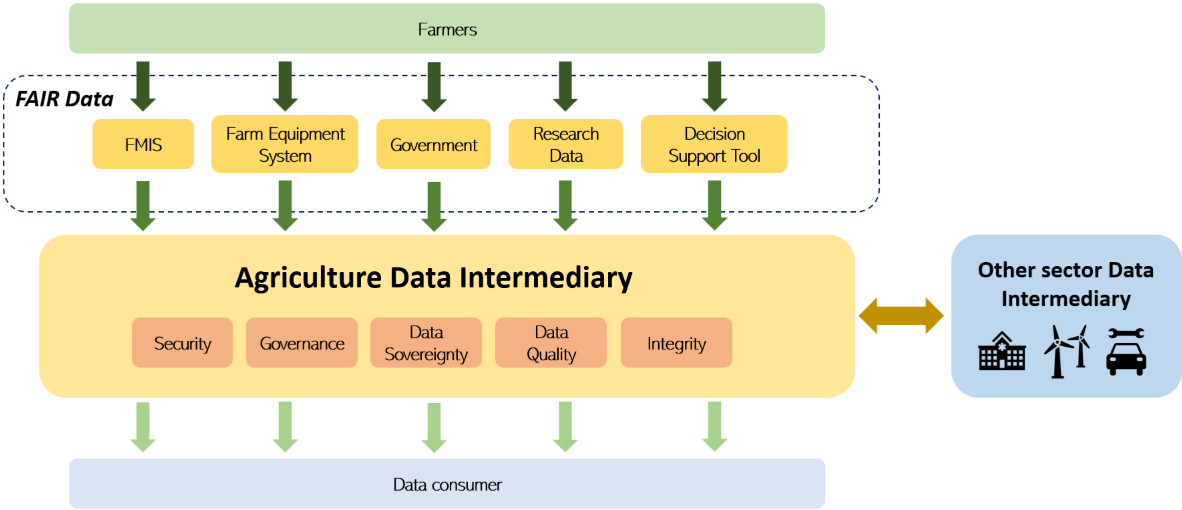


Figure 9. Data exchange concept system

3.2.5 Existing projects/efforts in interoperability and data exchange/saring .

Data variety is a natural consequence of the industry's growth. Moreover, the lack of interoperability in agricultural field operations makes it hard to extract value from data (Craker et al., 2018). To overcome this problem, some initiatives are done in terms of interoperability. The main purpose is to facilitate communication between tools and system information, using a standardized manner. ADAPT (AgGateway) consortium was created by 200+ agriculture industries to solve machine interoperability problems by replacing the current, fragile situation, where FMIS must support multiple hardware data formats, and each machinery manufacturer has to interact with multiple software companies. ADAPT is an open-source framework to support interoperability and can be applied inside a digital tool. COGNAC project initiated by Fraunhofer IESE providing software architecture reference to connecting agriculture data ecosystem. This project addresses the status quo in the digital ecosystem of agriculture by developing a comprehensive reference architecture (COGNITIVE AGRICULTURE, 2020). Demeter is a Horizon 2020 project (857202) supported by the European Union. This project aims to accelerate digital transformation in agriculture through the rapid adoption of advanced IoT technologies, data science, and smart farming, ensuring its long-term viability and sustainability (Demeter: Testbed Deployment, 2020). DataConnect (https://www.365farmnet.com/en/products/dataconnect/) initiated by CLAAS, Jhon Deere, CNH, and myfarm365 proposes non-proprietary cloud-to-cloud solutions to connect different machine manufacturers' portals. AgDataHub (https://agdatahub.eu/) is an agriculture data exchange intermediary developed by several organizations to provide secure data exchange between actors. This platform allows secure data exchange with API. The efforts to solve interoperability in the agriculture sector already existed a long time ago. In 1992 AgroEDI Europe (https://agroedieurope.fr/) association was created to facilitate economic and administrative data in the agriculture industries transfer using a standardized format. Below in Table 6 are listed projects in agriculture interoperability, transformation, and exchange.

Table 6. Projects in agriculture interoperability, transformation, and exchange

Project/ Association	**Initiator**	**Objective**	**Output**
ADAPT	AgGateway	Machine interoperability	Installable plug-ins
COGNAC (Cognitive Agriculture)	Fraunhofer IESE	Connecting agricultural data ecosystem	Software architecture's reference
DEMETER	European Union	Encourage farmers to implement IoT, extending interoperability specially machine to machine communication	20 experimental pilot projects in five categories
DataConnect	CLAAS, Jhon Deere, myfarm365, CNH	Unified cloud platform for interoperability between machines.	Cloud platform
AgDataHub	NumAgri	Focus on consent, data exchange, and data storage	Agri-trust: Consent for data exchange API-Agro: B2B data exchange platform
AgroEDI	La Coopération Agricole, l'Union InVivo, l'Association Nationale de Révision.	Data interoperability	DAPLOS, eDAPLOS + référentiels

Besides these existing projects, governments in other countries around the world also initiated a project to leverage data sharing and data openness. This happens because data circulation is considered important and may leverage the value of the data itself. The European Commission announced the European Strategies for Data in 2020 to create an efficient and secure single market for data sharing and exchange across sectors within the EU (Nagel & Lycklama, 2021) .For the agriculture sector, AgriDataSpace (https://agridataspace-csa.eu/) was an ongoing project to describe the digital agriculture landscape and define a roadmap toward the implementation of EU data space. United States Department of Agriculture (USDA) initiated the Open Data Framework aimed at building a framework to foster agricultural innovation and to support technological progress, production efficiencies, and environmental stewardship (USDA, 2023). This project is willing to ensure the creation of a neutral and secure data repository where producers, universities, and non-profit entities can store and share data. AgriGaia was funded by the German Ministry of Economic Affairs and Energy to build an agriculture AI ecosystem by implementing an innovative B2B platform for AI development in the agri-food sector

(*Gaia-X Position Paper*, 2021). Australian AgriFood Data Exchange (https://www.integritysystems.com.au/ozdata) designed by agrifood industry aims to enable safe data exchange to generate valuable insight and increase the industry's ability to compete in the global market. The Agriculture Data Exchange (ADeX) (https://adex.org.in/) is a collaborative project initiated by Telangana's Government and the Indian Institute of Science (IISc) that aims to allow data-driven services by facilitating data availability and exchange. In Belgium, a data space platform called DJustConnect (https://www.djustconnect.be/en) was developed by the public-private sector to make data available and provide safe, transparent, and controlled data transport. Table 7 below shows the details of these projects.

Table 7. Projects in agriculture data on other countries

Country/region	Project	Description	Status
European Union	AgriDataSpace	Provide a complete conceptual reference architecture, and reference technology for navigating heterogeneous repositories. Assess and share knowledge on those technological requirements.	Ongoing project 1/10/22 – 31/3/24
USA	Open Data Framework	A framework to create a neutral and secure data repository and cooperative where producers, universities, and nonprofit entities can store and share data	Request for applications 6/4/23
Germany	AgriGaia	B2B marketplace, enables a simple exchange of industry-specific data and AI algorithms, which are described with uniform standards	Ongoing project 1/1/21 - 12/31/23
Australia	Australian AgriFood Data Exchange (AADX)	An initiative designed and overseen by members of the agrifood industry to enable participants to share, reuse, and merge data from disparate systems in a secure environment on a permission basis.	Ongoing project: Aug2020 – 2023
India	Agriculture Data Exchange (ADeX)	Collaborative venture to create India's first data exchange platform for farmer services.	Launched (11/8/23)
Belgium	DJustConnect	Stimulate data exchange in the Agrifood sector with respect for the different stakeholders.	Launched on 2020

(data on August 25th, 2023)

3.2.6 Reference on other sectors.

To give a complete understanding of how data can be valorized, an analysis of existing data exploitation in different sectors such as industry, medical, and transportation (aviation) was conducted. The analysis was conducted starting by characterizing the data collected and current conditions from different sectors, existing problems/challenges, the differences to the agriculture sector, and in the end, a good example from this sector on how to manage massive data was described as a reference for possible adaptation to the agriculture sector. Below are complete analyses and descriptions of data utilization in different sectors.

