Data Sharing in Agricultural Supply Chains: Using semantics to enable sustainable food systems

Christopher Brewster [a,c,*], Nikos Kalatzis [b], Barry Nouwt [a], Han Kruiger [a] and Jack Verhoosel [a]
[a] *Data Science Group,TNO, Kampweg 55, Soesterberg, 3769 DE The Netherlands*
E-mails: Christopher.Brewster@tno.nl, Barry.Nouwt@tno.nl, Han.Kruiger@tno.nl, Jack.Verhoosel@tno.nl
[b] *Neuropublic SA, 6 Methonis Str, 18545, Piraeus, Greece*
E-mail: Nikos_Kalatzis@neuropublic.gr
[c] *Institute of Data Science, Maastricht University, Maastricht, 6229 EN The Netherlands*

Abstract. The agrifood system faces a great many economic, social and environmental challenges. One of the biggest practical challenges has been to achieve greater data sharing throughout the agrifood systems and the supply chain, both to inform other stakeholders about a product and equally to incentivise greater environmental sustainability. In this paper, a data sharing architecture is described built on three principles a) reuse of existing semantic standards; b) integration with legacy systems; and c) a distributed architecture where stakeholders control access to their own data. The system has been developed based on the requirements of commercial users and is designed to allow queries across a federated network of agrifood stakeholders. The Ploutos semantic model is built on an integration of existing ontologies, and the Ploutos architecture built on a discovery directory and interoperability enablers which use graph query patterns to traverse the network and collect the requisite data to be shared. The system is exemplified in the context of a pilot involving commercial stakeholder in the processed fruit sector. The data sharing approach is highly extensible with considerable potential for capturing sustainability related data.

Keywords: data sharing, supply chain, agrifood, graph pattern, ontology, Farm Management Systems

1. Introduction and Motivation

The agrifood sector is facing a series of immense challenges including economic, social and environmental problems and predicaments. From a financial perspective, agriculture is frequently uneconomic with farmers facing continuous pressures to cut costs, consolidate or find alternative income streams. From a social perspective, there are has been a flight from the land for a long time with both push and pull factors that continue across the world today. Recent decades have seen an immense growth in scientific awareness of the environmental impact of agriculture. The agrifood sector is responsible for over 34% of greenhouse gases, uses 70-80% of global freshwater, causes immense biodiversity loss, and is suffering from the annual loss of 27 billion tons of fertile land due to land degradation [1–4]. The Ploutos project (https://ploutos-h2020.eu/) has been designed as a minor contributor to mitigating these impacts taking an integrative approach by bringing together innovations in behavioural change and business models, as well data driven technological innovation with the overarching ambition to re-balance the value chain and achieve greater sustainability in all three dimensions. Data driven technological innovation largely focuses on

*Corresponding author. E-mail: Christopher.Brewster@tno.nl.

ways to achieve greater data sharing across different sets of stakeholders. This paper describes technological infrastructure which has been implemented based largely on semantic technologies to enable such data sharing among multiple agrifood stakeholders.

Data sharing in this context refers to the design of systems, standards and infrastructure that enables data to be shared between stakeholders both with supply chain relations but also off-chain stakeholders such as certification providers and food safety agencies. The importance of data sharing in the agrifood sector is widely recognised [5–10], but for our purpose here it is needed primarily for:

a. Monitoring purpose e.g. monitoring of agricultural certification (GlobalGAP, organic etc.) or environmental reporting (of great significance in the new EU CAP).
b. Promotional and reputation management purposes e.g. providing information and data to consumers and retailers concerning agricultural practices.
c. Enable different value chain relations e.g. providing mechanisms for payments for specific agricultural or environmental practices such as carbon farming or protection of biodiversity.
d. Tracking and tracing e.g. identifying the location or origin of a specific product, or optimising supply chain flows [11].

There are, of course, many other purposes for data sharing in the agrifood sector, both currently identified, and which could potentially arise, mostly revolving around increased or better decision support for farmers and other food chain actors. Data sharing for any purpose has usually meant architectures and systems which were centralised either explicitly or implicitly [12]. This is apparent in most solutions which centred around the use of the GS1 standards, and (although seemingly contradictory) most systems built using one form or another of blockchain technologies. The literature (e.g. [13]), but above all our analysis of the real-life agrifood requirements and the barriers and drivers to technology adoption, led us to assume the following principles:

A. Stakeholders should retain control of their data.
B. Interoperability must be maintained with conventional software, whether legacy or developed in the future.
C. Vendor lock-in must be avoided thereby ensuring both flexibility and future proofing of any given part of the overall system

Principle A, control of data, implies that data should remain as much as possible in the hands of the actors concerned (e.g. the farmers), that full control of access should be possible, but also that queries or requests for data should be for specific purposes. Principle B, interoperability with other systems, means there is no expectation that a specific system will be adopted by all stakeholders, only that interoperability with past and future systems should be designed in. Principle C, avoiding vendor lock-in, implies that a standards-based, semantic technology driven approach is inevitable, and this implies use or reuse of widely accepted ontologies.

