D4.4: Initial agri-food data-sharing reference architecture

WP4 – Data-driven Technological Innovation

Authors: Nikos Kalatzis, Christopher Brewster, Jack Verhoosel, Barry Nouwt, Han Kruiger, Antonis Georgousis

PLOUTOS

This project has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement No 101000594

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101000594

Document Information

G.A. No.	101000594	Acronym	Ploutos	
Full Title	Data-driven sustainable agri-food value chains			
Horizon 2020 Call	RUR-06-2020: Innovative agri-food value chains: boosting sustainability-oriented competitiveness			
Type of Action	Innovation Action			
Start Date	1st October 2020	**Duration**	36 months	
Project URL	Ploutos_h2020.eu			
Document URL	-			
EU Project Officer	Ivana Oceano			
Project Coordinator	Nikolaos Marianos			
Deliverable	D4.4: Initial agri-food data-sharing reference architecture			
Work Package	WP4 – Data-driven Technological Innovation			
Date of Delivery	**Contractual**	M6	**Actual**	M6
Nature	R – Report	**Dissemination Level**	P-Public	
Lead Beneficiary	Agricultural University of Athens (AUA)			
Lead Author	Nikos Kalatzis	**Email**	n_kalatzis@neuropublic.gr	
	NEUROPUBLIC	**Phone**	+30 2104101010	
Other authors	Nikos Kalatzis (NP), Christopher Brewster (TNO), Jack Verhoosel (TNO), Barry Nouwt (TNO), Han Kruiger (TNO), Antonis Georgousis (NP)			
Reviewer(s)	Valentina Manstretta (HORTA), Ali Zidi (Hispatec)			
Keywords	Technologies, innovations, technological requirements, gap analysis			

Document History

Version	Issue Date	Stage	Changes	Contributor
0.1	15/03/2021	Outlining the structure	Initial TOC and document structure	NP
0.2	15/4/2021	Requirements and initial architecture	Requirements and architecture added	NP, TNO
0.4	15/5/2021	Architecture finalised	Refinements and descriptions elaboration	TNO
0.5	7/6/2021	Consolidation of conclusions and introduction.	Consolidation and proof reading	NP, TNO
1	15/6/2021	Deliverable finalized	Deliverable reviewed by Ploutos partners	Horta, Hispatec, NP

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101000594

Table of Contents

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101000594

Table of Tables

Table of Figures

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101000594

Executive summary

The objective of this document is primarily to describe the Ploutos Data sharing framework which consists of a set of data processing and communication mechanisms. The target audience of this document are technically trained colleagues in this project, especially computer scientists and ontology engineers, as well interested scientists, researchers and developers in the wider ecosystem of agritech innovation where the focus on interoperability in multi-stakeholder systems.

In order to explain the foundations of the proposed data sharing approach, we will initially provide the high level design principles along with the technical requirements that have been extracted though an analysis process of the individual SIPs. Then the document elaborates on the actual data functional components through the provision and analysis of different architectural views.

This deliverable is structured as follows: In Chapter 2 a brief overview on Agri-food data sharing fundamentals are presented along with contemporary approaches on data interoperability. In Chapter 3 a detailed analysis on the design principles and the technical requirements which drove the specification of the Ploutos data-sharing framework are presented. In Chapter 4, the actual data sharing mechanisms to be developed within the Ploutos project are presented. Different architectural views (functional, informational, deployment) are presented that will support the upcoming development of these mechanism. It should be noted that the provided technical specification are an initial approach and it is expected to be updated and refined during the development phase. The implemented mechanisms will be tested and utilized for the needs of selected Ploutos SIPs. However, this mechanism aim to be generic enough and able to be utilized and serve any agri-food ICT system that addresses some minimum technical requirements (e.g. provide data through well specified APIs).

The two core entities of the Ploutos Data sharing framework are the Ploutos Interoperability Enabler (PIE) and the Ploutos Registration and Discovery Directory. As it will be further elaborated the PIE connects to existing information systems, retrieves on-demand the appropriate data sets, translates them to a standards-based data model and sends them to the authorized recipient. The Ploutos directory service essentially allows various PIEs to demonstrate their existence and the respective properties of the provided datasets.

In Chapter 5, the foreseen implementation technologies to be utilized are presented. Again this initial approach is expected to be further refined during the implementation phase. Chapter 6 presents the conclusions of this deliverable.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101000594

1 Introduction

1.1 Project Summary

The Ploutos project focuses on rebalancing the value chain for the agri-food system, transforming it into one that works for the benefit of society and the environment. The project will develop a Sustainable Innovation Framework that applies a systemic approach to the agri-food sector, building on three pillars: Behavioural Innovation, Sustainable Collaborative Business Model Innovation and Data-driven Technology Innovation. Exploiting a history of significant agri-food projects and the respective ecosystems around them, the project will deploy 11 innovative systemic Sustainable Innovation Pilots, where by adopting a Multi-Actor Approach innovative solutions and methodologies will be implemented, tested, assessed, generating practical learnings. The pilots cover a large range of agri-food ecosystems, across 13 countries, covering arable, horticulture (both open fields and greenhouses), perennials and dairy production among others. In each case, behaviour change, collaborative business modelling and data driven innovation will be integrated to deliver the most environmentally, socially, and economically sustainable solution. Moreover, a Ploutos Innovation Academy will be established as a vehicle for integrating the know-how, best practices and assessments developed across the project and derived from the Sustainable Innovation Pilots. Ploutos includes 33 partners, 22 of them being end-users, representing all relevant actors in the food system, including farmers, food industry companies, scientists, advisors, ICT specialists and policy makers.

1.2 Document Scope

Deliverable D4.4, “Initial agri-food data-sharing reference architecture” builds on the outcomes of D4.1 “Technology assessment and Ploutos technical requirements” and complements D4.2 “Data interoperability for the agri-food sector”. In Deliverable D4.1, we have described a range of current technologies which will be of relevance to the project and analysed the Sustainable Innovation Pilots (SIPs) for their core technological requirements. As noted in that document, Ploutos will focus its technological developments on interoperability architectures enable effective integration of stakeholders across pilots and supply chains, enabling technological solutions for farm management, precision agriculture and integrated supply chains to share data in an effective manner. Such a level of interoperability necessitates the adoption of an agreed semantic data model, or ontology. The data modeling approach, named the Ploutos Common Semantic Model (PCSM), to be utilized towards the realization of the Ploutos vision is presented in details D4.2. The PCSM is an ontology that incorporates existing well established (standardized when available) domain ontologies. It is also extendable allowing the incorporation of additional domain specific ontologies where in our case these domains are related to each individual SIP.

Having specified the PCSM it is necessary to specify and develop the respective data manipulation mechanisms that will allow the instantiation and utilization of the data model in a “real world” environment. These mechanisms – the Ploutos data-sharing framework - are specified within this document. The framework consists of two main functional modules: the Ploutos Interoperability Enabler (PIE) and the Ploutos Directory Service. Both modules aim to address a set of technical requirements which in turn reflect a set of real world prerequisites that dictate data-sharing in the agri-food domain.

1.3 Document Structure

The document is structured as follows:

Chapter 1 presents a summary of the project as well as the document scope and structure.

Chapter 2 presents Agri-food data sharing fundamentals, approaches on data interoperability along with dominant initiatives on Data Sharing.

This project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No 101000594

Chapter 3 elaborates on the design principles and the technical requirements which drove the specification of the Ploutos data-sharing approach.

Chapter 4 provides different views (functional, informational, deployment) of the Ploutos Data Sharing Framework

Chapter 5 presents the implementation technologies to utilized along with the implementation plan.

Chapter 6 presents the conclusions of this deliverable.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101000594

2 Agri-food data sharing fundamentals

In this section a brief but essential analysis is provided aiming to set the current status and background on data sharing for the agri-food sector. The aim is to identify the main concepts that characterize the ecosystem where the envisioned data-sharing mechanism will be deployed and operate.

Need for data sharing in agri-food: Information has always been shared informally in agricultural and food interactions. Whether it was the face-to-face visit of a consumer to their local butcher or fruit and vegetable market, or whether it was at agricultural fairs for food producers to share expertise, seeds and other useful data, the existence of information (and therefore data) sharing has been a fact of since the dawn of agriculture. With the growing complexity of the food system in the last 70+ years, the increasing globalisation of the food system, and above all the increasing length of supply chains, sharing of information and data between stakeholders has become both more difficult and more important. The vast number of stakeholders, the great heterogeneity of type and size, the great differences between different supply chains in length and complexity are all problems making data sharing in the supply chain more difficult. On the other hand, the need for data sharing has never been greater. Data is needed above all to respond to food crises (Brewster & Seepers, 2018), but equally in order to optimise supply chains and to enable traceability for a variety of purposes (Scholten et al., 2016). With the digitisation of farm practices and decision support systems (FMIS), data needs to be shared not just vertically up and down the supply chain but also horizontally e.g. between comparable farmers needing to share data benchmarking or productivity comparisons.

Challenges: Two practical aspects make data sharing particularly difficult. First, actors large and small in the food system have different ICT systems for undertaking various tasks including accounting, invoicing etc. These systems are not compatible between each other, and the requisite infrastructure has been until now missing (Brewster et al., 2017). Second, stakeholders at every stage from farm to fork are unwilling to share data for either privacy reasons, business confidentiality or suspicions concerning the uses that will be made of the data (Castle et al., 2016). For farmers and food producers, a number of key players have developed the "EU code of conduct" which lays out more explicitly contractual data sharing (cf. below), but this does not necessarily allay concerns (Copa-Cogeca et al., 2018; van der Burg et al., 2020). Regarding the supply chain, the EU's General Food Law (2002)[1] mandates the keeping of records but does not oblige these to be in electronic form let alone in an interoperable format, and there is no obligation to share data except under instructions from food agencies.

Core requirements: Current reality is that data, if exchanged at all, is shared in multiple formats and through various modalities (including manually handing over a usb key). Current systems and proposed architectures assume one-to-one interoperability between systems and at most allow for one actor to have access to all data (cf. Figure 1)

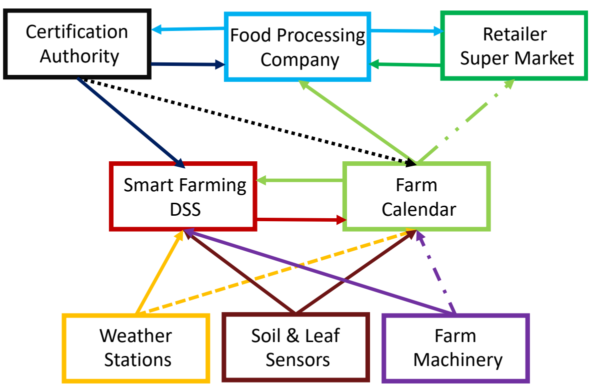

Figure 1 Current practices for data sharing (if it occurs).