1. Industry

a). Context of the data:

Data produced by industrial practices are mostly generated by sensors, similar to the agriculture sector. This data can be a representation of the production capacity of an industry. However, the industry data are mostly used by internal stakeholders for product research and development, quality control, etc, and, this data is rarely shared with other companies. This occurred because these data are considered to have a competitive value, so companies are afraid to lose control of data if they share their data ("Designing Data Spaces," 2022).

b). Comparison with the agriculture sector:

With the advancement of embedded systems, software technologies, and connectivity, data from industry/manufacture are capable of being produced and analyzed in real time. In agriculture, real-time monitoring may be performed using integrated sensors, such as weather monitoring sensors, embedded cameras, etc. However, the necessity of applying real-time monitoring in crop production is not as high as in the industrial/manufacturing sector, where at an industrial level, each batch of production may be fully monitored and controlled. Also in agriculture, the level of uncertainty and variability is higher because of climate conditions, except in greenhouse or plant factories. Moreover, the evolution of agriculture practices mostly corresponds to industry domain. Industrial 4.0 influences Agriculture 4.0, where integration IoT and Big Data were first implemented in the industry sector. With the needs of traceability and sustainability in agrifood industries, it was influenced the data utilization (collect, process, analyse) in agriculture sector, and transform the agricultural practices into massive utilization of IoT and Big data (Liu et al., 2021).

c). Challenges:

The Industrial Revolution 4.0 is still ongoing and industrial production processes have become more autonomous because of the implementation of IoT, big data, AI, and blockchain technology (Liu et al., 2021). This digitalization leads to massive data production, however strong confidential issues between stakeholders are one of the constraints in terms of data sharing. Also, there is a need to harmonize data from cross-national policies and regulations.

d). Innovation:

To integrate wide adaptation of technologies, a three-dimensional model that describes a step-by-step migration to achieve Industrial 4.0 was presented in 2015, and this model is known as “Reference Architecture Model Industrie 4.0” (RAMI 4.0) (Contreras J et al., 2017). This model is based on standards and represents each role and functionality of each element in a factory. An adaptation of this kind of reference architecture may be beneficial in the agriculture sector to accelerate digitalization. Also, digital twins and Asset Administration Shell (AAS) are considered key elements in Industry 4.0 (Nakagawa E et al., 2021). A reference architecture of digital twins in agriculture is explained by Falcão et al., (2023)to address the challenge of digital interoperability in the agricultural data ecosystem.

2. Medical

a). Context of the data:

Medical data is the most sensitive data because it’s related directly to a person and if not managed well can cause serious loss and damage to someone. On the other hand, medical records are important to trace a person’s medical treatments and as a driver for innovation and collaboration. Today, the implementation and utilization of medical data are mostly managed by the government and in collaboration with private vendors.

b). Comparison with the agriculture sector:

The medical sector is considered important and the intervention from the government in the health sector is massive. Also, medical practices are mostly regulated strictly. The data are also considered more complex because of the enormous volume and variety. And the collecting data practices are heterogeneous, but still in standardized manners with common ontologies.

c). Challenges:

As we know, the population is growing and the advanced medical technologies and innovation, the data generated are also growing. The data can be produced by various entities such as medical facilities, drug stores, researchers, healthcare equipment, drug manufacturers, and many more. On the individual level, a challenge toward more sustainable and traceable healthcare arises to prevent unnecessary hospitalization. To address this challenge, shared healthcare information among professionals using Electronic Health Records (EHR) is considered as one of the solutions (Burnel, 2018). Reflecting on this phenomenon in the healthcare system, an adaptation of sharing farm data on soil or crop health is beneficial to give the best recommendations on different agriculture solutions.

d). Innovation:

The Electronic Health Records in France called système national d'informations inter-régimes de l'Assurance maladie (SNIIRAM) are covering 99% of the French population. Then, in 2017 new platform système national des données de santé (SNDS) was established to allow the recording of patient drug reimbursement data, date and paramedical interventions, laboratory test, hospitalization data, and personal data (birth date, gender, etc) (Scailteux et al., 2018). This documentation allows professionals to treat better the patient based on personal medication records. In agriculture, farm records are mostly defined by digital providers and there is a possibility of unlinked data. Good data consolidation in agriculture can be started from common standards and proper documentation that are being applied to all actors, and this could be done by regulation enforcement, just like in the healthcare system. Also, after the COVID-19 pandemic, the urgency of cross-border health data is raised. The European Data Strategy is willing to revolutionize health data by establishing a European health data space, where in this ecosystem actors will allow secure and confident data exchange for innovation in the medical sector (Nagel & Lycklama, 2020). In general, the collaboration between public and private stakeholders may bring a better impact.

3. Transportation (Aviation)

a). Context of the data:

Data in the aviation domain are used to improve service, flight management, aircraft maintenance, airline operations, and many more. With the emergence of big data technologies, the process of data collection, storage, and computation has become easier (Biliri et al., 2019).

b). Comparison with the agriculture sector:

The aviation sector is mostly dominated by already big actors because the cost of producing a commercial aircraft is high and the strict regulations, especially in terms of safety and security. Because of this, there isn't much variety in terms of aircraft brands, making the consolidation data easier. Also in the aviation domain, safety is the priority, and stakeholders have the same motivation to ensure passengers' safety. As a result, collaboration and partnership to create innovation to enhance aircraft safety are common practices between manufacturers.

c). Challenges:

At the business level, data produced by airlines is enormous, and the ability to store and analyze the data is considered one of the challenges (Al-Azab et al., 2021). Also in the ecological issues, an efficient flight is considered important to reduce carbon emissions. Efficient and punctual flight management and reducing unnecessary flights may be impactful in this matter.

d). Innovation:

A platform 'Flightradar24' (https://www.flightradar24.com/provides) an aircraft monitor where the public can track an aircraft and its route. This is a good example of how data openness is achievable. Just by connecting a tracking sensor and receiver a complete open tracking platform can be created. The only limitations are the coverage area by receivers and the compatible transponder on aircraft. Another example is a tool 'SkyWise' developed by Airbus to enhance safety by optimizing analysis tools and aircraft maintenance. This open-source tool is used by more than 100 airlines. Lessons learned from the aviation domain are the spirit of data utilization for greater goods, and t big actors can initiate innovation for public interest.

3.3 Study case: Terres Inovia decision tool

3.3.1 Introduction (context)

Rapeseed crops or *colza* are common crops in France. Farmers normally start to sow the rapeseed in August or September and rapeseed is considered a good starter for planting because it can fix nitrogen in the soil. However, if rapeseed grows too fast, it becomes more sensitive to cold and can result in stem frosting. This condition can lead to stem canker risk caused by *Leptosphaeria* fungi, and this condition can result in yield loss in the end. To avoid this condition, a fungicide/growth regulator can be applied when rapeseed is in between 5-6 leaf growth stage. Also, weather conditions have a great impact, where stem cankers are mostly found in warm and wet autumn (Brachaczek et al., 2021). TerresInovia developed a simple Decision Support Tool to give recommendations on whether or not to apply growth regulators with the level of risk. To use this tool, farmers just have to create or log in with their account. This tool is simply designed with only yes or no questions, and multiple answers questions. The idea is to provide effective questions for the input of algorithms, and it's considered a natural behavior of users to prefer answers short and easy rather than complex and long questions. In general, this tool estimates the risk of autumn elongation of rapeseed and indicates the benefit or not of applying a growth regulator, depending on the precocity of the six-leaf stage, the sensitivity of the variety to elongation, the density, available nitrogen, and rapeseed size. The architecture of the tool, interface, and the result of recommendation can be seen in Figure 10, 11, and 12 below.

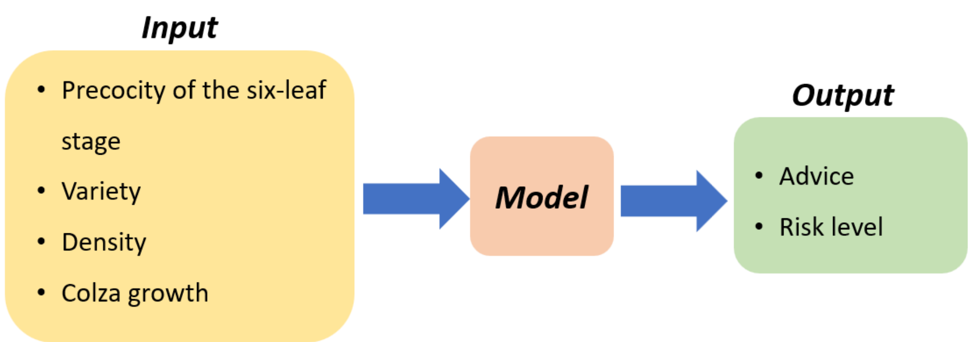


Figure 10. The architecture of TerresInovia decision support tool for autumn rapeseed elongation