From a practical perspective data sharing has meant the application of semantic technologies, and to a greater or lesser extent the adoption of the FAIR data principles [14]. As demonstrated below, this has meant the use of ontologies, URIs, and some degree of reasoning. The *Findability* of data is ensured through the use of a registry and a discovery directory, the *Accessibility* of data is ensured by the use of standardised protocols integrated with access control, *Interoperability* of data is ensured by the (re)use of widely used ontologies/vocabularies that are accessible online, and the *Reusability* of data is ensured by using community standards and by ensuring data provenance is a consequence of data ownership and control. Obviously here the target is the relevant agrifood community rather that the wider scientific research community.

The rest of this paper describes the technology developed and deployed for the project and is structured as follows: Section 2 provides a review of related work, followed in Section 3 with an overview of the use cases (pilots) which have informed the design choices and the consequent requirements. Section 4 describes the semantic model developed for the project. In section 5, the overall architecture is presented followed by a worked example in Section 6. A discussion performance and feasibility is followed by conclusions and future work.

2. Related Work

This section presents related work addressing a) agrifood related vocabularies and ontologies, b) data sharing/-traceability systems using semantics, c) querying across federated data using semantics.

Agrifood related ontologies: There is a considerable body of work building ontologies for the food and agriculture domain which has gone hand in hand with the development of Linked Data (and "Linked Open Data") in the agri-food domain. The major effort here has been AGROVOC developed by the FAO and maintained by a network of institutes around the world [15]. It is nowadays the most comprehensive multilingual thesaurus and vocabulary for agriculture. AGROVOC has now been partially mapped onto the US National Agricultural Library of the USDA and the CABI thesaurus in the form of the GACS ontology which has mapped and integrated the top 15,000 concepts [16]. Other recent work in this area has also focused on developing ontologies for sharing of research data including the Crop ontology initiative, the Agronomy Ontology (AgrO), and the Plant Trait Ontology (TO) supported by CGIAR [17]. FOODON integrates a number of existing ontologies but its focus seems to be again on research data although its ambition is to provide a mechanism for data integration across the food system. Considerable efforts have been put into extending and integrating the FOODON ontology with various other ontologies extending its utility to areas such as nutrition, and integrating it with the Foodex2 standard from EFSA [18, 19].

As mentioned, most work on ontologies for the agrifood domain has up to now mostly been targeted towards the clear definition of domain concepts and terms in the form of a vocabulary for the annotation of research publications or research data sets [17]. As a result, there is little or no use yet of ontologies for supporting the actual sharing and exchange of data across the agrifood chain. Only a few papers about the use of ontologies for traceability and data analysis in the dairy sector [20, 21]. In addition, a few innovation projects have dealt with the use of ontologies in the horticultural supply chain and greenhouse domain [22]. However, the use of semantic standards for information exchange by standardisation organisations, like GS1, ISO or AgGateway is not yet common practice or even beginning to be picked up. More generally under the auspices of the EC funded projects like ATLAS (https://www.atlas-h2020.eu/) and especially DEMETER (https://h2020-demeter.eu/), there is support for greater use of semantic standards and technologies where the data and data models are explicitly specified, where URIs are widely used, and where data integration is consequently made far easier [23].

Data sharing architectures: The use of semantics for data sharing or data exchange in general has a long history epitomised by the Linked Data initiative and more recently the FAIR data movement. There has been a tendency for this to be largely a focus for researchers rather than commercial applications with the exception of the life sciences. Within the food and agriculture sector, the majority of efforts at data sharing have focussed on ontologies for research data and assumed relatively centralised approaches to repositories (e.g. the CGIAR platform for Big Data [17]), or else have been based on XML. Three major data standardisation efforts in the agri-food areas are ICAR (for livestock and dairy production), ISOBUS (for machine-to-machine data sharing) and AgGateway (for FMIS applications). These are currently available as XML standards (with codebooks) but are slowly moving towards JSON-LD versions of their standards. [24] make the case for using Linked Data principles and a variety of ontologies so that data can be integrated for farmer decision support. [5] emphasises the importance of semantics for integrating IoT derived data in agriculture, while [25] similarly uses ontologies to ensure data integration for supply chain data. These papers, as do most others, assume centralised architectures. The Linked Pedigree architecture was proposed to enable data sharing across supply chains by formalising the GS1 EPCIS standards as ontologies and enabling SPARQL queries across distributed triple stores, and this work partly inspires the technological approach described below [26, 27].