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101000594

This means systems are fragmented/isolated - act as "vertical silos" – impeding the creation of cross-domain, cross-platform and cross-organisational services and innovation. Thus, one core requirement is **data interoperability** which means that **data standards** are used to represent shared data. The second core requirement is that there should be some form of **common architecture** that provides the infrastructure for data to be shared, queried, or retrieved across a multiplicity of data providers. In view of the, perfectly justified, concerns of different stakeholders concerning who has access to data and the use that is made of that data, a third important requirement is that any system enabling data sharing also allows for **access control**, allowing data providers to determine who has access.

2.1 Data interoperability

Data interoperability means that data on one computer system can be transferred to another system and used in the same way. This has been both a challenge and an ambition for computer systems designers at least since the invention of SGML in the 1980s which was designed to enable document sharing. There are different levels of interoperability (Ouksel & Sheth, 1999):

- Technical Interoperability: usually associated with communication protocols and the infrastructure needed for those protocols to operate.
- Syntactic Interoperability: usually associated with data formats and encodings, e.g., XML, JSON and RDF.
- Semantic Interoperability: associated with a common understanding of the underlying meaning of the exchanged content (information).

If all three above are achieved then "organisational interoperability" may be possible, where organisations can collaborate at a data level effectively. Achieving technical, syntactic and semantic interoperability is hard for a number of reasons, both technical and social. When successful this means in practice that data sharing is possible, and does not necessitate one-to-one adaptors/translators or application specific APIs. At an abstract level, a scenario as shown in Figure 2 is the overall ambition:

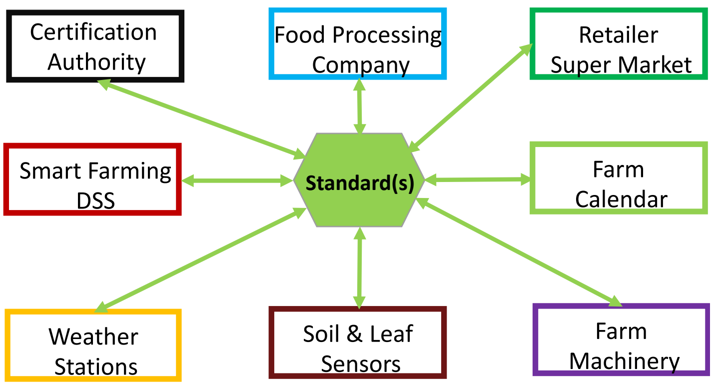

Figure 2 The ambition for data interoperability

The most widely used standards[2] in agriculture and food are usually *syntactic* standards with limited or underspecified semantics. Examples include ISOBUS for farm machinery, a number of different UN/CEFACT standards including those used in EDI, and the GS1 standards EPCIS and GTIN. This has resulted in frequent absence of interoperability, some of which has been mitigated over time. So, the AEF holds "plugfests" for farm machinery so that vendors and purchasers can check on the effective interoperability of farm machinery

[1] https://ec.europa.eu/food/horizontal-topics/general-food-law_en

[2] For a review of relevant standards on which the Ploutos Common Semantic Model has been based cf. Deliverable D4.2 from the Ploutos Project.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101000594

using ISOBUS. FOODEX2 although defined as an XML message standard, has extensive codebooks providing details on both permissible values and the interpretation thereof. More *semantically* specified standards in the agrifood sector include AGROVOC, largely used to annotate documents, AGRO for agronomic experiments, and FOODON which is a rigorous ontology built on the OBO Foundry principles.

For the purposes of the Ploutos project we have a developed the Ploutos Common Semantic Model which is largely based on existing semantic standards, ontologies, that are in wide or growing use across the agrifood system. Details are presented in Deliverable D4.2 while for reasons of completeness a short summary is presented in section 4.2.

2.2 Data Initiatives enabling Data Sharing

There are a number of initiatives, especially at the European level, which are worth mentioning because they form part of the wider context in which data sharing and data standardisation is taking place.

The FAIR Data Principles: The Life Sciences have for a long time been at the forefront of initiatives to promote principles and practices which will make it easier for researchers to share and reuse data. The FAIR principles are a set of principles (Wilkinson et al., 2016) which have received wide attention and a mostly positive response from the research community, not just in the biological and life sciences but also in many other research areas. The principles emphasise the importance of metadata about data or other research outputs, so as to allow them to Findable, Accessible, Interoperable, and Reusable. There is no longer an expectation that all data be open or directly linked to other data sets because *access* is recognized an important parameter. But from a scientific perspective, the key is to ensure that the descriptive metadata is appropriate, interoperable and permanently available. There is a growing awareness in other sectors that these principles are of great utility including agriculture (Caracciolo et al., 2020), especially in research data but we would argue the principles are equally of relevance in the food production and supply chain processes. The principles have been challenged in some cases from an ethical perspective and also from the perspective that the principles may reflect a colonial attitude (Boeckhout et al., 2018; Carroll et al., 2021).

Code of Conduct for agricultural data sharing: As a result of a collaboration between COPA-COGECA[3] and CEMA[4] together with other important associations representing different parts of the agrifood sector, a "code of conduct" was proposed in 2018 (Copa-Cogeca et al., 2018). The code focussed on non-personal data, and tried to make explicit the rights of "data originators" and "data owners" mainly by emphasising the need for commercial contracts make clear rights and responsibilities. This code was very advanced for its time and has helped to make regulators (largely the EC) think more carefully about the kind of ambitions to have concerning data sharing in the agrifood space. The code has come under some criticism in that it does not address the issue of the inherent power imbalances that exist in the food system (van der Burg et al., 2020).

Common European Data Space in Agriculture: The EC has a strategic position to support the creation of a number of "data spaces" including one for agriculture[5]. They see data as coming from a number of different sources (cf. Figure 3) and would like to see an aggregation or federation of these different sets of data. This will be supported by means of a number of different funding "instruments" and other related initiatives including a cofounded partnership "Agriculture of Data" over coming years.

[3] Farmers and agrifood cooperatives association https://copa-cogeca.eu/

[4] Trade association of European agricultural machinery manufacturers https://www.cema-agri.org/

[5] http://dataspaces.info/common-european-data-spaces

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101000594

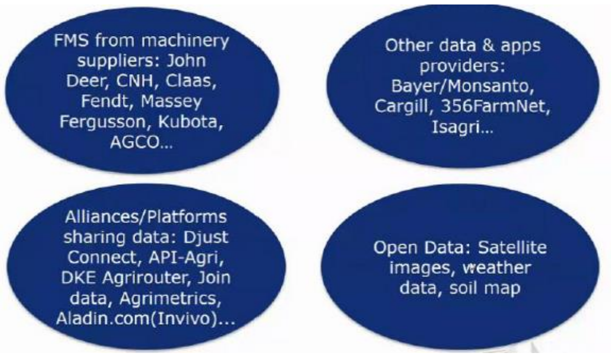

Figure 3 The range of data sources relevant to a data space in agriculture (from EC presentation)

In order to realise this ideal, there exist a number of challenges including technical, social and business-related issues to overcome. The AIOTI[6] noted particularly the need to federate at a data level rather than at a platform level, the need for security mechanisms to ensure farmers have control over their data, the importance of interacting with legacy systems, and issues concerning interaction of farm data with personal data from a regulatory perspective[7].

There are number of social, political and regulatory issues which need to be borne in mind. It is widely acknowledged by farmers and other food producers that they do not want the sharing of data to result in data monopolies and result in further consolidation. There is strong resistance to the possibility that data may be used to further monitor farm activity. And the social and ethical implications of these technologies will need to be considered further by technologies as well as regulators and policy makers.

[6] https://aioti.eu/

[7] AIOTI WG6 Smart Farming and Food Security position paper, available at https://ec.europa.eu/digital-single-market/en/news/expert-workshop-common-european-agricultural-data-space-0

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101000594

3 Design Principles

This section elaborates on the high level principles and the technical requirements that will drive the specification of the Ploutos data sharing framework.

3.1 SIPs analysis technical outcomes

This deliverable builds on the work conducted by T4.1 - documented in deliverable D4.1 “Technology assessment and Ploutos technical requirements” – where existing, well-established but also emerging technologies and innovations in the agri-food sector are presented along with the outcomes from a number of interviews with the various Ploutos SIPs. Building on the technologies’ assessment of T4.1, existing data modelling and data management standards will be partially or wholly integrated in order to develop the Ploutos data sharing mechanism that will facilitate heterogeneous data integration and interoperability for the agri-food domain. It should be noted that the analysis on Ploutos data models that will be utilised for harmonising the description of the various information entities are documented in deliverable “D4.2: Data interoperability for the agri-food sector” (Verhoosel et. al, 2021). This deliverable (D4.4) will elaborate on the data management architecture and mechanisms - that will also integrate the data models provided by D4.2 – and it will provide the necessary specifications that will drive the implementation of the actual data sharing mechanisms. These mechanisms will then be applied and evaluated in support of a number of selected Ploutos SIPs.

Information in table 1 provides the outcomes of the technology types to be utilised in Ploutos SIPs according to the replies received. These outcomes are extracted by the SIP-dedicated workshops and from information that was directly provided by the SIP participants conducted in the context of T41.

Technology Type	SIP ID
“Smart farming” or “IoT” or “Data logging equipment for agricultural machinery”	1, 3, 4, 5, 6, 7, 8, 11
Traceability	1, 2, 3, 4, 5, 7, 9, 10, 11
Data interoperability, “Data sharing”, “Data driven technology"	1, 2, 3, 5, 6, 7, 9, 10, 11
Decision Support Systems (DSS)	2, 9, 10
Insurance platform	2, 10
Data driven technology	5
API for satellite imagery	8
Algorithms for detection of measures from the raw data	8

Table 1. Technology types to be utilised in Ploutos SIPs.

As it is evident smart farming services are dominant in almost all SIPs and will be provided by technologies provided by various SIP members. These services are mainly facilitating data collection through various sources namely, IoT sensors, weather monitoring/forecasting services, farmers’ field book, etc. The collected data are then processed and visualised illustrating the most important information aspects, both current and historic, on a farm level. These data collections are also utilised for supporting the decision making process with regards to various cultivation practices. Often the decision making process is assisted by intelligent computational approaches incorporating statistics, machine learning, and artificial intelligence. These data are often collected, processed and maintained by systems also known as Farm Management Information Systems (FMIS). Table 2 present the various, existing, commercial, Farm Management Information Systems that have been reported to serve the needs of the respective SIPs.