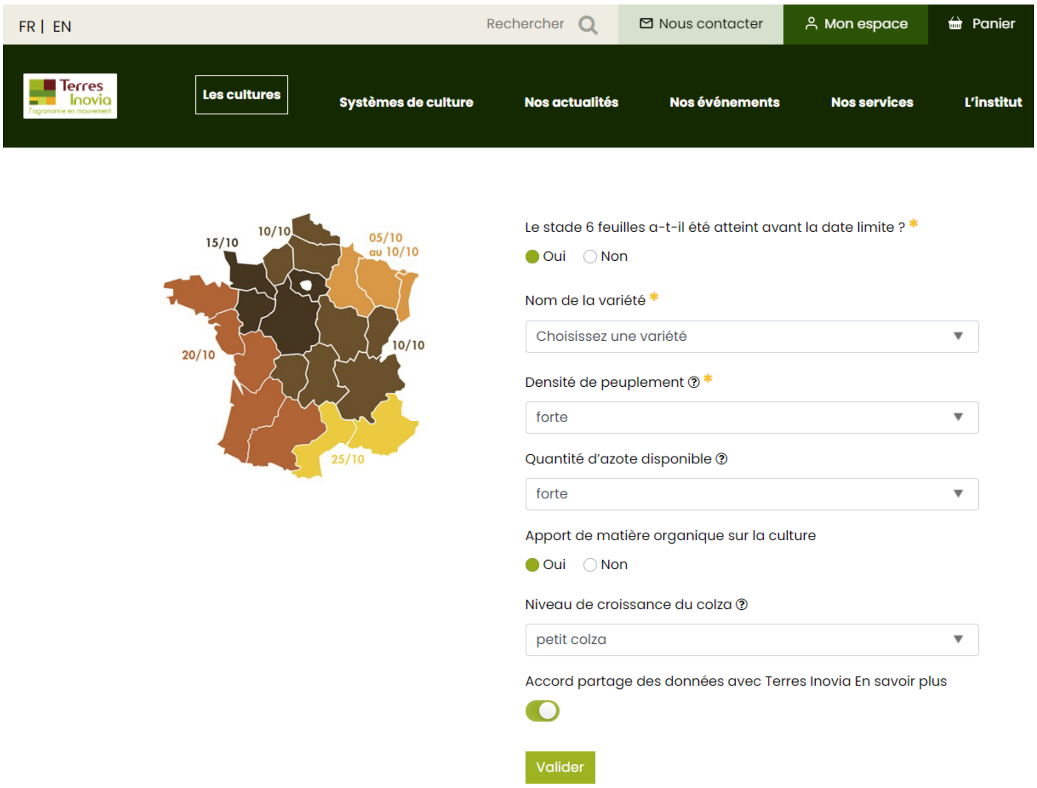


Figure 11. Interface of the tool's form

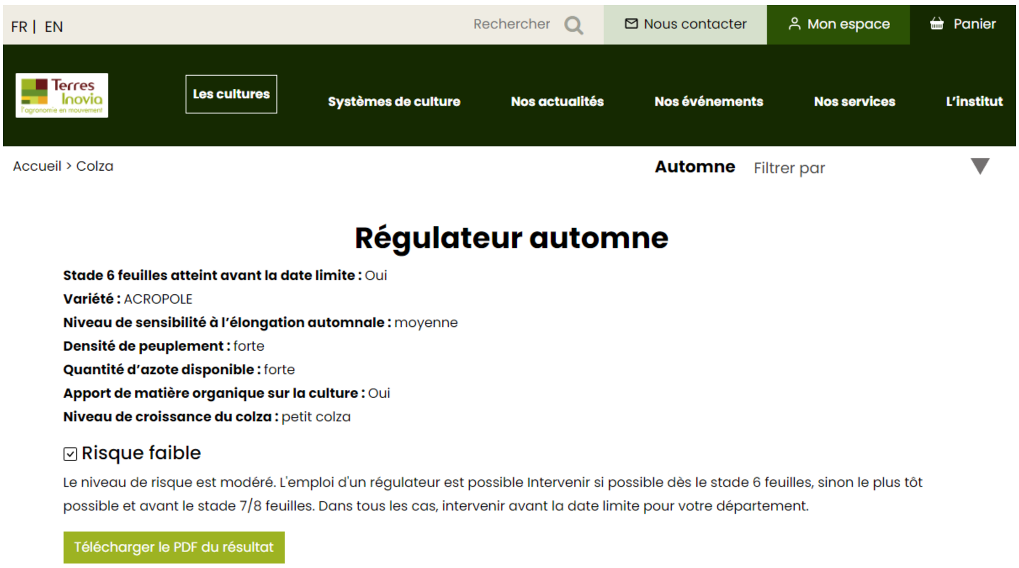


Figure 12. Result/recommendation from the tool

The process of data fusion in this study case was initialized by inspecting and analyzing the data collected from the tool. The objective of data inspection is to understand the data before practicing further analysis. The data analyzed for this study case has been anonymized, so the privacy of the users is secured. First, the data's size and dimension are measured. Then the data cleaning process was performed by removing null values and repetitive records (filled by the same person multiple times). After the data was cleaned, some descriptive statistic was performed to understand the distribution of the data. Below in figure 13 are the bar plot of top 10 varieties selected by users on the data and figure 14 the data record sample.

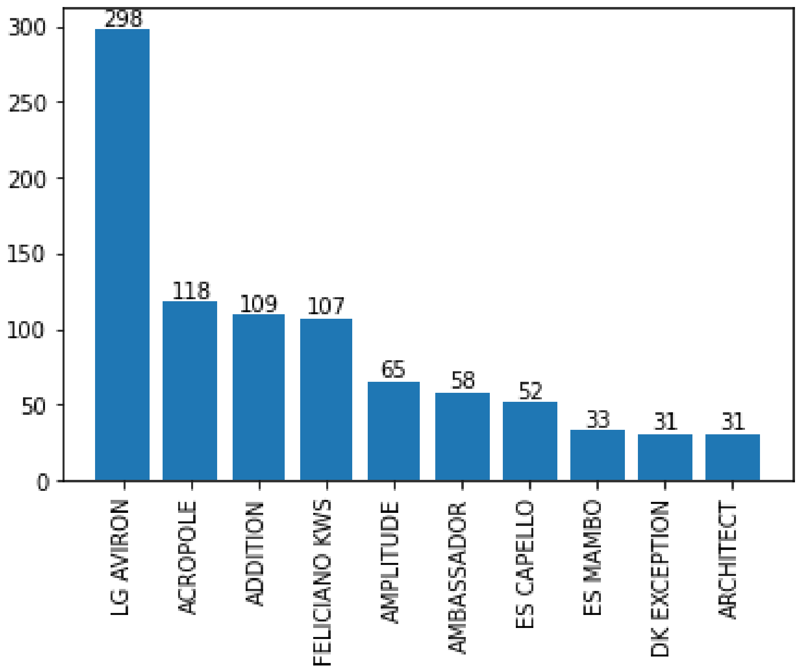


Figure 13. Distribution of top 10 varieties selected by users


```
{'adviceText': "Le niveau de risque est moyen.\nL'emploi d'un régulateur
est possible à conseillé.\n\nIntervenir si possible dès le stade 6 feuil
les, sinon le plus tôt possible et avant le stade 7/8 feuilles. \nDans t
ous les cas, intervenir avant la date limite pour votre département.",
  'apportCulture': 'false',
  'avantDateLimite': True,
  'companyId': 20099,
  'createDate': 1679590954367,
  'densitePeuplement': 'forte',
  'elongationAutomnale': 'moyenne',
  'gnisCode': '580E319',
  'groupId': 0,
  'isRGPD': False,
  'modifiedDate': 1679590954367,
  'niveauCroissance': 'gros colza',
  'nomVariete': 'ALEZZAN',
  'qteAzote': 'forte',
  'regulateurEntryId': 5137678,
  'riskLevel': 'moyen',
  'userId': 3877353,
  'userName': '1730c3dda5302b8881a558f1ac0465015eb6016f743ea51f65780b78a
887129b'}
```


Figure 14. Anonymized TerresInovia data sample

3.3.2 Feature modeling

The development of this tool was performed by analyzing the potential of data fusion in some possible aspects. The Idea was connecting relevant open-source data as an input to improve the decision support tool. As we know, the model was fed by simple input, however, there are some possible improvements to potentially optimize the results. Moreover, the improvements must be made by not changing the essential objective of the tool, and not adding more complexity to the users. Initially, there are 6 inputs in the tool:

- Limit date of 6 leaf-stage (*avantDateLimite*): Can be observed easily by farmers, and the determining method using binary question(yes/no) was considered as the most effective method. Identification of 6 leaf-stage using with high accuracy can be done using Terrestrial LiDAR Point Cloud in filed scale (Hu et al., 2022).
- Variety names (*nomVariete*): The process of rapeseed's variety auto-detection is complicated and there's no proven research for detecting enormous varieties. However, there's possible fusion with varieties database.
- Crop density (*densitePeuplement*): In the tool, the density is stated by high (forte) or low (faible), there's a possible fusion using satellite data to measure density.
- Nitrogen quantity (*qteAzote*): Needed a lab test or based from historical data of fertilizer application. For possible automation, it can be achieved by connection with others tool who own this information (FMIS).
- Organic's matter contribution (*apportCulture*): Binary question (yes/no) are considered the best method to address this matter.
- Crops growth level (*niveauCroissance*): Are stated by big rapeseed (gross colza) and small colza (petit colza). There's a possibility for automation using satellite images to identify better the growth level.

Besides, there is one important input that we considered important but not included for generating advice. This input is weather data. As explained, humidity can influence rapeseed's growth process, and risk of frost. Therefore, based on that consideration on the inputs, these are possible data fusion with open-source data:

a. *Connection to GNIS database for more valid varieties*

Based on the tool, users have to choose one variety on the question form. Different varieties may have different resistance to the elongation conditions, and it is also considered in generating results for advice and risk level. The varieties are listed and certified by Seeds and Plants inter-professional association (SEMAE) and the database can be accessed on SEMAE platform (https://www.semae.fr/catalogue-varietes/). This platform provides a catalog of databases from different crop varieties. Connecting the tool with the SEMAE database may give benefit by providing updated and valid existing varieties. The process of data fusion with this database can be made by linking the selected varieties (*gnisCode*) from the tool to SEMAE rapeseed varieties database. In this study case, matching the variety's names (*nomVariete)* and GNIS code (*gnisCode)* with SEMAE rapeseed database was conducted. First, a dictionary made from *nomVariete* and *gnisCode*. Then, this dictionary was compared with a dictionary made from SEMAE database with the assumption that these two dictionaries should be the same. In the end, these two dictionaries give the same result, and the querying process can be done by replacing *nomVariete* with SEMAE's varieties name based on the *gnisCode.* Below Figures 15 and 16 show the result from two different dictionaries.

