

Project VDPA+: Mieux Valoriser les données des pratiques agricoles

“A project that aims to leverage the values of the agriculture’s data”

Dafa YUMNA, Estelle MILLET, Jean-Eudes HOLLEBECQ

1. 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 (Findable, Accessible, Interoperable, Reusable) 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

2. Introduction

Alliance H@rvest’s goals

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. There’s also, the academic concern by organizing conferences and courses. Moreover, project VDPA+ is a part of data and artificial intelligent topic. Alliance H@rvest organized and funded Project VDPA+ to respond to the actual condition where agriculture data are considered not well-valuated caused of the data consolidating reference model absence.

Context: Under valorized agriculture data.

1. Status quo of data consolidation model.
2. Challenges of cybersecurity and data proprietary.
3. High expectations for supplying the decision support tool, traceability, etc.

Objectives

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.

3. Methodology

The project was divided into three phases:

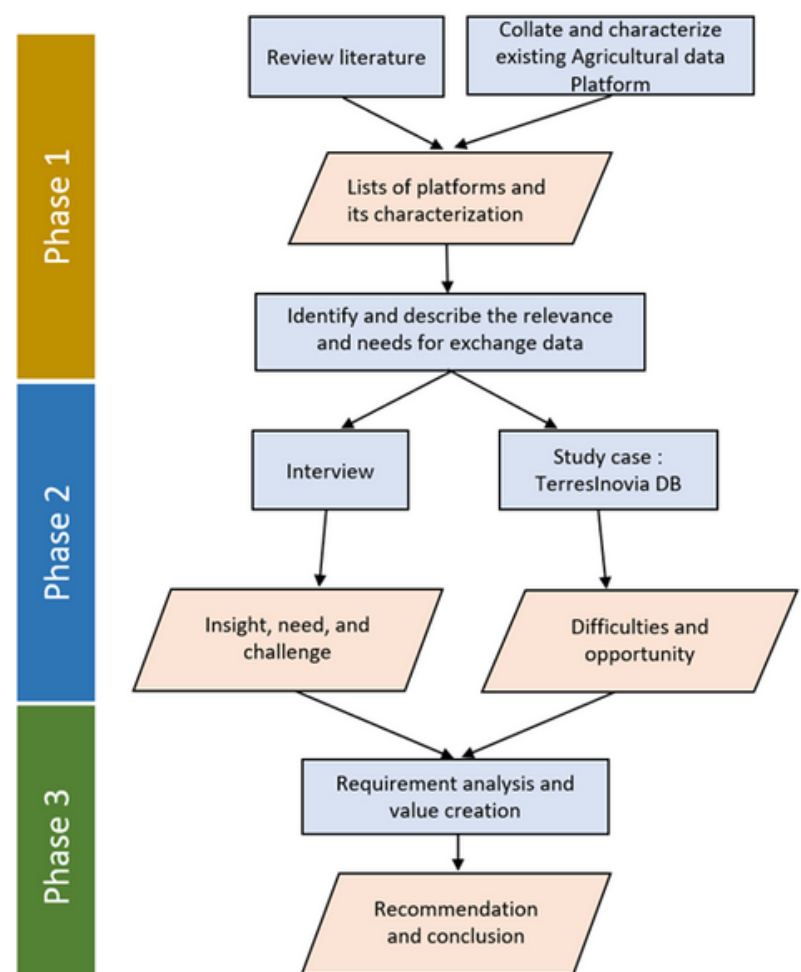
Phase 1: characterize and create a map of existing agriculture digital platforms

Phase 2: to determine needs and creation of added value

Phase 3: was to create a possible schema of data valorization.

The research was designed to answer the relevant questions with 3 different holistic approaches:

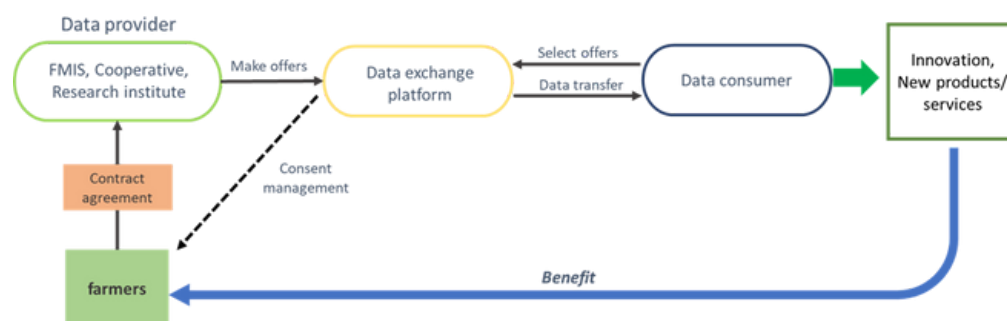
1. **Review** (publication, web, database) to understand the state of arts data utilization and valorization in agriculture domain.
2. **Interview** with relevant actors to Identify actual challenges and constraints
3. **Data fusion study case on Terres Inovia decision tool** to analyze and give a viewpoint of the applicable level of data fusion from actual data collected by public applications and open-source data