This project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No 101000594

FMIS or DSS or Data platforms in Ploutos SIPs.	SIP ID
Gaiasense (Neuropublic)	1, 3, 7
vite.net DSS (Horta)	10
granoduro.net (Horta)	2
ERP Agro (Hispatec)	4
NADIA (IoT solution, Anysolution)	11
EvenKeel (for data sharing)	5

Table 2. Data platforms that will participate in Ploutos SIPs

According to (Fountas et. al, 2015) FMISs are systems that aim to keep track of farm activities and to manage the large amount of information generated through the use of ICT solutions. There are currently hundreds of FMISs available on the market that can support decision making by finding the best practices for farm management. According to (Sorensen et. al, 2010) the FMIS is defined as: "a planned system for the collecting, processing, storing and disseminating of data in the form of information needed to carry out the operations functions of the farm." Each FMIS usually focuses on one or multiple domains of the agricultural sector, for example, livestock or arable farming. In general, an Information System aims to support decision making by providing timely information about the planning, control and operational functions of an organization (Watson et al. 1991). In a similar manner, the FMIS does the same for the agricultural domain usually aiming to reduce the production costs, maintain high quality and to comply with the agricultural standards.

The second outcome derived by the SIP-dedicated workshops is that almost all SIPs are aiming to achieve "Data interoperability" and "Traceability services" during the SIP realisation. These are also the main objectives of the data management mechanisms that will be developed by WP4 and will serve the various SIPs. As it will be analysed in the following sections the WP4 data sharing mechanisms will be generic enough to enable data interoperability by binding with existing systems while the initial implementation will be tested in the context of selected Ploutos SIPs.

3.2 High level principles

In this section a set of high level design principles that will dictate the overall functionality of the data sharing mechanisms are presented. This list is an extension of the design principles on semantic interoperability mechanisms presented in (Kalatzis, Routis, et. al, 2019):

Standardised semantics: The information model to be selected for facilitating data interoperability should comply with standards and demonstrate all these features that are required for operating in the targeted application domain. Multiple standardised data models possible to be applied on the same application domain complicates things - introducing the need for cross-standard interoperability mechanisms.

Seamless integration: The interoperability mechanisms should not affect the existing operation of the IoT deployment or to affect it as minimal as possible. Real world, operational deployments tend to be complicated including various customised engineering approaches. Data Interoperability Enablers to be Plug-n-Play and light-weight (as possible).

Information completeness: Data interoperability is associated with the use of data translation. The interoperable data model and the data conversion mechanism should be able to support the available proprietary/customised information elements as much as possible. Although it might not be necessary to expose the full set of information entities as some of them might only useful for internal use the interoperability mechanisms should have the capacity of modelling and managing the full set of data.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101000594

Data volume optimisation: Use of semantics increase overall data volume. Operational IoT systems tend to minimise the overall data volumes using simplified formats modelled for example in CSV and JSON. Translating to XML, RDF, OWL, etc. introduces a significant overhead.

Control of information flow: The administrative entity that owns or manages the data should also control which information elements are leaving their cyber-premises. Easy to comprehend and clear administration and configuration operations are necessary

Security and Privacy protection of individuals: Keep balance between monitoring of every-day activities for various application domains without violating human rights (e.g. privacy). Advances on sensing technologies and data collection mechanisms make feasible the deployment of vast sensor networks that can potentially become intrusive and violate established regulations (e.g., GDPR)

Reuse: Already significant efforts invested on data interoperability (even in parallel). Not to reinvent the wheel, not to build another data platform.

3.3 Software requirements specification

This section specifies the technical requirements that are related to the data collection and data management functionalities that are necessary for the realization of the PLOUTOS pilots. The requirements are divided into categories, functional and non-functional, and are recorded according to the VOLERE[8] standard. The VOLERE standard defines a set of types of each requirement as it is presented in the Table below:

Functional	Functional	FUNC
	Data	DATA
Non-functional	Look and Feel Requirements	L&F
	Usability Requirements	USE
	Performance Requirements	PERF
	Security	SEC

Table 3. Categories of requirements according to Volere Standard

The following requirements have been extracted based on the analysis of the SIPs combined with the specified high level design principles. These technical/software requirements will then be mapped to the actual functional component of the Ploutos data-sharing framework.

Requirement ID:	**R1**	**Version:**	**0.3**	**Last Update Date:**	**10/1/2021**
Title	Ensure Semantic Interoperability of heterogeneous data items				
Description	Heterogeneous data items should be modeled based on if possible standarised data modeling approaches. Use of core ontologies in combination to domain specific ontologies allows the expandability of the data model capabilities.				
Involved stakeholders/actors	Solution providers, technology providers, PLOUTOS advanced services				
Type	Functional				
Priority Level	Mandatory				
Identified by Partner(s)	NP, TNO				
Status	Confirmed				
Category	DATA				
Comments/Remarks					

8 https://www.volere.org/templates/volere-requirements-specification-template/

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101000594

Requirement ID:	R2	Version:	0.3	Last Update Date:	10/1/2021
Title	Ensure Syntactic Interoperability with the various data sources.				
Description	The system should provide a set of formal data format specifications and the ability to exchange information in order to communicate on a technical abstraction level with the various data sources.				
Involved stakeholders/actors	Solution providers, technology providers				
Type	Functional				
Priority Level	Mandatory				
Identified by Partner(s)	NP, TNO				
Status	Confirmed				
Category	DATA, FUNC				
Comments/Remarks					

Requirement ID:	R3	Version:	0.3	Last Update Date:	10/1/2021
Title	Support seamless integration with legacy information sources				
Description	The interoperability mechanisms should not affect the existing operation of the systems or to affect it as minimal as possible. The PLOUTOS interoperability enabling modules needs to be versatile and easy to adapt to the underlying system. Use of translators that will convert legacy data to the semantically annotated (e.g. OWL).				
Involved stakeholders/actors	Solution providers, technology providers				
Type	Functional				
Priority Level	High				
Status	Confirmed				
Identified by Partner(s)	NP, TNO				
Category	DATA, FUNC				
Comments/Remarks					

Requirement ID:	R4	Version:	0.3	Last Update Date:	10/1/2021
Title	Ensure Interoperability in data provision by legacy data sources				
Description	PLOUTOS platform will provide a common query language interface via a interoperable/standarised API that will allow to extract data from heterogeneous databases and data sources. The service API should be able to direct the query to a specific database or source, using query languages like SPARQL (able to provide integrated view over different datasets) or SQL syntax. The API should represent results in a standard format (e.g., JSON or XML), and it should support, if possible, content negotiation to allow the clients to specify their preferred representation for results..				
Involved stakeholders/actors	Solution providers, technology providers, end users				
Type	Functional				
Priority Level	Mandatory				
Identified by Partner(s)	NP, TNO				
Status	Confirmed				
Category	DATA, FUNC				
Comments/Remarks					

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101000594

Requirement ID:	R5	Version:	0.3	Last Update Date:	10/1/2021
Title	Asynchronous exchange of data				
Description	The system should be able to operate in Asynchronous mode. Asynchronous operation is realized based on notification mechanisms to register data consumers upon specific events. Often a publish/subscribe process is followed.				
Involved stakeholders/actors	Solution providers, technology providers				
Type	Functional				
Priority Level	High				
Identified by Partner(s)	NP, TNO				
Status	Confirmed				
Category	DATA, FUNC				
Comments/Remarks					

Requirement ID:	R6	Version:	0.3	Last Update Date:	10/1/2021
Title	Support of enhanced descriptions for data items				
Description	The data modeling approach to be implemented for the needs of PLOUTOS should be expandable and to foresee the assignment of additional descriptions in the form of metadata.				
Involved stakeholders/actors	PLOUTOS advanced services, Data providers, end users				
Type	Functional				
Priority Level	High				
Identified by Partner(s)	NP, TNO				
Status	Confirmed				
Category	DATA, FUNC				
Comments/Remarks					

Requirement ID:	R7	Version:	0.3	Last Update Date:	10/1/2021
Title	Make data discoverable/findable				
Description	Data provision mechanisms should ensure that the various data items/ data sets will be discoverable with the use of quering mechanisms. This requirement is aligned with the FAIR principles. The use of Linked Data mechanisms will support this functionality.				
Involved stakeholders/actors	PLOUTOS advanced services, Data providers, end users				
Type	Functional				
Priority Level	Mandatory				
Identified by Partner(s)	NP, TNO				
Status	Confirmed				
Category	DATA, FUNC				
Comments/Remarks					

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101000594

Requirement ID:	R8	Version:	0.3	Last Update Date:	10/1/2021
Title	Ensure availability of datasets				
Description	The datasets/data items should be available for retrieval through the use of standardised well documented APIs along with robust data provision mechanisms.				
Involved stakeholders/actors	PLOUTOS advanced services, Data providers				
Type	Functional				
Priority Level	Mandatory				
Identified by Partner(s)	NP, TNO				
Status	Confirmed				
Category	DATA, FUNC, PERF, SEC				
Comments/Remarks					

Requirement ID:	R9	Version:	0.3	Last Update Date:	10/1/2021
Title	Ensure confidentiality and integrity of datasets				
Description	The confidentiality and integrity of datasets/data items should be ensured through the use of information security mechanisms (e.g. access control, cryptography).				
Involved stakeholders/actors	PLOUTOS advanced services, Data providers				
Type	Non-functional				
Priority Level	Mandatory				
Identified by Partner(s)	NP, TNO				
Status	Confirmed				
Category	PERF, SEC				
Comments/Remarks					

Requirement ID:	R10	Version:	0.3	Last Update Date:	10/1/2021
Title	Ensure ethical data utilisation				
Description	The utilization of data sets/data items should confront with current (legal/ethical) regulations. (Code of Conduct, GDPR) This requirement will be addresses via the implementation of both technical and administrative/regulatory means.				
Involved stakeholders/actors	PLOUTOS advanced services, Data users				
Type	Non-functional				
Priority Level	Mandatory				
Identified by Partner(s)	NP, TNO				
Status	Confirmed				
Category	SEC, DATA				
Comments/Remarks					

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101000594

Requirement ID:	R11	Version:	0.3	Last Update Date:	10/1/2021
Title	The data provider has the means to control the data sharing mechanisms and policies				
Description	The administrative entity of the Information Management system that provides the various data sets has the appropriate means for total control on which data items are shared and with whom.				
Involved stakeholders/actors	PLOUTOS advanced services, Data users				
Type	Functional				
Priority Level	Mandatory				
Identified by Partner(s)	NP, TNO				
Status	Confirmed				
Category	FUNC, SEC, DATA				
Comments/Remarks					

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101000594

4 Ploutos Data Sharing Framework

Having already presented in the previous sections the main requirements and concepts of the Ploutos data-sharing approach, in this section the Ploutos data sharing framework is presented. The description of the architecture is based on ISO/IEC/IEEE 42010:2011 methodology which involves specifying the architectural viewpoints that address stakeholders' concerns formulating their requirements and creating for them consistent architectural views using architectural models. The following viewpoints will be presented:

1. High-level functional viewpoint which will describe interactions, relationships and dependencies between the system and its environment that will interact with the system itself, other systems, users, and developers. In addition, it will describe the main functional elements of the architecture, interfaces and interactions.