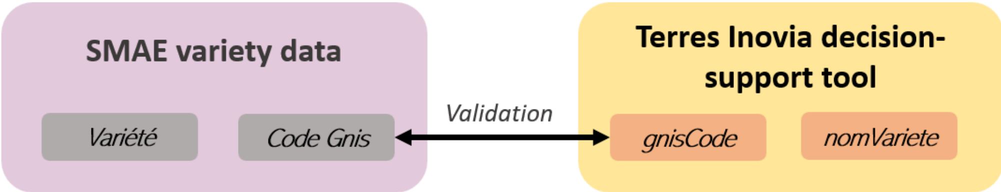
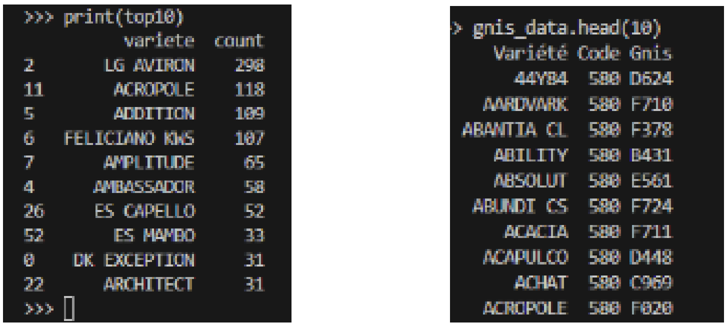


Figure 15. Integration schema between SEMAE data and Terres Inovia data



(a) (b)

Figure 16. (a) Query result from tool data and SEMAE data, (b) dictionary from SEMAE data

From the analysis of querying procedures, there is a point to be noticed. The process of pulling SEMAE database has been done by manual access, by downloading the CSV file of rapeseed varieties data through the SEMAE's website. This process is considered as one of constraints or limitations because the data are not provided by API and caused the integration process to be done manually. Therefore, the developers of tool should consult the database regularly for a variety of updates. With the API, integration process may be performed easily, and provide practical effort for an automation process.

b. Integration with geo location data

One of important aspect taken into account for determining recommendation and risk level from tool are the density of crop (*densityPeuplement).* In this tool, users can state the density by two levels: low (*faible)* and high (*forte)*. However, there are possible improvements for the tool and giving more specific information by measuring biomass density from satellite images. The calculations of biomass on the farm may give better accuracy of the density, rather than only stating the level of density by low or high. But, to get into that information, users have to input the plot boundaries information which is not provided in the tool. Users may draw their plot boundaries in the tool interface or upload their plot boundaries shapefile. Therefore, the important aspect of this tool was to provide simple and not complex questions to accommodate users, and those solutions considered may bring more complexity. So, another possible solution was foreseen to provide plot boundaries without adding more complexity to the tool. The idea was to querying existing data about plot boundaries from open source and masking it with satellite data. The data about registered farm can be retrieved on *Institut national de l'information géographique et forestière's* platform (https://geoservices.ign.fr/rpg) which provide anonymous data on plot boundaries in France. The feature has been developed so users can drop a pin inside their plot, and based on that information a query process was performed on RPG data to gain plot boundary. Then, satellite image was masked with that plot boundary to gain a relevant raster image of a farm. After that, a calculation of the vegetation index may be conducted to determine the level of rapeseed density. Figure 17 shows the schema of geolocation's integration.

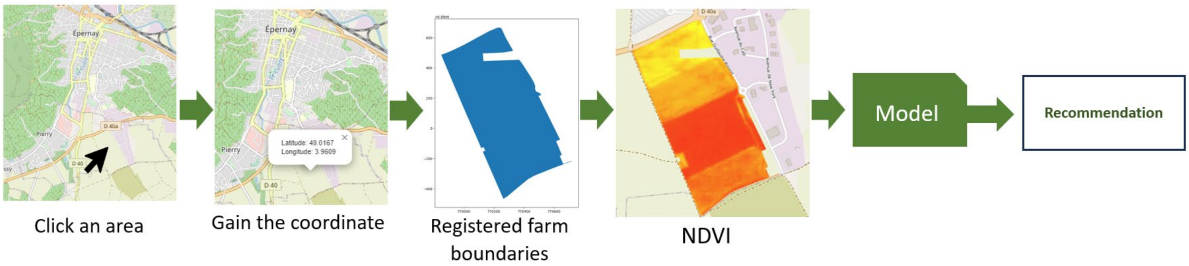


Figure 17. Geolocation integration schema

The first consideration for applying this feature was the size of RPG data. The size of RPG data is big, and the consequence is the querying process will take some time. So, to reduce the processing duration, this study case used the Grand-Est region RPG data. For the improvement, the developers may provide more location-specific questions so users can select only one region to avoid querying data on the whole France region. Second, the RPG data are updated annually. Even when this study case was carried out (September 2023), the latest version of RPG is still in 2021. Thus, there are possibility the farm border has changed, and the tool querying the outdated data. Lastly, the satellite image may have a cloud coverage problem even if we apply a cloud coverage filter. This occurred because the cloud coverage filter selects an image based on the total cloud percentage on one captured image, while our masked image is only a small part of the complete image. Also, the image resolution may be taken into account for generating better recommendations. In general, the issues encountered in this feature mostly concern data quality, and as data consumers the best we can do is pay more attention to possible bad results and minimize the error by providing more input detail and a better model.

c. Integration with weather data

As mentioned earlier, another factor that influences the elongation is weather conditions. Adding historical or forecast weather data to the model may give a better improvement of recommendations and risk levels. Weather data can be retrieved by providing information such as starting date and location. In the tool, the date information is represented in *createDate,* and there's no information about location. However, information about location by integrating geolocation data of farm boundaries may be implemented to generate weather data. The process of pulling weather data can be performed using API in Open-Meteo platform (open-meteo.com) with a combination of these parameters (*createDate* and farm coordinate). The only consideration for adding a weather feature is the limited requests per second on the free version. An upgrade to a subscription version is required to receive larger requests.Figure 18 shows the schema of weather data integration.

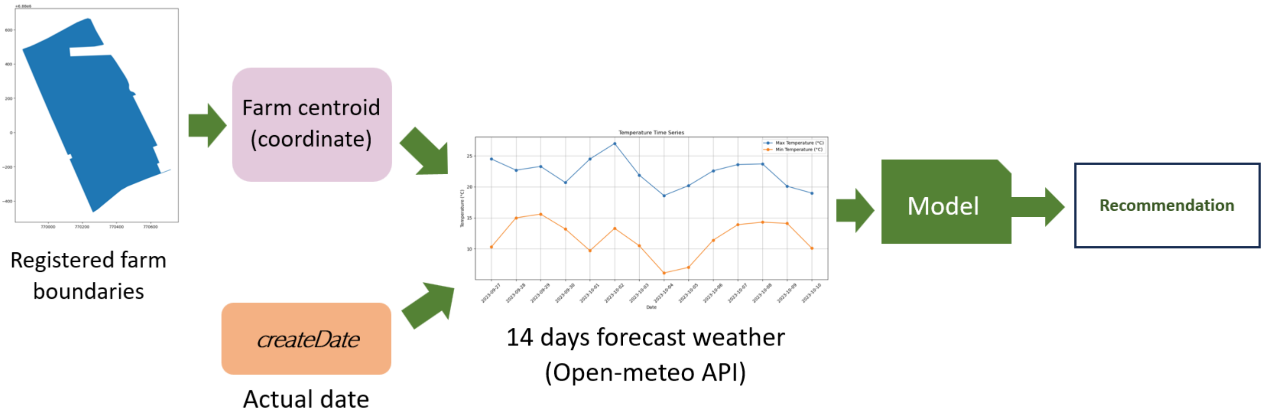


Figure 18. Weather data integration schema

3.3.3 Reference architecture

The developments described earlier may improve the tool in giving better advice. In total, there are two additional inputs (vegetation index and weather) may be given for the model. Validation with SEMAE data does not generate new feature but only provide a more valid and updated variety's name and code. Also ensure a better interoperability between several databases dealing with the same need of information. On the other hand, these additional features (vegetation index and weather) can be implemented by adding only one input to the tool. As an effort to generating FAIR data, the farmers crop's location are the key to precisely identify the crops, and become more robust when combined with date. This important feature is farm coordinate which does not yet exist on the tool. Moreover, this feature can be adjusted to the tool with a simple method where the user only needs to drop a pin inside their farm. This method is considered not to increase the complexity of the existing tool. Also, weather data can be generated, based on this location data. Full reference architecture from developments can be seen in Figure 19 below. Unfortunately, the development cannot be continued until model development. This occurred because the ground truth data are not observed so the model improvement cannot be performed. Analysis of ground truth data and these available variables should be performed to obtain the proper model. However, this study case showed that data fusion can be performed on a simple decision support tool by integrating relevant variables on open-source data. As a result, new variables can be added to improve the model and have the possibility to giving better recommendations. These variables may become a main or supplementary input. In the tool, the crop density and colza growth are measured by nominal scale (high, moderate, low). With the integration of location data and vegetation index, these inputs may be measured with a holistic model and may increase the accuracy. Moreover, the climate variable which is not taken into account by the current tool can be added to increase the precision of risk, based on climate conditions in selected locations.