4. Result

4.1 Agriculture data characterization and mapping

1. **Data in agriculture domain are diverse and heterogeneous**-> different actors, motivations, collecting methods, and technologies.
2. **Data are not diffused/shared**-> mostly collected, stored and consumed exclusively by services providers.
3. **Current efforts to leverage data sharing**-> data-related regulations (GDPR, Data governance Act, Data Act), data intermediary and data consent management (AgDataHub, AgataConsent), Web3 technologies (OKP4).
4. **Solutions to overcome big data problem in agriculture** -> FAIR principles application, encourage the actors by providing concrete use case and describe the benefits.



Different data types from farmers flows to different digital tools. Data exchange intermediary act as a facilitator of exchanges. Currently, data exchange with intermediary still not achieved wide application. These digital tools might be connected to open-source data. Image above refers to AgDataHub model.

4.2 Opportunity, challenge, constraint, limitation

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 questionnaires used are designed to cover general issues and are not personalized between different respondents.

Opportunities:

1. The existence of new applicable data regulations.
2. Digitalization of the agriculture sector.
3. Development of data exchange platforms.
4. Massive data innovation in agriculture sector.

Limitations:

1. Regulations and law.
2. Data size.

Challenges:

1. How to get good quality data
2. Increase data fusion application and consolidation
3. Encourage data sharing to gain more benefit.

Constraints:

1. Skepticism from stakeholders in sharing data.
2. Diverse and unharmonized data.
3. Limited knowledge of good data governance practices.

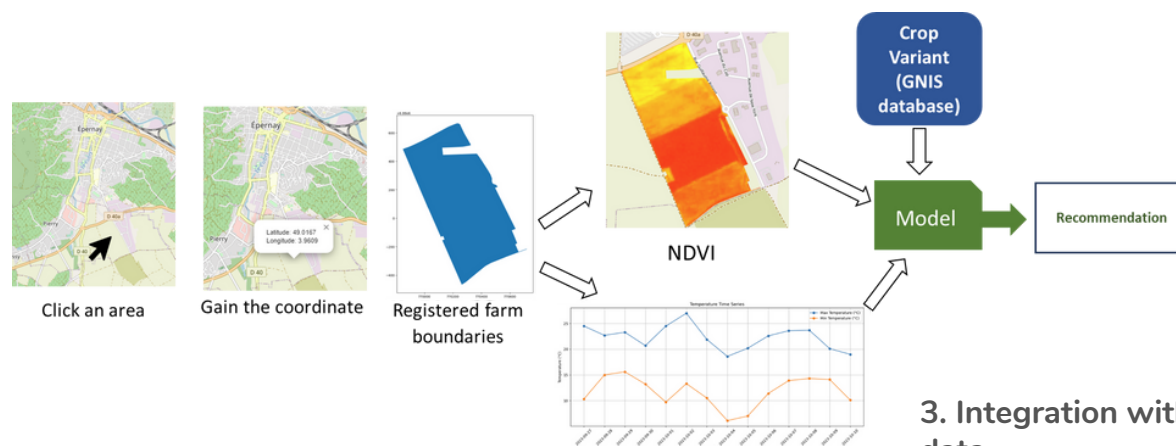
4.3 Terres Inovia decision tool

This study case was carried out using a dataset 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. As the result, three developments were conducted.

1. Connection to variety database (SEMAE) for more valid varieties

Added value: Validate and update varieties

Constraint: Variety database (SEMAE) does not provide an API version, the data must be downloaded manually (interoperability issues)



2. Integration with geo-location data

Added value: Enrich the input for measuring crop density and colza growth, possible to give more accurate results.

Constraints: Big data size for querying farm boundaries, unupdated data, and low resolution satellite image (data quality issues)

3. Integration with weather data

Added value: Enrich the input for measuring risk by adding weather variable

Constraint: Limited requests (need to pay for having larger requests per second or for commercial purposes) (accessibility issues)

5. Conclusion and recommendation

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 Recommendation

1. **Collaboration with other actors (public or private) who also work in this topic** -> Wider the impact with more participation from key actors, and communication on project's result.
2. **Reflection on upcoming/existing efforts of data valorization**-> Participation on AFNOR standardization's agenda, EU project (AgriDataSpace) (if possible)
3. **Develop use case (with regards to existing regulation)** -> Providing actionable methods and analyse the benefits/constraint to specific agriculture issues.
4. **Increase education and encourage FAIR principles implementation** -> Key to tackle big data issues in agriculture domain.