2. Information viewpoint which will describe the way that the Ploutos data sharing framework will handle data heterogeneity. The main goal of the introduced semantic modelling framework is to provide a common communication language between stakeholders. Although the details of the Ploutos information model are elaborated in “D4.2: Data Interoperability for the Agri-Food Sector” (Verhoosel et. al, 2021) in order for this document to be self-contained a short overview is also presented here.

4. Deployment viewpoint which will describe how and where the system is deployed, considering hardware and physical dependencies. It provides consistent mapping across the existing and emerging technologies and the functional components specified in the Function View.

5. Message sequence diagrams and information flows based on the Ploutos data sharing framework for selected indicative use cases. The use cases are based on selected SIP scenarios.

4.1 High-Level Functional View

According to the technical requirements one of the core objectives of the Ploutos data sharing framework is to achieve the controlled and technically sound flow of data among the various information providers and consumers without at the same time disturbing the current operations of the underlying systems. In order to achieve this the **Ploutos Interoperability Enabler** (PIE) is introduced which aims to be generic enough in order to be deployed as a plug and play extendable module on top of the targeted system with minimal customisation efforts. The PIE operates in combination with the **Ploutos Registration and Discovery Directory** in order for the various PIE to announce their existence and their capabilities and hence be discoverable by entities interested to exchange data. One of the core characteristics of this approach is that data are not shared with a common third party but are exchanged directly among peers. As it is illustrated in figure 4 each information source maintains a specific adaptation of the PIE. In a nutshell, the PIE is able to receive queries for data selections from remote systems, to fetch the appropriate data sets from the underlying system, to translate these data to a common data model and to transfer the data to the remote system. The Ploutos Registration and Discovery Directory acts as a catalogue where the various PIEs are registering their existence along with operational properties/capabilities. The functionalities of these two core components are elaborated at the following sections.

This project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No 101000594

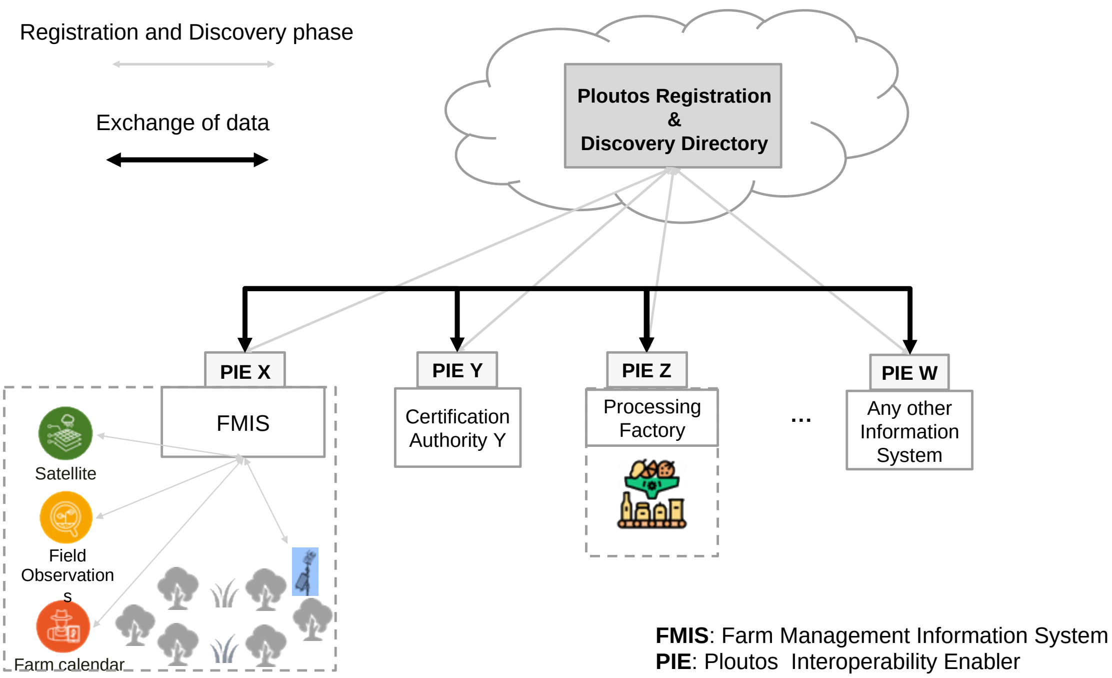

Figure 4. High level view of the Ploutos data sharing architecture

4.1.1 Ploutos Interoperability Enabler

The Ploutos Interoperability Enabler from a deployment perspective is hosted within the underlying system's (e.g. FMIS, DSS, data collection service) administrative cyber premises as a trusted service while the overall functionality and data sharing is feasible to be controlled by the administrators of the system. The PIE consists of a set of functional components which aim to extend the functionality of existing systems with specific features. The PIE's main role within the Ploutos architecture is allowing knowledge bases to exchange data in an interoperable manner with other participants of the Ploutos data sharing network. A Knowledge Base (KB) can be any service, application or platform that: 1) needs certain knowledge in order to function, 2) provides certain knowledge that others might need, or 3) both. Examples of Knowledge Bases are: a service that provides a forecast of local temperatures when given a GPS location, an app that gives insight into the supply chain of tomatoes, a platform that manages different sensors on a farm or a database that stores a farmer's planning.

A functional component diagram of the PIE along with the potential interactions with external systems is presented in Figure 5 and analysed in the following paragraphs.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101000594

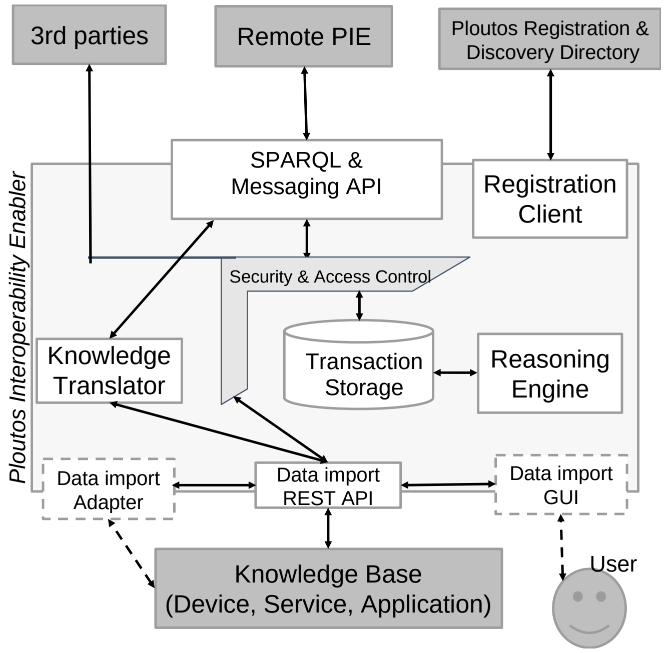

Figure 5. Functional component diagram of the Ploutos Interoperability Enabler (PIE)

Knowledge Mapper: It will enable interoperability at semantic level through a data translation service that will realise the conversion of data streams provided by the hosting system to selected standardised data formats and vice versa. The translation functionality of the "Knowledge Mapper" will be adapted depending on the properties of a selected common data model, however for the needs of the Ploutos project and SIPs the translation will be realized according to the properties of the Ploutos Common Semantic Model (PCSM) [D42].

Data import REST API: This component provides the main mechanisms for the PIE to exchange data sets from the underlying system on which is deployed. The main gateway of the PIE is a REST API that incorporates the necessary security mechanisms. It is expected that that the underlying "Knowledge Base" will consume this REST API in order to exchange data with the PIE. However, given the expected heterogeneity of the systems that the PIE will interact additional mechanisms are provided in order to facilitate the connection with these systems. In case the "Knowledge Base" already provides each own API the optional component "Data Import Adapter" is utilized which acts as an intermediate among the "Data Import REST API" and the API of the underlying system. In case the "Knowledge Base" doesn't provides any APIs or it is not feasible to consume the PIE REST API an alternative approach is supported by the PIE where a "Data Import Graphical User Interface" allows the manual importing of data.

Security and access control: Within the Ploutos ecosystem ensuring security and access control on data exchange is considered a transversal issue that spans the whole ecosystem. Security related functionality provided by PIE will be based on existing best practices ensuring Confidentiality, Integrity and Availability of data. The security mechanisms will be enforced at all the crucial data exchange phases (e.g. when a data query is received by the PIE, when data exchange with the "Knowledge Base" is taking place). The PIE will provide the means for enforcing the dictated authentication, access control and data governance policies during data exchange.

Registration client: PIE will support the underlying "Knowledge Base" to advertise its existence using metadata descriptions of the respective type of services that it offers (e.g. available information types, data utilisation policies, location, timespan of data sets). This registration process along with the advertisement of the PIE-Knowledge base capabilities will occur in an automated and periodically manner. The Registration Client periodically contacts the "Ploutos Registry and Directory service" in order to provide notifications on potential updates on PIEs capabilities. The registration process specifies how exchange of data and

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101000594

knowledge will be realized. For example, the following questions are answered during the registration phase of a PIE:

- What knowledge can be requested from me?
- What knowledge will I publish to the network?
- What knowledge will I request from the network?
- To which knowledge will I subscribe?

For example, a temperature sensor might regularly publish temperature measurements to the interested parties, and will respond to requests for the current temperature. A thermostat app might subscribe to knowledge about temperature measurements in a room, or request the current temperature. It might also publish current temperature preferences of a user. A heating system might subscribe to both the knowledge about temperature preferences and temperature measurements to be able to optimally control the temperature.