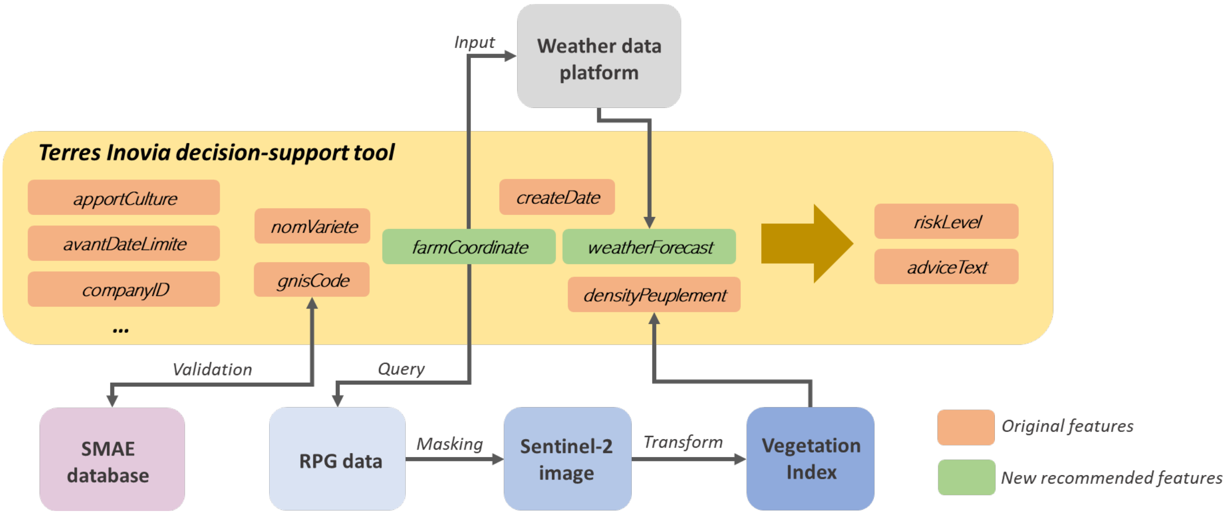


Figure 19. Reference architecture of possible development on Terres Inovia decision support tool

To conclude, process of automation and fusion in Terres Inovia tool are possibly done, and this process can improve decision models and agricultural practices. It's important to notice that the developments methods to interconnect these data are not the main problems. Based on the process of collection until integration, we faced the problems of interoperability, data quality, and accessibility. These problems can be countered by massive application of FAIR principle. Nevertheless, the result of the recommendation may become an input to another tool. For example, agronomist might be interested to use the data for development/assessment of a variety in certain conditions. Also, integration with the FMIS to scheduling a regulator application for farmers. Then, to connect the results from Terres Inovia tool to other tools, the implementation of FAIR principle plays an important role. Refers to David et al. (2020), these are parameters that should be considered:

a). Findable

For the identification process, a unique, global, and persistent identifier is needed. Moreover, the metadata should be persistent and linked to the authority (Terres Inovia) and the datasets itself. The metadata should describe the data format/type and respect to standards/dictionary. Then it should be searchable using relevant keywords. So the data consumer can find and locate the data/metadata easily

b). Accessible

Data can be stored in repositories, and efficiently accessible to various uses and users (using commong protocol: http). Also, some restrictions can be applied to different levels of data access for sensitive data. For example, the full data only accessible by the administrators, and only non-sensitive data can be accessible for public consumption.

c). Interoperability

The data/metadata should use standard vocabularies/thesaurus/onthologies/data dictionary. Provide several different formats such as CSV, JSON, XLSX, etc. Even provide an API versions to facilitate interoperability are better solutions.

d). Reusability

The metadata can be enriched by description of potential reuse actions and collecting procedures for traceability and sustainability. Also important to justify the legal resuse restriction.

From all the parameters explained, the missing link today is recommendation or agreement to use a specific ontologies/terminology in agriculture. By defining it, the process of interoperability might be easier. The solution is not to create new ontologies, but validate the existing one, and linking the data (Arnaud et al., 2020).

4. Discussion

4.1 Perspective from stakeholders

Based on the result of characterizing and mapping of current condition of agriculture data landscape, the research was intended to seek another insightful opinion from different actors. In total, there are 11 respondents who give their perspectives on this matter. They come from different backgrounds such as academic, business, data intermediary, non-lucrative organization, etc. Table 8 shows the list of respondents and their domains. Their

Table 8. List of respondents

No	Domain	Organization name
1	Academic	UniLaSalle, Télécom Paris
2	Innovation	Sofiproteol, Terres Inovia, AgreenTech Valley
3	Standarization body	AgroEDI Europe
4	FMIS	WIUZ
5	Data exchange enabler	AgDataHub, AgataConsent, OKP4
6	Farmers' federation	FNSEA

4.1.1 Opportunity and upcoming trends

From our findings in the interview process, we learn that innovation and digitalization trends in the agriculture sector will increase. Also, existing laws about data protection and exploitation will encourage a healthy environment to generate more knowledge from data. Apparently, advanced farming equipment has fully integrated with sensors and cloud technologies. Basically, the data are already being collected and stored. Hence, data are the fuel for innovation, and the data can be valuable when it's used. These data hold magnificent potential if circulated. Moreover, the existence of a trusted third party (AgDataHub) for data exchange scales up the ambition for releasing the true potential of data. Not only data intermediaries, a tool that relies on blockchain technology are also being developed (OKP4). With this tool, stakeholders may build their own data orchestration to generate knowledge without exchanging the data. The algorithm will directly process the data, and the data will not be transferred to another information system. In general, the existence of proper regulations and mature technologies will foster the data valorization process. But still, it will take years before it will be fully adopted.

4.1.2 Challenges

Currently, these stakeholders produce, collect and process data in different ways and motivations. Academic stakeholders often use open-source data or collect data directly from farmers and data from private partners for research purposes. Business-oriented stakeholders often produce data by direct collection to users for product/service innovation. Non-lucrative organizations do not process data on a regular basis, and farmers' federations only manage the database of their members. With this different motivation and method of collecting data, a common challenge is posed on how to get data with good quality. At this moment, data is not well circulated. They are mostly stored and consumed for internal use. An interest in having access to good quality data is rising as data-driven innovation emerges. On the other hand, big data for agriculture is underexplored compared to other research focuses such as food processing, innovation technology, market competitiveness, and sustainable agriculture (Suryaningrum A et al., 2023). The next challenge posed is to increase the fusion and integration of data generated by agriculture practices. This data fusion is not expected only in agriculture data, but also in other domains such as environment, economy, industry, and many more. The fusion is considered will foster innovation, leverage the value of a product, and optimize farming practices. Lastly, the challenge of encouraging good data governance and data sharing. Education, communication, and mediation with other stakeholders are essential to building supportive conditions for data sharing among actors.

4.1.3 Constraints and limitations

The first constraint that should be managed is the diverse and unharmonized data. As we know, data collected in the agriculture domain varies and is often unharmonized and it might cause by different sensors, collecting processes, data management methods, etc. The second barrier is the skepticism from stakeholders about sharing data. The data considered may have a competitive value that can differ a product/service from the other. Indeed, some stakeholders are still not confident that their data will not use against them by their competitor. So, data portability and traceability are the main keys to this constraint. At this moment, these features are still being developed by service providers. Lastly, the low literation and implementation of good data governance and stewardship are considered as a constraint. Then, cross border regulations can be a limitation, because of different local norms and interpretation. Also, the data size in agriculture can be considered as a limitation. Imagine if data from all farmers in France are stored in one platform, it will be enormous. The proposed method is having a decentralized data storing system, and then being connected by govern an access to the database.

4.2 Requirements and recommendations

Based on the analysis of the current agricultural data landscape, the effort of data valorization is still happening exclusively on the digital tools level, and lack of connectivity between each actor. As we know, the synergy of the data may lead to better agricultural practices such as reducing input on land, better irrigation management, animal welfare, and many more. But, the realization is remarkably complex, and intensive efforts are needed. Rozenstein et al. (2023) described that there are legal, technical, economic, and social barriers to data-driven systems. First, the legal aspects have affected the process from upstream to downstream. It defines the regulation on how data should be managed and who can have access to that. Apparently, the legal aspect of data in the European region is defined in GPDR, Data Governance Act, and Data Act. Even though the regulations already exist, mistrust and a big debate on data ownership still exist. The Data Act approached these issues by covering the right on data access and share data for users, the obligation of third parties and data holders, interoperability obligation, and other regulations in regard to data portability. However, this legal approach still does not fully cover ag-data issues, and more domain-specific are expected to come (Atik, 2022). In general, the efforts to break legal barriers have already begun in the EU region. To accelerate the transformation and reach wide application, education about existing regulations to all stakeholders is considered necessary. Second, technical barriers occurred because of incompetent users in utilizing/maintaining the tools, and the inexistence of relevant technologies. Data generated by sensors will influence the output in a data-driven system (Linaza et al., 2021) and uncalibrated sensors will mislead the knowledge-generating process. As we know, agriculture data are heterogeneous and often clustered because of the exclusivity of existing digital tools. Innovations to overcome these issues exist (AgDataHub and OKP4), but these digital solutions are still being developed to reach maturity and wide application. Besides, the true major issues today are data quality and enormous data volume in agriculture. Therefore, the application of FAIR principles can tackle these issues by improving the decision-making process and handling data volume growth issues at the same time (Ali & Dahlhaus, 2022). Lastly, economic and social barriers naturally coming from hesitation to use/implement data sharing and can be solved by demonstrating the benefits of information technology tools (Rozenstein et al., 2023). To be precise, provide a concrete implementation of existing technologies by developing a new use case or application on data valorization/integration with a data intermediary or blockchain solution to leverage the value of data and perform better agricultural practices. For example, a use case of generating environmental impact from agricultural practices with data-driven technology.