SPARQL & Messaging API: Each PIE maintains a SPARQL & Messaging API ensuring interoperability on syntactic level. The SPARQL Query Language is a Declarative Query Language (like SQL) for performing Data Manipulation and Data Definition operations on Data represented as a collection of RDF Language sentences/statements. Messaging functionality is achieved using the popular open source Apache ActiveMQ framework.

Reasoning engine: Two types of reasoning are realized within the Ploutos data sharing framework: a) reasoning to infer new data and b) reasoning for orchestration of data exchange. In the first case and in the context of the Semantic Web rule based analysis on the collection of existing data/facts allows to infer new facts and knowledge. As it is analysed in (Bienvenu et. al, 2020), this is a standard process for the semantic web and a plethora of reasoners[9] are available to be reused. For the needs of PIE, reasoning will be employed in selected cases where the enrichment of the knowledge base is considered useful for the needs of the SIP. The second case of reasoning refers to the orchestration of data exchange and will be analysed along with the "Registration and Discovery Directory" in the following section.

4.1.2 Ploutos Registration and Discovery Directory

The Ploutos Registration and Discovery Directory (PRDD) will be deployed at a cloud server system being accessible through Internet. The core objective of this directory service is to allow the registration/discovery of the various PIEs along with their characteristics and to support the orchestration of knowledge discovery. As it is already described, data sharing within the Ploutos ecosystem is ensured through the use of a common language expressed in the form of an ontology or knowledge model. The domain's knowledge model is written in RDF/OWL, which allows to take advantage of the reasoning capabilities that are available for these models. The complementary use of PIE and PRDD provides the necessary awareness about the supply and demand of knowledge in the network allowing the use of reasoning to orchestrate the knowledge supply on-demand. This means that, given a specification of knowledge that is requested, a PIE can figure out the appropriate knowledge base to get it. It should be noted, that this approach allows the complementary distributed querying of knowledge bases in order to serve one query. In addition, and given that the PIE is aware of changes in the network, new knowledge bases can be dynamically added to the network. In summary, the use of PRDD and PIE provides the following advantages:

• Knowledge orchestration removes the need to implement compatibility between all pairs of knowledge bases in the network by hand.

[9] https://www.w3.org/2001/sw/wiki/OWL/Implementations

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101000594

• Changes in the knowledge network are handled seamlessly by synchronizing information about knowledge interactions.

• Established open-source Semantic Web technologies are leveraged to provide knowledge models and reasoning capabilities.

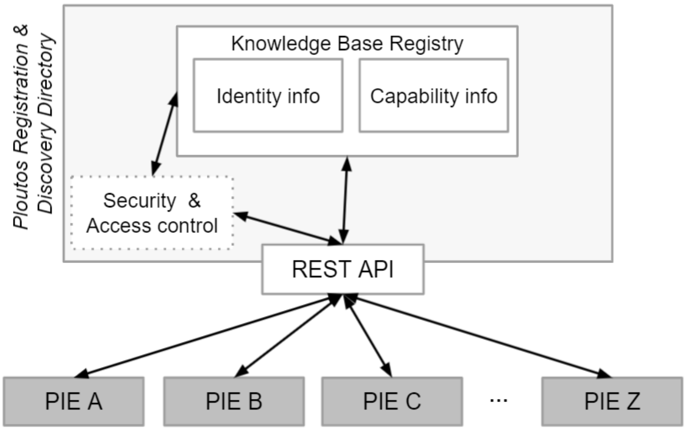

Figure 6. Functional component diagram of the Ploutos Registration and Discovery Directory

In figure 6 the functional components and their potential interaction of the Ploutos Registration and Discovery Directory are presented.

REST API: Communications to and from the PRDD are facilitated through a REST API. The REST API allows PIEs to register their identity and capability information. This metadata is also periodically requested by every PIE to update its internal state according to the latest updates in the Knowledge Network. This metadata is used by a particular PIE to determine which other PIEs should be involved in a data exchange.

Security and access control: This component ensures that only authorized entities are allowed to issue queries and retrieve insights on the properties of the registered PIEs. It should be noted that the PRDD only maintains meta data, so even a malicious reference has been added for a specific PIE, the targeted PIE is the entity that will finally grant access to the requested data or not. It is foreseen that standard security mechanisms will be employed for protecting the registry of PIEs such as Rules Based Access Control and cryptography.

Knowledge Base Registry: This component handles the registration process of the PIEs. When a new PIE appears on the Ploutos ecosystem it will contact the PRDD in order to announce 1) its existence with Identity information, 2) the way that it can be reached (e.g. IP address, URL, port), 3) the capabilities that it requires from other PIEs, and 4) the capabilities that it offers, and 5) the necessary access policies. The same process is applicable when an update on the condition of a PIE occurs.

4.2 Ploutos Common Semantic Model

Based on the requirements of the various SIPs and their data sharing intentions, a first version of a semantic model with common concepts and relations has been developed, the PCSM. In the next subsection, the guiding principles for the modelling and selection of concepts and relations in the PCSM will be described. Then, in the next subsection, the common concepts are briefly listed and the overall design of the PCSM in the form of an ontology is described. More details on the PCSM can be found in (Verhoosel et. al, 2021).

4.2.1 Guiding principles

For the design of the PCSM, a following guidelines have been used:

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101000594

1. The PCSM is based on semantic technologies, like RDF and OWL, because it is currently the best way of intuitively defining formal semantics (OWL) and provides the flexibility for modular reuse of existing data models or extending them.
2. The PCSM should be a small, core model that covers the main common concepts in the agrifood domain ranging from the farm via the supply chain to the consumer.
3. The concepts and relations for the PCSM are selected from the requirements of the SIPs. When most of the SIPs require a certain concept, it is part of the PCSM, e.g., the concepts farm, farmer, parcel, and soil that are required by most of the SIPs.
4. Existing ontologies that already define the required concepts are reused by the PCSM as much as possible. Nonetheless, we define a specific Ploutos namespace for the PCSM, namely https://www.tno.nl/agrifood/ontology/ploutos/common#prefixed as *ploutos*, that is used to inherit the concepts of these existing ontologies.
5. Existing ontologies are only reused when they have a clear formal OWL structure that is publicly available and accessible or downloadable in a .owl, .ttl or .rdf format, for instance at the W3C website or the AgroPortal (http://agroportal.lirmm.fr). Consequently, no reuse of proprietary ontologies of different projects will be done.
6. Vocabularies and thesauri/taxonomies that simply define and list a large set of hierarchical terms will not be reused in the PCSM other than using the *rdf:isDefinedBy* property to point to the definition of the concept in a vocabulary.
7. Reuse of existing concepts and properties in the PCSM is done using the *rdfs:subClassOf* or *owl:equivalentClass* construct for concepts and the *rdfs:subPropertyOf* construct for properties.
8. Concepts and properties that are required by the PCSM but are not yet part of existing ontologies will be added as concepts and properties to the *ploutos* namespace.
9. The well-known ontology design pattern called Part-Observation-Property pattern is used in which as much as possible concepts are expressed in a *ploutos:partOf* relation with another concept and measurements are defined as observations of observable properties of features of interest. More details on this pattern can be found in the following subsection.

A consequence of these guiding principles is that existing data models that are not proper ontologies are not explicitly inherited and extended by the PCSM. Unfortunately, extensive data models like NGSI-LD and vocabularies like AGROVOC fall into this category, as they are either vocabularies that can still be pointed at using a *rdf:isDefinedBy* or XML-based data models that cannot be reused as existing ontology, because they do not have the RDF-based linked data structure.

Finally, we follow the concepts and terms of existing ontologies and thus these terms might not fit perfectly for the agrifood supply chain domain, e.g., the term Observation and partOf relation. Therefore, we distinguish between model-technical terms and domain-specific terms and try to find a balance between an elegant model and recognizability for domain experts.

4.2.2 Main PCSM concepts

The following is a selection of high-level concepts that are required for the SIPs and will be part of the PCSM. They are selected based on the occurrences in the analysis in Chapter 4. In general, if a concept occurs in three or more SIPs, it is included here. We distinguish different types of concepts to be included in the core PCSM. Table XX gives an overview.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101000594

Category	Common concept	Specialization
Geographical-related concepts	Farmer	
	Farm	
	Parcel	
	Crop	
	Location*	
	Logistics unit*	
Material-related concepts	Soil	
	Fertilizer	
	Pesticide	
Action-related concepts	Operation	Product Operation
		Parcel Operation
	Observation	
Environment-related concepts	Temperature	
	Humidity	
	Precipitation	
	Wind speed*	
	Solar radiation*	

Table 4. The common concepts that are defined in the PCSM. The concepts marked with '' are currently not explicitly defined in the PCSM but may be added later.*

To model these common concepts, the existing ontologies listed in Table 3 are reused. Unfortunately, not all data models or ontologies we considered were compatible on a technology level and were therefore discarded. Obviously, there are also other ontologies that might become available soon in the necessary format. For instance, the EPPO[10] Global Database maintained by the EU EPPO organisation contains a lot of plant and pesticide information that can be reused easily once there is an ontology available that contains this information in an RDF and OWL-structure.

[10] https://gd.eppo.int/

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101000594

Prefix	Name	Base URI
ENVO	Environment Ontology	http://purl.obolibrary.org/obo/envo.owl#
s4agri	SAREF4AGRI	https://saref.etsi.org/saref4agri/
SSN	Semantic Sensor Network	http://www.w3.org/ns/ssn/
SOSA	Sensor Observation Sample Actuator	http://www.w3.org/ns/saso/
OM	Ontology of units of Measure	http://www.ontology-of-units-of-measure.org/resource/om-2/
Weather	BIMERR Weather Ontology	https://bimerr.oit.linkeddata.es/def/weather#

Table 5. Existing ontologies reused in the PCSM

4.2.3 Example of PCSM concepts

In this section we briefly describe the design of the PCSM and show an example diagram of the Ploutos concept Farmer, Farm, Parcel and Crop and how they reuse concepts of existing ontologies and their properties.

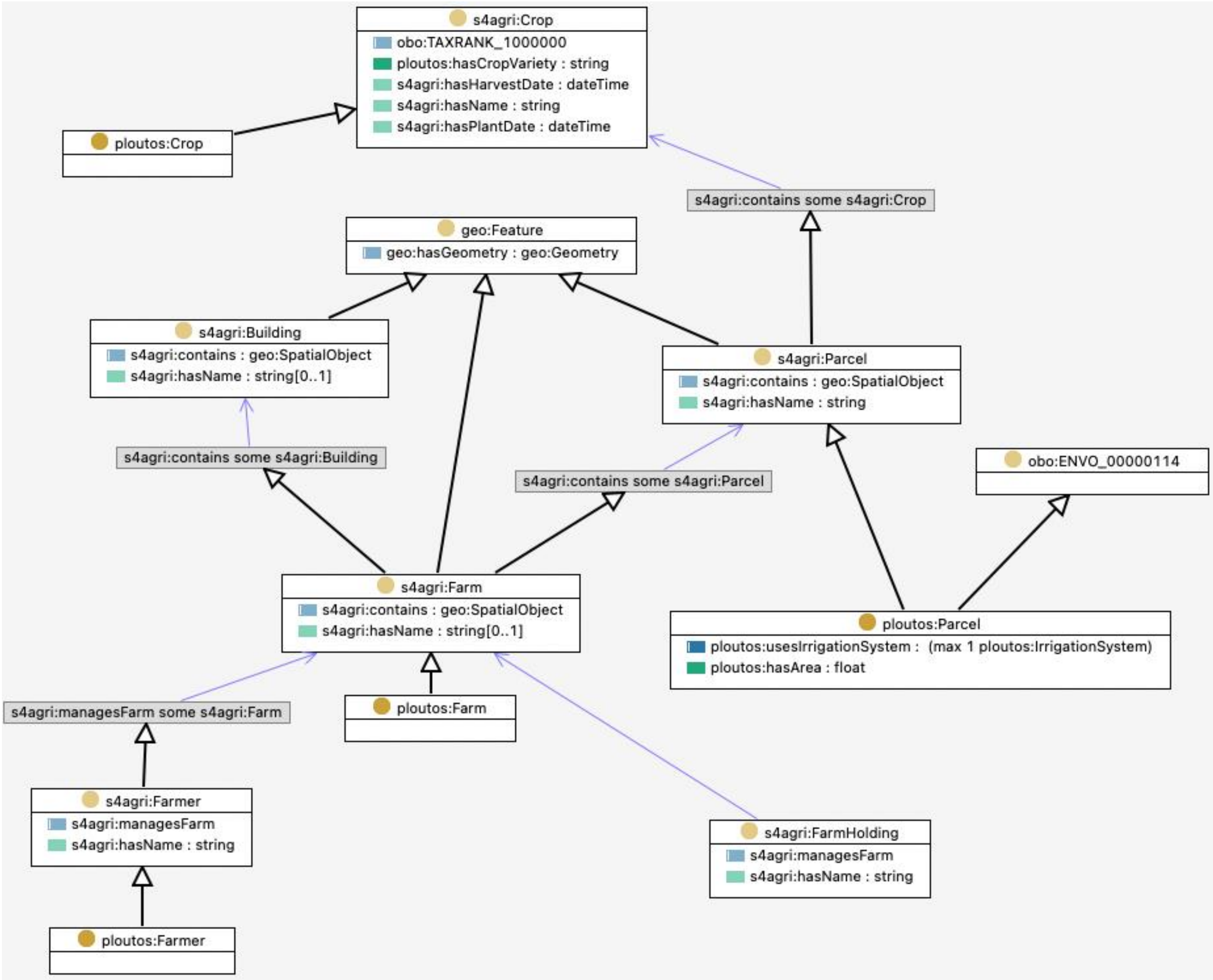

Figure 7. OWL Diagram of the Ploutos concepts Farmer, Farm, Parcel and Crop

For modelling the Ploutos concepts farmer, farm, parcel and crop we reused similar concepts in the saref4agri ontology and the ENVO ontology. A diagram of how these concepts relate can be found in Figure 7. As can be seen, the *rdfs:subClassOf* property is used to relate the Ploutos concepts with the concepts of reused existing ontologies. Thereby, instances of these Ploutos concepts directly inherit the properties of these reused concepts. However, where needed, additional properties are defined for Ploutos concepts, for

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101000594

instance the property *ploutos:hasArea* for concept *ploutos:Parcel*. In addition, the *s4agri:contains*property is reused to model the fact that a farm contains one or more parcels and that a parcel contains zero or more crops. Also important to note is that the ENVO ontology is mostly used to give extra meaning and explanation to a concept that is nicely defined by ENVO. So, the *ploutos:Parcel* concept is an *rdfs:subClassOf* of *obo:ENVO_00000114* which is the concept "agricultural field" that has a clear definition in the hierarchical structure of ENVO including synonyms like "cropland" or "grassland".

4.2.4 Usage of the PCSM in the Knowledge Engine

The PCSM is used by the Knowledge Engine as the schema for the data that is being exchanged. The Knowledge Engine works with any ontology described in the Web Ontology Language (OWL). For Ploutos we will use the PCSM, which is based on ontologies from the SAREF family, for the main common concepts. Any additional concepts that are needed for a specific use case can be defined by creating your own ontology that reuses the PCSM. The data itself is described in the Resource Description Framework (RDF). The Knowledge Engine uses graph patterns in terms of the PCSM concepts and properties to describe what data can be exchanged by a single Knowledge Interaction.

The Knowledge Engine uses Binding Sets to transmit the actual data. Data in RDF is always described as a set of triples. Each triple has the format subject–predicate–object. Individuals in RDF have an identifier in the form of a URI. The graph pattern describes what data can be exchanged using a certain Knowledge Interaction. A graph pattern is basically a piece of RDF (serialized as Turtle), where certain the subject, predicate and object can all be substituted by variables. When we have a graph pattern and a piece of data, we can express the piece of data in the context of that graph pattern as a binding.

4.3 Deployment View

The Deployment view depicts the key components of the Ploutos data-sharing framework from a system engineer's point of view. It is concerned with the topology of software components on the physical layer as well as the physical connections between these components. In figure 8 the PRDD.jar corresponds to the JAVA implementation - packaged as a jar file - of the Ploutos Registration and Discovery Directory service. In a similar manner, the Ploutos Interoperability Enabler corresponds to the PIE.jar files. It is important to mention that different instances of the PRDD.jar service can be deployed in various server systems in order to avoid a single point of failure during the PIE discovery process.

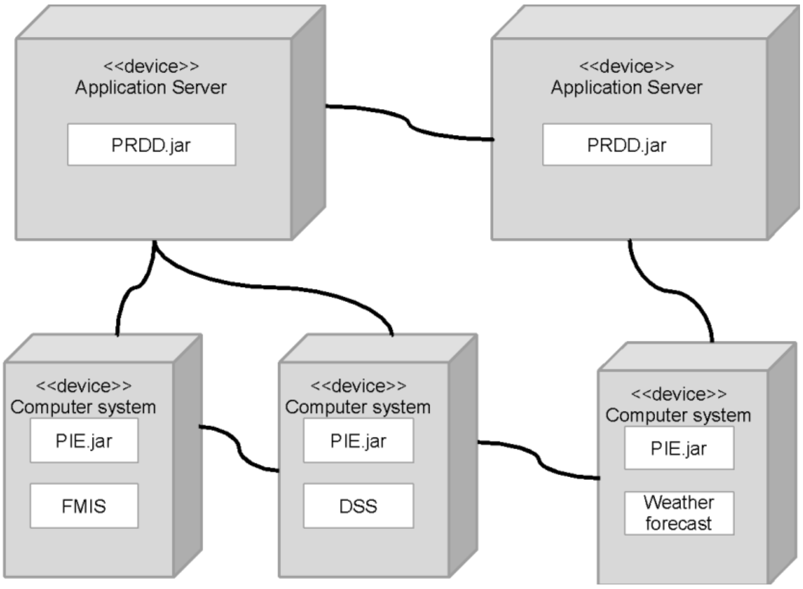

Figure 8. Deployment view of the Ploutos Data Sharing Framework.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101000594

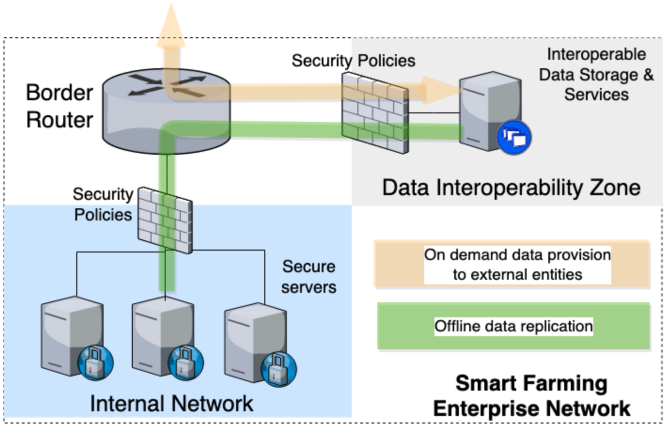

Figure 9. A deployment view of the Data Interoperability Zone

Figure 9 presents a deployment approach for the data interoperability services which have been in more details presented in (Kalatzis, et al. 2019). Based on this, the PIE is expected to be deployed on the administrative cyber-premises of the underlying smart-farming solution but at or near the local network perimeter following a deployment approach similar with the Science Demilitarized Zone (Science DMZ) (Crichigno, et al. 2018). This approach ensures that the internal operations of the Smart Farming enterprise network will remain intact while any special security and data transfer policies can selectively be applied to the portion of the network designed to facilitate the transfer of interoperable data.

4.4 Ploutos data sharing framework instantiation

In this section the specified Ploutos data sharing framework is being instantiated based on one of the Ploutos SIPs. This instantiation allows to specify the sequence of interactions among the key functional components and identify the information flow. For this exercise Ploutos SIP1 has been selected as an indicative use case that includes key entities -acting as knowledge bases- namely a Farm Management Information System, a Certification authority, and a food-processing factory. In more details SIP1 is entitled "Supporting a frozen fruit value chain consisting of small farmers, for optimising production, reducing environmental footprint and re-using data for certification and subsidies". According to SIP1 execution plan the following challenge is targeted:

"The vast majority of fruit producers in Greece face sustainability problems, as they have small and fragmented farms with fields that are often in different microclimate zones and face high inputs costs, lack of resources for investments and lack of financing. They take their everyday decisions based on either their intuition, or the acquired knowledge of their ancestors, facing these challenges with increased use of plant protection products, fertilizers and water, increasing production costs, risking overall quality, while depleting natural resources, and harming the environment. They need a way to reduce production costs and increase their revenues at an environmental friendly way. Proodos Farmers' Union in Pella, Greece, is such an example. ALTERRA is a food processing company that works closely local producers and cooperatives like Proodos to produce high quality frozen fruit products for a wide variety of food producers, food distribution networks and retailers /supermarkets. They usually offer contracts to farmers offering a better price for high quality products but they need a way a) to support the production of these products and b) to prove the high quality to their customers and build up their brand name ripping additional value. In addition, being able to get a certificate like GLOBALG.A.P. Farm Sustainability Assessment (GGFSA), they have access to new markets and buyers that offer higher price."