5. Conclusion

5.1 Conclusion

Digitalization and data innovation in the agriculture sector are showing positive trends, and in correspondence with proper underlying laws/regulations on how to treat, exchange, and valorize data, it will be an opportunity to optimize agricultural practices. Agriculture data are collected by various actors, methods and motivations. Then, the impact is agriculture data domain are diverse and unharmonized. Efforts to harmonize data can be made by closing the gap in each existing system, and the application of the FAIR principle is considered as a solution. FAIR principle should be applied as soon as data from farmers are being collected to unleash the potential of data. Also, open data ecosystem is still considered far from ideal. Immature technologies, slow adaptation of law, and unconfident actors are still the main contributors to this current condition. However, some innovations are being developed in terms of data exchange and data ecosystem. Lastly, the study case revealed that performed data fusion can give more variables to be taken into consideration when building a model and may leverage the value of the outputs. From interoperability, data quality, and accessibility are the problems encountered when conducting the study case.

5.2 Perspectives

For the continuance of the research, a strong suggestion to was proposed to develop a topic (in compliance with existing regulations) to solve big data problems in agriculture Therefore, to develop a better condition of data valorization, this topic should also encourage other actors to apply FAIR principles. This can be achieved by revealing the implication and benefit of applying the FAIR principles and specifications of good quality agriculture data for exchange or sharing. Then, collaboration with other actors (public or private) who work in this theme are also considered may accelerate the creation of better data valorization condition. The shortcoming of this research is that there was no direct opinion exchange with farmers, which has an important role as users, data sources, and data holders. Farmers perspectives were delivered by respondents that had an interaction/related with them.

Bibliography

ADS AGRICULTURAL DATA SPACE (ADS) A PUBLICATION OF THE FRAUNHOFER LIGHTHOUSE PROJECT "COGNITIVE AGRICULTURE." (2020.).

Al-Azab, M. R., Mohamed, H., Al, M., Abd, H., Samie, E., & Associate, M. (2021). *Big Data Analytics in Airlines: Opportunities and Challenges Big Data Analytics in Airlines: Opportunities and Challenges ARTICLE INFO ABSTRACT* (Vol. 21, Issue 4). https://www.researchgate.net/publication/356647231

Arnaud, E., Laporte, M. A., Kim, S., Aubert, C., Leonelli, S., Miro, B., Cooper, L., Jaiswal, P., Kruseman, G., Shrestha, R., Buttigieg, P. L., Mungall, C. J., Pietragalla, J., Agbona, A., Muliro, J., Detras, J., Hualla, V., Rathore, A., Das, R. R., … King, B. (2020). The Ontologies Community of Practice: A CGIAR Initiative for Big Data in Agrifood Systems. *Patterns*, *1*(7). https://doi.org/10.1016/j.patter.2020.100105

Beddar-Wiesing, S., & Bieshaar, M. (2020). *Multi-Sensor Data and Knowledge Fusion -- A Proposal for a Terminology Definition*. http://arxiv.org/abs/2001.04171

Biliri, E., Pertselakis, M., Phinikettos, M., Zacharias, M., Lampathaki, F., & Alexandrou, D. (2019). Designing a Trusted Data Brokerage Framework in the Aviation Domain. *IFIP Advances in Information and Communication Technology*, *568*, 234–241. https://doi.org/10.1007/978-3-030-28464-0_21

Birner, R., Daum, T., & Pray, C. (2021). Who drives the digital revolution in agriculture? A review of supply-side trends, players and challenges. *Applied Economic Perspectives and Policy*, *43*(4), 1260–1285. https://doi.org/10.1002/aepp.13145

Brachaczek, A., Kaczmarek, J., & Jedryczka, M. (2021). Warm and wet autumns favour yield losses of oilseed rape caused by phoma stem canker. *Agronomy*, *11*(6). https://doi.org/10.3390/agronomy11061171

Burnel, P. (2018). The introduction of electronic medical records in France: More progress during the second attempt. *Health Policy*, *122*(9), 937–940. https://doi.org/10.1016/j.healthpol.2018.07.022

Cattaneo, G., Micheletti, G., Glennon, M., & Croce, C. La. (2020.). *THE EUROPEAN DATA MARKET MONITORING TOOL*.

Consolidated text: Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) (Text with EEA relevance). (2016.). https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A02016R0679-20160504&qid=1696926479090ELI:http://data.europa.eu/eli/reg/2016/679/2016-05-04

Craker, B., Danford, D. D., Ferreyra, R. A., Nelson, K., Rhea, S. T., Stelford, M. W., & Wilson, J. A. (2018). *ADAPT: A Rosetta Stone for Agricultural Data*. www.aggateway.org

Cravero, A., Pardo, S., Galeas, P., López Fenner, J., & Caniupán, M. (2022). Data Type and Data Sources for Agricultural Big Data and Machine Learning. In *Sustainability (Switzerland)* (Vol. 14, Issue 23). MDPI. https://doi.org/10.3390/su142316131

D5.3 Testbed Deployment, System Extensions and Applications for Pilot Round 1 v1.6. (2020.).

Daum, T., Ravichandran, T., Kariuki, J., Chagunda, M., & Birner, R. (2022). Connected cows and cyber chickens? Stocktaking and case studies of digital livestock tools in Kenya and India. *Agricultural Systems*, *196*. https://doi.org/10.1016/j.agsy.2021.103353

David, R., Mabile, L., Specht, A., Stryeck, S., Thomsen, M., Yahia, M., Jonquet, C., Dollé, L., Jacob, D., Bailo, D., & FAIRness, al. (2020). Literacy: The Achilles' Heel of Applying FAIR Principles. *CODATA Data Science Journal*, *19*(32), 1–11. https://doi.org/10.5334/dsj-2020-032ï

Denitza, T. (2017). The Rule of Law and EU Data Protection Legislation. *The ORBIT Journal*, *1*(1), 1–16. https://doi.org/10.29297/orbit.v1i1.16

Designing Data Spaces. (2022). In *Designing Data Spaces*. Springer International Publishing. https://doi.org/10.1007/978-3-030-93975-5

Falcão, R., Matar, R., Rauch, B., Elberzhager, F., & Koch, M. (2023). A Reference Architecture for Enabling Interoperability and Data Sovereignty in the Agricultural Data Space. *Information (Switzerland)*, *14*(3). https://doi.org/10.3390/info14030197

Fountas, S., Carli, G., Sørensen, C. G., Tsiropoulos, Z., Cavalaris, C., Vatsanidou, A., Liakos, B., Canavari, M., Wiebensohn, J., & Tisserye, B. (2015). Farm management information systems: Current situation and future perspectives. *Computers and Electronics in Agriculture*, *115*, 40–50. https://doi.org/10.1016/j.compag.2015.05.011

Gaia-X_DSBC_PositionPaper. (2021).

Giray, G., & Catal, C. (2021). Design of a data management reference architecture for sustainable agriculture. *Sustainability (Switzerland)*, *13*(13). https://doi.org/10.3390/su13137309

Hu, F., Lin, C., Peng, J., Wang, J., & Zhai, R. (2022). Rapeseed Leaf Estimation Methods at Field Scale by Using Terrestrial LiDAR Point Cloud. *Agronomy*, *12*(10). https://doi.org/10.3390/agronomy12102409

J*, M., & M, I. (2019). Role of Big Data in Agriculture. *International Journal of Innovative Technology and Exploring Engineering*, *9*(2), 3811–3821. https://doi.org/10.35940/ijitee.A5346.129219

Janssen, S. J. C., Porter, C. H., Moore, A. D., Athanasiadis, I. N., Foster, I., Jones, J. W., & Antle, J. M. (2017). Towards a new generation of agricultural system data, models and knowledge products: Information and communication technology. *Agricultural Systems*, *155*, 200–212. https://doi.org/10.1016/j.agsy.2016.09.017

Liu, Y., Ma, X., Shu, L., Hancke, G. P., & Abu-Mahfouz, A. M. (2021). From Industry 4.0 to Agriculture 4.0: Current Status, Enabling Technologies, and Research Challenges. *IEEE Transactions on Industrial Informatics*, *17*(6), 4322–4334. https://doi.org/10.1109/TII.2020.3003910

Mcfadden, J. (2022). *THE DIGITALISATION OF AGRICULTURE A LITERATURE REVIEW AND EMERGING POLICY ISSUES The Digitalisation of Agriculture: A Literature Review and Emerging Policy Issues*.

Mildebrath. (2023). *EU policies-Insights Understanding EU data protection policy SUMMARY*.

Nagel, L., & Lycklama, D. (2020). *OPEN DEI Position Paper Design Principles for Data Spaces*. https://doi.org/10.5281/zenodo.5244997

Osinga, S. A., Paudel, D., Mouzakitis, S. A., & Athanasiadis, I. N. (2022). Big data in agriculture: Between opportunity and solution. *Agricultural Systems*, *195*. https://doi.org/10.1016/j.agsy.2021.103298

Otto, B. (2015). Quality and Value of the Data Resource in Large Enterprises. *Information Systems Management*, *32*(3), 234–251. https://doi.org/10.1080/10580530.2015.1044344

Porter, M. E., & Heppelmann, J. E. (2014). *HBR.ORG SPOTLIGHT ON MANAGING THE INTERNET OF THINGS How Smart, Connected Products Are Transforming Competition*.