Based on this generic description SIP1 will pursue the following objectives:

1. Establishing gaiasense (http://www.gaiasense.gr), a Smart Farming (SF) solution in the area of Proodos, helping the farmers to reduce the application of inputs (water, fertilisers, pesticides), thus reducing the

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101000594

production costs, improving the fruit quality (less chemical residuals), while reducing the smaller environmental footprint of the production.

2. Connecting the gaiasense and the systems of Alterra with the Ploutos data sharing and traceability solution that will help in collecting all the needed data/proof to secure certification and sustainability related labels, like GGFSA, the "Macedonian Land Products" label of Central Macedonia Region (policy maker) and the "Gaiasense4earth" environmental performance label.

3. Measuring the additional value created by the new collaboration between the two value chain links (Proodos and Alterra) that enabled securing the certificates and labels based on the collected and shared data, and proposing a fair redistribution among them.

4. As an extra benefit, the farmers will be able to use the collected data as a secondary evidence in the context of the new monitoring and eligibility criteria controls (traffic lights system) of the Future CAP.

Objective number 2 (in bold) has been selected in order to be addressed by the Ploutos data sharing framework. The following steps illustrated in figures 10-12 elaborate on the sequence of exchanged messages towards the realization of the objective:

a. Registration of information sources

In this scene three main entities are participating, the FMIS that serves a number of selected farms of the Proodos Farmers association, the Certification Authority Y (capable to issue GlobaG.A.P. certificates) and the Alterrra frozen fruit Processing Factory. At this stage we assume that a PIE has already been deployed to each information system and that each PIE need to registere with the Ploutos Registration and Discovery Directory. The exchanged messages are illustrated in figure 5.

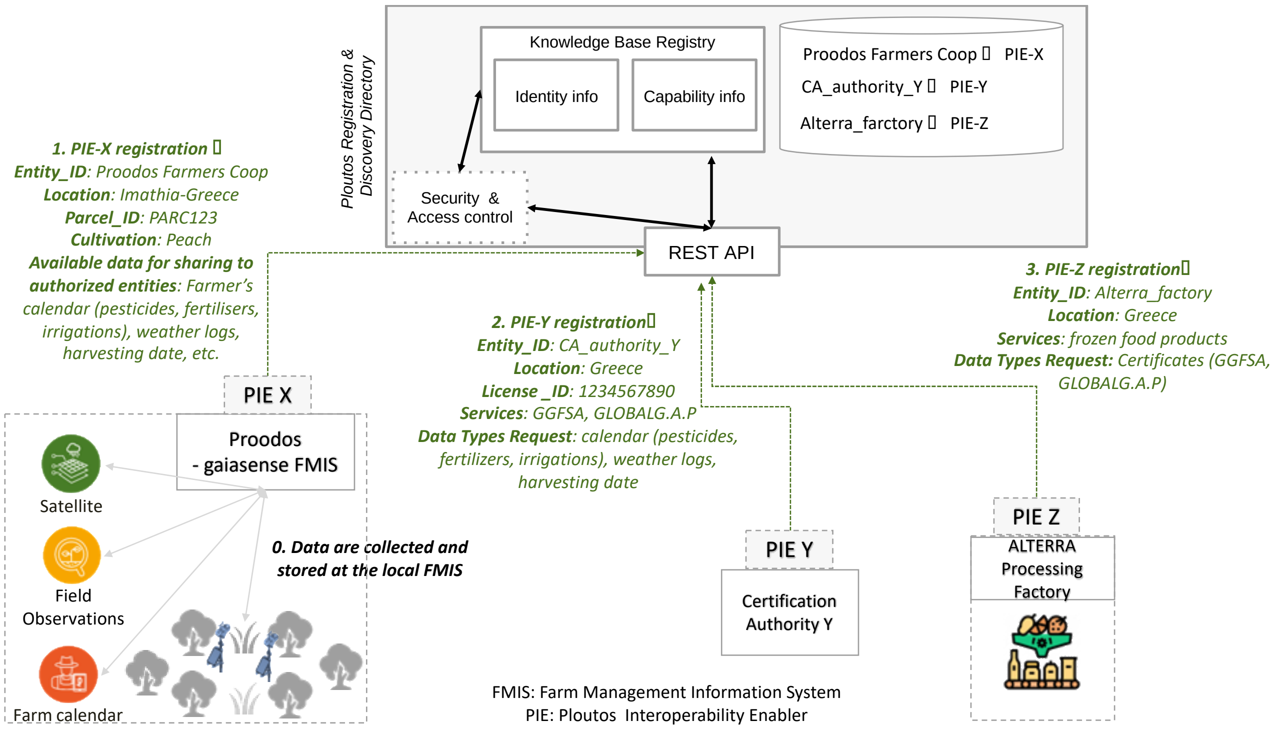

Figure 10. Registration of information sources

Each PIE demonstrates its self by sending a message that contains a unique PIE identifier, the entity that serves, the physical location (area) that is deployed, the parcel IDs (and polygon's coordinates) that the data collections refer to, along with the available data types for sharing or to be requested.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101000594

b. Data Discovery & Exchange - Issuing a Certificate

At this stage PIE X and PIE Y exchange a sequence of messages and data sets in order a certificate of good agricultural practices to be issued. The process is initiated by PIE Y (Certification Authority) that queries the Registration and Discovery Service in order to retrieve a data set of farming practices (e.g. farmer's field-book along with additional evidences) for a specified parcel. The Ploutos Registration and Discovery Service replies with a PIE identifier (PIE-X) capable to provide the requested data set. Then, PIE-Y requests the data set directly by PIE-X. The security and access control mechanisms verify the authorization of data access and then grand the access and transfer of data. An illustration of the exchanged messages are provided in figure 11.

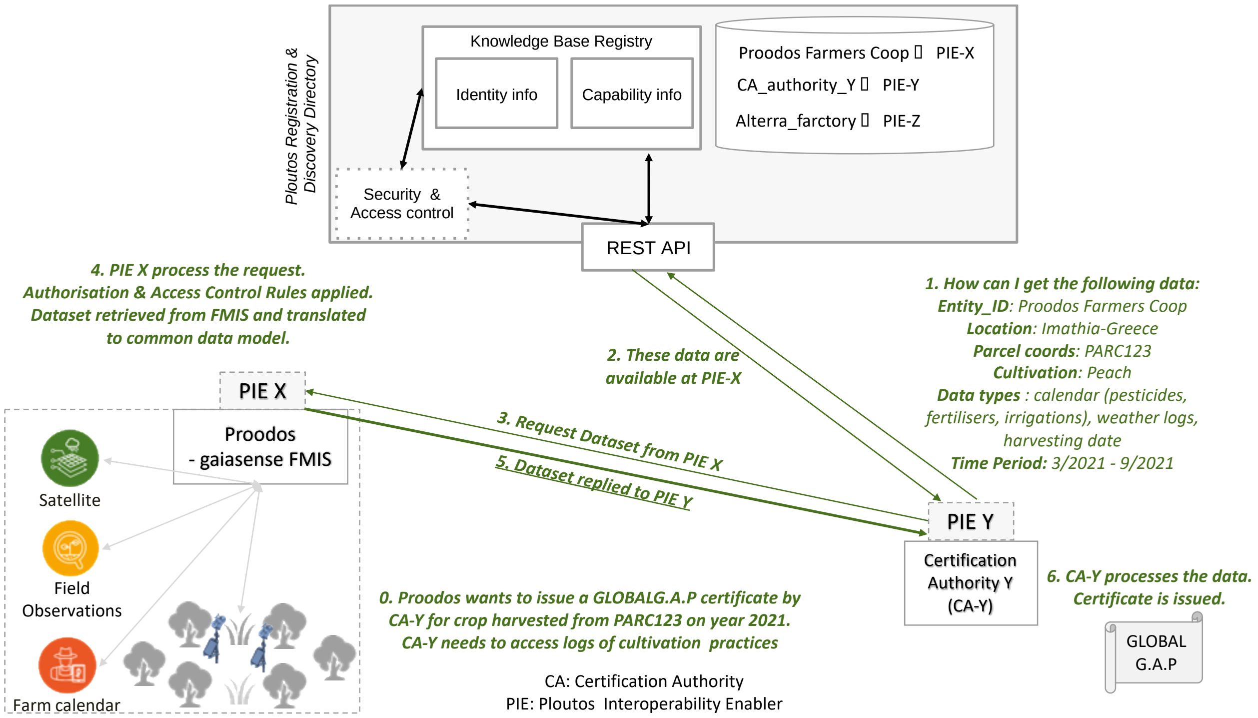

Figure 11. Data Discovery & Exchange - Issuing a Certificate

c. Data Discovery & Exchange - Retrieving a Certificate

At this stage the Alterra Frozen Fruit processing factory has received an amount of fresh fruits cultivated by farmers that are members of the Proodos Farmer's association. In the previous stage a certificate of good agricultural practices was issued for Proodos Farmers however now this certificate needs to be retrieved by Alterra. Through PIE Z deployed at the Altera factory the query of the date item that corresponds to the certificate is issued to the Ploutos Directory service. The Directory service replies that certificate-data are available through PIE Y. Then PIE Z directly contacts PIE Y in order to retrieve digital evidence of the certificate. The sequence of messages is illustrated in figure 12.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101000594