Proposal for a REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL on European data governance (Data Governance Act) COM/2020/767 final Languages and formats available. (2020).

Proposal for a REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL on harmonised rules on fair access to and use of data (Data Act) COM/2022/68 final Languages and formats available. (2022). https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=COM%3A2022%3A68%3AFIN&print=true

Ruohonen, J., & Mickelsson, S. (2023). Reflections on the Data Governance Act. *Digital Society*, *2*(1). https://doi.org/10.1007/s44206-023-00041-7

Scailteux, L.-M., Droitcourt, C., Balusson, F., Nowak, E., Kerbrat, S., Dupuy, A., Drezen, E., Happe, A., & Oger, E. (2018). *Pharmacoepidemiology French administrative health care database (SNDS): the value of its enrichment Enriching SNDS*.

Schweihoff, J., Jussen, I., & Möller, F. (2023). *Trust me, I'm an Intermediary! Exploring Data Intermediation Services*. https://www.researchgate.net/publication/371990357

Snyder, H. (2019). Literature review as a research methodology: An overview and guidelines. *Journal of Business Research*, *104*, 333–339. https://doi.org/10.1016/j.jbusres.2019.07.039

Sunding, David, and David Zilberman. 2007."The Agricultural Innovation Process: Research and Technology Adoption in a Changing Agricultural Sector." In Handbook of Agricultural Economics, Volume 1A, edited by Bruce L. Gardner and Gordon C. Rausser, 207–61.New York, NY: North-Holland.

Top, J., Janssen, S., Boogaard, H., Knapen, R., & Şimşek-Şenel, G. (2022). Cultivating FAIR principles for agri-food data. *Computers and Electronics in Agriculture*, *196*. https://doi.org/10.1016/j.compag.2022.106909

USDA. (2023). *REQUEST FOR APPLICATIONS Open Data Framework*. https://doi.org/10.233

Wilkinson, M. D., Dumontier, M., Aalbersberg, Ij. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J. W., da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., … Mons, B. (2016). Comment: The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, *3*. https://doi.org/10.1038/sdata.2016.18

Appendices

Respondent's list

No	Name	Organization	Position	Alliance H@rvest
1	Nicolas MANIEZ	AgreenTech Valley	RnD Innovation Project Manager	Yes
2	Jérôme DANTAN	UniLaSalle	Professor and Researcher (IoT and Big data)	Yes
3	Isabelle GATTIN	UniLaSalle	Professor and Researcher (Soil science)	Yes
4	Baptiste FAINEANT	Sofiproteol	Project Manager, Innovation Fund and Sustainable Sectors	Yes
5	Laurent DECREUSEFOND	Télécom Paris	Professor and Researcher (mathematics & statistics)	Yes
6	Nicolas VANNETZEL	AgDataHub	Responsible Support Client	No
7	Marie BEURET	AgroEDI Europe	Executive Officer	No
8	Lucas PREZELIN	OKP4	Business Development Manager	No
9	Héloise GILLES	FAST-Agri (AgataConsent)	Directrice	No
10	David TALEC	WIUZ	Responsible Partenaire	No
11	Melchior BIZOT-ESPIARD	FNSEA	Chargé de mission innovation et prospective	No

Questionnaire

- Q1. What types of data does your company/organization collect and analyze? And how to collect them? Do you collect/give data from/to external sources? (buying/selling/open source)
- Q2. What are of the major challenges or constraints you face right now in terms of data valorization?
- Q3. Do you foresee in the future any trends or developments in this subject? and how do you plan to adapt into that kind of situation? What is missing today?
- Q4. Are there any specific regulations or legal considerations that impact the way you handle your data?

- Q5. What are the important aspects to manage if we want to apply the open data or data exchange with other organizations?
- Q6. What is your opinion about open data movement where companies share the data together?
- Q7. Do you have an example of data fusion you would like to test?
- Q8. How can VDPA+ project help you?

List of platforms

Type	Platform	Author	Functionality	Input data	Output (meta) data	Objective
Farm equipment system	MyJhonDeere	Jhon Deere	· Fleet management	· GPS	· Working plan	Help farmers organize their machinery work, and maintenance scheduling.
			· Field data recording	· Fuel consumption	· Machine report	
				· Working duration	· Location history	
				· Machine performance		
	MyPLM Connect	NewHolland	· Field data recording	· Agronomic data	· Working plan	Monitor the productivity and performance, cost efficiency, data analysis, farm management
			· Data sharing	· Machine data	· Machine report	
				· Personal information		
	AFS Connect	CNH	Field data recording	GPS data	Working plan	Manage users machinaries resource and records field agronomic data.
			Fleet management	Agronomic data	Machines' location	
				Machines performances data		
	CLAAS Connect	CLAAS	Field data recording	GPS data	Working plan	Manage users machinaries resource and records field agronomic data.
			Fleet management	Agronomic data	Machines' location	

			Spare parts shop	Machines performances data		
Farm Management Information System; field and herd operation management	Doctofarm	ISAGRI	· Transfer information	· Text message	· Farm portfolio	Allows technicians and farmers to work remotely through a chat, or to create georeferenced and shared observations
			· Consultation	· Image		
	MesParcelles	Chambre d'Agriculture	· Field activity record	· Fertilization data	· Fertilization and phytosanitary scheduling	Protect from disease and record fertilizer and phytosanitary use
				· Phytosanitary data		
				· Satellite images		

	Field manager	Xarvio	Field monitor	Satellite images	Biomass map (every 3 day)	Help farmers to monitor their fields and recommend better crop management to increase effeciency and productivity
			Seeding	Yield map	Weather data	
			Nutrition	soil map	irrigation recommendation	
				weather data		
	FieldView	BAYER	Data connectivity in real-time, record historical data, sharing data	Machines data	Field health analysis	Providing a connected tool to visualize machines' works to a cloud platform
				Satellite images	Variable rates recommendations	Giving recommendation about variables rate application
					Repository	
	MyEasyFarm	MyEasyFarm	Integration and visualization of different FMIS	Data from other FMIS	Data visualization from different FMIS	Consolidate different digital solutions into a platform.
				Machines data		
				Satellite images		

				Weather data		
	WIUZ	WIUZ	Planting to harvesting management	Farm data boundaries	Operations management	Recommending farms operation with actual resources
			Online consultation			
	Agreo Vigne	SMAG	Manage vineyards operations	Vineyard's report (on-land audit by agronomist)	Vineyard map	Helps user manage their resources and increase products' values by tracebility
			Financial/economic management	Working records (employee, working hours, etc)	Resource monitor and management	
			Product traceability (certification HVE)			
	SMAG Farmer	SMAG	Manage operations from planting to harvest	Machines data	Machines performance	Providing and consolidating informations for farmers, and giving recommendations.
				weather data	Crops growth maps	
				Satellite images	Weather prediction	
				Plot data	Recommendations	

				Production cost		
	Cattle management	SMAG	Manage and monitor cattle milk and slaughter operations	Machines data	Data visualization from machines	Manage herds operations.
					Repository	
Farm Management Information System; Best practice	BeAPI	BeAPI	· Soil mapping	· Soil conductivity	· Soil heterogeneity map	Optimizing fertilizer appliance based on soil heterogeneity data.
				· Soil fertility	· Fertilizing recommendation	
				· Yield data	Soil potential map	
	GeoFolia	ISAGRI	· Providing several tools for farm optimization	· Satellite image	· Fertilizing recommendation	Help farmers to making decision
				· Yield data	· Risk alert	
				· Fungicide data		
	Farmstar	Farmstar	· Fertilizer recommendation	· Satellite image	· Recommendation	Optimizing yield, reduce cost
				· Chlorophyll and biomass measurement		

	Weenat	SmartAg	With their module, records and forecast weather on specific location	Weather sensor data (windspeed, rain, temperature,	Recommendation based on the weather monitoring and forecast	Help farmer to decide the best time to apply fertilizer and prepare from natural events
	Movida	BAYER	· Fungicide application management	· Weather data	· Spraying plan	Optimizing the fungicide appliances to vineyard
	TankMix	YARA	Customization of fertilizer, phytosanitary, and fungicide products.	Users inputs	Recommendation product	Help farmers to find the right product and give recommendations based on tested mix.
			Product recommendation	Laboratory test	Mixing results	
	Avizio	Syngenta	Recommendation for wheat's fungicide application	plot data (sow, soil type, variety, cultivation method)	Recommendation map	Predict the risk of yield loss, and recommend fungicide application on wheat.
	Sencrop	Sencrop	Weather monitoring	Weather sensor data (windspeed, rain, temperature,	Recommendations based on the weather monitoring and forecast	Help farmer to decide the best time to apply fertilizer and prepare from natural events
Agriculture digital marketplace/sales	Isanet marche	ISAGRI	· Market portal for farmer and distributerur	· Product offer	· Price graphs	Product selling with secure transaction. Help farmers to decide which crop to be planted based on market stock.