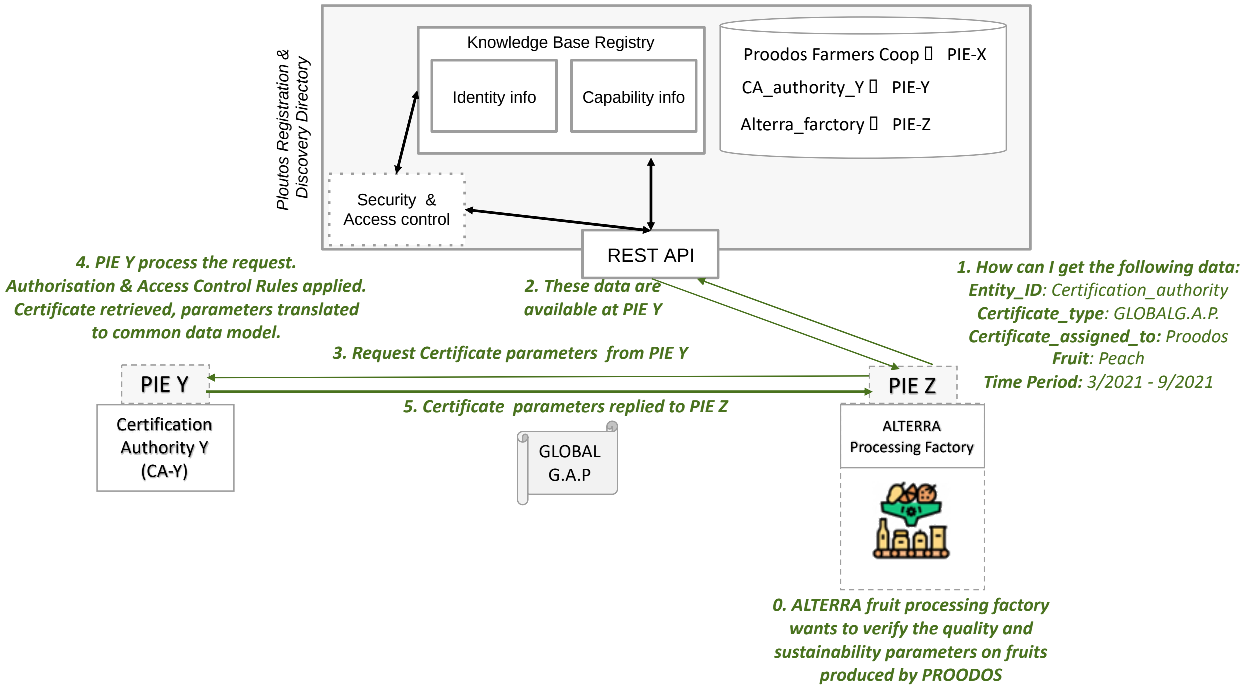

Figure 12. Data Discovery & Exchange - Retrieving a Certificate

5 Implementation

5.1 Technologies

The following technologies will be utilized for the development of the Ploutos data sharing components.

- Ploutos Interoperability Enabler: JAVA 8
- Ploutos Directory and Registry: JAVA 8
- Platform-independent deployment: Docker container images
- Knowledge model and graphs: W3C RDF and OWL
- Knowledge caching: Apache Jena Fuseki triplestore (open source)
- Query/response: W3C SPARQL language
- Reasoners:
    - Openllet reasoner (open source version of Pellet) for OWL reasoning
    - Apache Jena Generic rule reasoner with combined backward (SELECT) and forward (INSERT) chaining
    - Built-ins: generic Apache Jena built-in functions + user-developed built-ins
- Publish/subscribe messaging services: Apache ActiveMQ Broker/ Websockets+MQTT
- Access control: Ontology-based access control (OBAC) with read/write access policies

5.2 Implementation Plan

According to the Ploutos project Gantt plan the first release of the Ploutos Interoperability enablers will be ready for testing on September 2021 (M12). The code will be escorted by a report entitled “D4.6: Initial agri-food data-sharing platform implementation. It should be noted that this release will have the role of a proof of concept, however it will be followed by regular releases (every two months) where offered functionality will be incremented. On a similar manner the respective documentation of the services will further be elaborated. The final outcomes of the Ploutos data sharing services will be documented in Deliverable D4.5 “Final agri-food data-sharing reference architecture” to be delivered on M23 along with the D4.7: “Final agri-food data-sharing platform implementation” that will escort the final release of the code.

Traceability related functionality will in principle be based on the data sharing components presented in this document. However, customization of the specified services will be necessary according to the specificities of tracking and tracing within the food-chain. The design specifications and implemented services will be documented in D4.8 “Initial distributed ledger agri-food traceability and data sharing service specifications” to be delivered in M12 and D4.9 “Final distributed ledger agri-food traceability and data sharing service specifications” to delivered in M24.

This project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No 101000594

6 Conclusions

This deliverable focus on the Ploutos data sharing framework. The framework has been specified considering SIPs requirements (specified in D4.1), best practices on agri-food data sharing, and a set of software/technical requirements. One of the core objectives of the Ploutos data sharing framework is to achieve the controlled and technically sound flow of data among the various information providers and consumers without at the same time disturbing the current operations of the underlying systems. In order to achieve this the Ploutos Interoperability Enabler (PIE) is introduced which aims to be generic enough in order to be deployed as a plug and play extendable module on top of the targeted system with minimal customisation efforts. The PIE operates in combination with the Ploutos Registration and Discovery Directory in order for the various PIEs to announce their existence and their capabilities and hence be discoverable by entities interested to exchange data. One of the core characteristics of this approach is that data are not shared with a common third party but are exchanged directly among peers.

This deliverable sets the high level descriptions of the functional components that will be developed during the next months. As the development process evolves this documented will be regularly updated in order to deflect all the design decisions and technical/implementation aspects of the Ploutos data sharing mechanisms. On M23 the revised and complete version of this deliverable will be published aiming to act as a reference for achieving interoperable data-sharing for the agri-food sector in EU.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101000594

7 References

Bezuidenhout, L. (2020). Being Fair about the Design of FAIR Data Standards. *Digital Government: Research and Practice*, *1*(3), 1–7. https://doi.org/10.1145/3399632

Bienvenu, M., Leclère, M., Mugnier, M.-L., & Rousset, M.-C. (2020). Reasoning with Ontologies. In P. Marquis, O. Papini, & H. Prade (Eds.), *A Guided Tour of Artificial Intelligence Research: Volume I: Knowledge Representation, Reasoning and Learning* (pp. 185–215). Springer International Publishing. https://doi.org/10.1007/978-3-030-06164-7_6

Boeckhout, M., Zielhuis, G. A., & Bredenoord, A. L. (2018). The FAIR guiding principles for data stewardship: Fair enough? *European Journal of Human Genetics*, *26*(7), 931. https://doi.org/10.1038/s41431-018-0160-0

Brewster, C., Roussaki, I., Kalatzis, N., Doolin, K., & Ellis, K. (2017). IoT in Agriculture: Designing a Europe-Wide Large-Scale Pilot. *IEEE Communications Magazine*, *55*(9), 26–33. https://doi.org/10.1109/MCOM.2017.1600528

Brewster, C., & Seepers, R. (2018). *Food Integrity and Data Sharing along the Supply Chain: Overview* (Project Deliverable D17.2). FoodIntegrity Project.

Caracciolo, C., Aubin, S., Jonquet, C., Amdouni, E., David, R., Garcia, L., Whitehead, B., Roussey, C., Stellato, A., & Villa, F. (2020). 39 Hints to Facilitate the Use of Semantics for Data on Agriculture and Nutrition. *Data Science Journal*, *19*(1). https://doi.org/10.5334/dsj-2020-047

Carroll, S. R., Herczog, E., Hudson, M., Russell, K., & Stall, S. (2021). Operationalizing the CARE and FAIR Principles for Indigenous data futures. *Scientific Data*, *8*(1), 108. https://doi.org/10.1038/s41597-021-00892-0

Castle, M., Lubben, B., & Luck, J. (2016). Factors Influencing Producer Propensity for Data Sharing & Opinions Regarding Precision Agriculture and Big Farm Data. *Presentations, Working Papers, and Gray Literature: Agricultural Economics*. https://digitalcommons.unl.edu/ageconworkpap/48

Crichigno, J. , Bou-Har E., and Ghani, N. A Comprehensive Tutorial on Science DMZ, in IEEE Communications Surveys & Tutorials, 2018

Copa-Cogeca, CEMA, CEETTAR, ESA, Fertiliser Europe, F., FEFAC, ECPA, EFFAB, & CEJA. (2018). *EU Code of conduct on agricultural data sharing by contractual agreement*. Copa-Cogeca. http://www.ecpa.eu/sites/default/files/documents/AgriDataSharingCoC_2018.pdf

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

Kadadi, A., Agrawal, R., Nyamful, C., & Atiq, R. (2014). Challenges of data integration and interoperability in big data. *2014 IEEE International Conference on Big Data (Big Data)*, 38–40. https://doi.org/10.1109/BigData.2014.7004486

Kalatzis, N., Marianos, N., & Chatzipapadopoulos, F. (2019). IoT and data interoperability in agriculture: A case study on the gaiasenseTM smart farming solution. *2019 Global IoT Summit (GIoTS)*, 1–6. https://doi.org/10.1109/GIOTS.2019.8766423

Kalatzis, Nikos, Routis, G., Marinellis, Y., Avgeris, M., Roussaki, I., Papavassiliou, S., & Anagnostou, M. (2019). Semantic Interoperability for IoT Platforms in Support of Decision Making: An Experiment on Early Wildfire Detection. *Sensors*, *19*(3), 528. https://doi.org/10.3390/s19030528

Ouksel, A. M., & Sheth, A. (1999). Semantic interoperability in global information systems. *ACM SIGMOD Record*, *28*(1), 5–12. https://doi.org/10.1145/309844.309849

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101000594

Scholten, H., Verdouw, C. N., Beulens, A., & van der Vorst, J. G. A. J. (2016). Defining and Analyzing Traceability Systems in Food Supply Chains. In *Advances in Food Traceability Techniques and Technologies* (pp. 9–33). Elsevier. https://doi.org/10.1016/B978-0-08-100310-7.00002-8

Sørensen, C. G., Fountas, S., Nash, E., Pesonen, L., Bochtis, D., Pedersen, S. M., Basso, B., & Blackmore, S. B. (2010). Conceptual model of a future farm management information system. *Computers and Electronics in Agriculture*, *72*(1), 37–47. https://doi.org/10.1016/j.compag.2010.02.003

van der Burg, S., Wiseman, L., & Krkeljas, J. (2020). Trust in farm data sharing: Reflections on the EU code of conduct for agricultural data sharing. *Ethics and Information Technology*. https://doi.org/10.1007/s10676-020-09543-1

Verhoosel, J., Brewster, C., Kruiger, H., Nouwt, B., (2021). “D4.2: Data Interoperability for the Agri-Food Sector”, H2020 Ploutos projects.

Watson, H. J., Carroll, A. B., & Mann, R. I. (Eds.). (1991). *Information systems for management: A book of readings* (4th ed). Irwin.

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J. 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). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, *3*, 160018. https://doi.org/10.1038/sdata.2016.18

Zhai, Z., Martínez, J. F., Beltran, V., & Martínez, N. L. (2020). Decision support systems for agriculture 4.0: Survey and challenges. *Computers and Electronics in Agriculture*, *170*, 105256. https://doi.org/10.1016/j.compag.2020.105256

This project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No 101000594

END OF DOCUMENT

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101000594