					· Stock graphs	
	Frais et local	Chambre d'Agriculture, ministère de l'Agriculture et de l'Alimentation	Giving maps of local shops	Shops database	Maps of shops and their products	Facilitate both consumer to find nearest local product shops and also help shop owners to increase selling.
	aladin.farm	Invivo	Market place for agriculture products and services	Products/services data	Products/services detailed information	Facilitate distributors to sell their products/services and also facilitate costumers to buy a products
	Agri Maker	FAST Agritech	Information portal for farmers	Database of digital tools	Catalog of digital tools, news and updates	Providing information about updates, information, and existing tools to help farmers optimize their farming practices.
Decision support tools platform	Terres Inovia	Terres Inovia	Informations and recommendations for sunflower, rapeseed, soja	Users inputs	Recommendations	Helps farmers by providing information on each specific problems (giving insight for decision-making)
			Weed, insects, and disease information		Informations	

	ARVALIS	ARVALIS	Information and recommendations for various crops (plant physiology, disease, seeding, etc)	Users inputs	Recommendations	Helps farmers by providing information on each specific problems (giving insight for decision-making)
					Informations	
	B-GNR (Carburant)	Chambre d'Agriculture	· Carburant analysis	· Fuel consumption	· Recommendation	Identify priority paths to save fuel on each farm. B-GNR is particularly suitable for livestock farms.
				· Tillage		
				· Meadows management		
Data exchange and consent management tool	AgDataHub	AgDataHub	Data exchange intermediary	data offers from data providers	Data offers metadata	Facilitate secure and trusted exchange between data providers and data consumers
	Agata Consent	FAST Agritech	Consent management			Provide consent management between users and data consumers
Management and administrative	LEA	SMAG	Contractors' operation management	Orders' receipts	Oders' constracts docimentation	Facilitate management with equipments' contractors
	Replace.Me		Workers management			Manage workers by finding replacements.

	DematAgri		Archive and documentation	Digital document		Provide secure digital document management tool
Cloud infrastructure services	Microsoft Azure for Agriculture Data Manager	Microsoft	· Ingest, store & manage farm data	· Sensor's data	· Models	Drive innovation through insight, reduce environmental impact, optimize agriculture operations
			· Build & run models	· Satellite image	· Insight	
				· Weather data		
	AWS Solution	Amazon	· Customize a cloud computing architecture	· Sensor's data	· Models	Providing customizable architecture for agricultural needs with cloud computing technology.
				· Satellite image	· Insight	
				· Weather data	· Architecture	
	OKP4		Customize a cloud computing architecture with blockchain technology			Providing customizable architecture for knowledge/value creation.

List of open databases

No	Author/ platform name	Type of Data	address	Format	Meta data	API	Accessibility	Frequency	Context
1	Open-Meteo	String, float, Boolean	https://open-meteo.com/	JSON, CSV, XLS	Yes	Yes	Everyone can access; 10.000 request per day (max) for free and non-commercial use	Data updated every 1-6 hours, depends on the data provider	Open-Meteo utilize open-data weather forecasts provided by national weather services.
2	FAOSTAT	String, integer, double	https://www.fao.org/faostat/en/#data/	CSV, XLS	Yes	No	Everyone can access; No limitation	Annually	The quality of data varies considerably between countries depending upon their individual data collection methodologies.
3	Copernicus SciHub	Image	https://scihub.copernicus.eu/dhus/search?q=*	Jp2	Yes	Yes	Require account creation; Very big size data (up to 20GB per file)	Updated every 2-3 days	The quality of data varies considerably depending on cloud cover condition
4	GéoServices (RPG)	Shapefile	https://wxs.ign.fr/agriculture/geoportail/wmts?SERVICE=WMTS&VERSION=1.0.0&REQUEST=GetCapabilities	shp	Yes	No	Everyone can access; no limitation	Annually	These data have been produced by the Services and Payment Agency (ASP) since 2007.
5	European Commission	All format available, depends on the data	https://data.europa.eu/en	Depends on the data	Yes	No	Everyone can access; No limitation	Not specified, depends on publisher	The data were published by EU member organization. The quality of data is various on each country member.
6	Ministère de la Transitio	String, integer, double	https://www.data.gouv.fr/fr/datasets/donnee-secheresse-propluvia/	CSV	Yes	No	Everyone can access; No limitation	Updated daily	The data publish a level of drought periodically of France area.

	n écologique								
7	FranceAgriMer	String, text	https://visionet.franceagrimer.fr/Pages/accueil.aspx	xlsx, PDF	Yes,	No	Everyone can access; No limitation	Annually, monthly	Providing data from various category (crop production, livestock, price, land use, etc). Some datasets are not completely filled due to confidentiality.
8	AGRIVOC (FAO)	String, text	AGROVOC: AGROVOC Multilingual Thesaurus (fao.org)			no	Everyone can access; No limitation		Semantics vocabularies for different name of species
9	GéoPortail	Image	https://www.geoportail.gouv.fr/donnees/carte-des-sols	pdf	Yes	No	Everyone can access; account creation needed for export a file	Varies (based on map type)	Visualize different type of maps.
10	Research Data Gouv	dataset	https://entrepot.recherche.data.gouv.fr/	Varies, based on data	Yes	No	Everyone can access the metadata; account creation needed for export data	Varies, based on data	Providing research dataset and publish a data. The data are variated since they are collected with different method.
11	World soil dataset (FAO)	dataset	https://www.fao.org/soils-portal/soil-survey/soil-maps-and-databases/harmonized-world-soil-database-v12/ru/		Yes	No	Everyone can access; require software installation to open the dataset		Dataset about soil type in all country with scale 1 : 5.000.000
12	Climate reanalysis data	dataset	https://cds.climate.copernicus.eu/cdsapp#!/search?type=dataset	Zip, tar, gz	Yes	Yes	Everyone can access; No limitation	annually	Data about climate in past, present and future (prediction).

13	SEMAE (plant variety database)	dataset	https://www.semae.fr/catalogue-varietes/	CSV	Yes	No	Everyone can access; No limitation	annually	
14	ADEME	string	https://data.ademe.fr/datasets?topics=TQJGtxm2_	CSV, JSON, XLSX, ODS	Yes	Yes	Everyone can access; No limitation	annually	Data about ecological and environment; example: nitrogen application on soil, electricity use, housing, etc
15	E-Phy	dataset	https://ephy.anses.fr/	CSV, XML			Everyone can access; No limitation		Phytopharmacies data catalog
16	THEIA	dataset	https://catalogue.theia-land.fr/		Yes,	No	Everyone can access; account creation needed for download a file		Geographic product
17	UPOV	dataset	GENIE Database (upov.int)			No	Evryone can access the limited free data; 750CHF annual subcription needed to have full access		Plant variety data and license
18	(SIG_GR) Ecological status/potential of surface water bodies	dataset	https://data.public.lu/fr/datasets/ecological-status-potential-of-surface-water-bodies/	geojson, shp	Yes	Yes	Everyone can access; No limitation		Ecological status/potential of surface water body

19	DataGrandEst	dataset catalog	https://www.datagrandest.fr/data4citizen/portail?facet.field=%5B%22organization%22,%22tags%22,%22themes%22,%22features%22%5D&rows=12&start=0#	csv, json, XLSX, geojson, kml, WMS	Yes	Yes	Everyone can access; No limitation		Provide dataset catalog to help users find available data in Grand-Est region
20	Open data - Département du Calvados	dataset catalog	https://open.isogeo.com/s/434296189d794266a04a8678e256f0c4/HiLeps3y1qrSx6aTDmKyp0spGPix0		Yes	No	Everyone can access; No limitation		Provide dataset catalog to help users find available data in Calvados region
21	Data Economi Gouv	dataset catalog	https://data.economie.gouv.fr/explore/?sort=modified		Yes	Yes	Everyone can access; No limitation		
22	Alim'Confiance	dataset	https://www.data.gouv.fr/fr/datasets/resultats-des-controles-officiels-sanitaires-dispositif-dinformation-alimconfiance/	csv	Yes	Yes	Everyone can access; No limitation		Data about sanitary control
23	UMR 1069 SAS INRAE - L'institut Agro	dataset catalog	https://www.data.gouv.fr/fr/organizations/umr-1069-sas-inrae-linstitut-agro/#/datasets						
24	ADES-Water quality	dataset catalog	https://ades.eaufrance.fr/Recherche	PDF, txt	Yes	no	Everyone can access; No limitation		Data about water quality in France
25	Pesticides on undergro	dataset	https://www.data.gouv.fr/fr/datasets/pesticides-dans-les-eaux-souterraines/	txt, csv	Yes	No	Everyone can access; No limitation		Pesticide content in underground water

	und water								
26	L'Indicateur de Fréquence de Traitements phytosanitaires (IFT)	dataset	https://www.data.gouv.fr/fr/datasets/doses-de-reference-indicateur-de-frequence-de-traitements-phytosanitaires/#/resources	csv, pdf	Yes	Yes	Everyone can access; No limitation	weekly	
27	EUPVP - Common Catalogue Information System	dataset catalog	https://ec.europa.eu/food/plant-variety-portal/index.xhtml;jsessionid=c6dwDaaS7gi1hq4b3Gk5_tBN1JgBVgrpTuTn_JhflrOoyEHYS2fw!1235683713	xlsx	Yes	No	Everyone can access; No limitation		EU common catalog for plant variety