Interoperability and legacy systems: The issue of making existing legacy systems interoperable is not new and has been researched as part of the data integration topic [28]. The goal of data integration is focused largely on the need for answering user queries over distributed relational data sources and using SQL as the query language. The data integration landscape has changed with the increase in (types of) data sources and applications that can benefit from it. Other approaches that use semantic technology propose query-rewriting of the overall query to integrate distributed heterogeneous data sources [29]. This is different from our approach in that the data sources are integrated using semantic domain knowledge instead of a user query. Our focus lies on composing web service APIs of the distributed legacy data sources. Applying semantic technologies to web services has been the topic of extensive research. Two topics can be distinguished, although they are often investigated together. On the one hand, research has focused on giving semantic descriptions of the inputs/outputs and pre/post conditions of web services to be able to automatically discover or select relevant services [30]. On the other hand, research focused on how to automatically compose multiple web services together [31]. Despite the many proposals to include semantics into the descriptions of web services [32–34], many of today's web service APIs do not describe their formal invocation, pre- and post-conditions, and semantic input/output. The research on semantic web service composition in general has investigated several techniques to achieve this, and the approach described in our paper is similar to rule-based 'planning' [35–37]. However, in contrast to these approaches, the approach taken in this paper uses graph patterns of a common semantic model to be able to automatically compose distributed heterogeneous data sources.

In the area of Internet of Things, with the growth of IoT devices efforts have been devoted to overcoming interoperability challenges [38, 39]. In this context the W3C has launched the Web of Things Working Group (https://www.w3.org/WoT/) to counter the fragmentation arising from the use of IoT devices and support the use of semantic standards. This work largely builds on the Semantic Sensor Network (SSN) ontology (https://www.w3.org/TR/vocab-ssn/). The Open Mobile Alliance and a variety of telecommunication standards bodies have supported the Next Generation Service Interface (NGSI) [40]. The NGSI API and the associated NGSI context model has been adopted by key organisations in the IoT standardisation efforts . For an example of using this for data translation from multiple sources see [41] which has also influenced the architecture of our interoperability enablers described below.

3. The use cases and requirements

Project Pilots: The use cases which have driven the technological approach described below have arisen from the pilots of the Ploutos project, mentioned above. This three-year project, begun in Oct 2020, aims to integrate innovation in behaviour change and in business modelling together with data driven technologies to help rebalance the agrifood value chain and make the agrifood system more sustainable. The project has involved 33 partners including 22 end-users in 11 pilots, representing a variety of actors in the food system, including farmers, food industry companies, advsors, and ICT providers. The pilots cover a range of agri-food ecosystems, covering arable, horticulture (both open fields and greenhouses), perennials and dairy production among others. The objective is for each case, behaviour change, collaborative business modelling and data driven innovation to be integrated and to deliver the most environmentally, socially, and economically sustainable solutions. Table 1 provides an overview of the Ploutos pilots, the countries involved and the targeted agrifood sector. In each pilot, 3-4 partners are participating each one demonstrating complementary expertise (e.g. farmers, farmers associations, food processors, retailers and ICT providers).

Requirements Analysis Process: Each pilot analysed through a number of interviews with key participants and with the use of questionnaires aiming to extract user and technical requirements. From the analysis, it was evident that most of the Ploutos pilots demonstrated the need for more effective mechanisms for "data capturing", "data driven farm advisory services" and "to establish data flows" between key domain players such as "farmers", "processors", "retailers" and "consumers". Only those pilots which a) expressed a clear need for sharing and combination of data between different stakeholders, and b) demonstrated clear requirements for modelling data in a common vocabulary, were chosen so that six out of the eleven pilots have been the initial focus of development. For each of these pilots, an analysis was undertaken of their overall motivations, workflows and high-level architecture. This enabled the identification of the concepts which needed to be included in the common semantic model. A set of competency questions was defined as a result, as well as identifying relevant standards (vocabularies. ontologies) that could be reused. Here further details are provided for Pilot 1, as an example, although the full analysis can be found in [42].

Prototypical Pilot: Pilot 1 is typical in its complexity and requirements. The main pilot participants are located the Pella region of Northern Greece, and involve two farms of about 10Ha in total, which cultivate peaches. A fruit and vegetables processing factory participates in the pilot in the same area and here the harvested peaches are processed, packaged and sold as a frozen product. Farmers have to tackle a number of sustainability problems, as their farms are small, fragmented and in different micro-climate zones, while there is a lack of financial resources for ICT related investments. Current production methods suffer from increasing inputs consumption and costs with a consequent increase environmental footprint. Fruit processors operating in the area are following various schemes where better prices are offered for farmers with high quality products. The quality is evaluated by examining the characteristics of the fruits (e.g. taste, shape) but also by checking the cultivation practices followed. It is important for the fruit processors to have demonstrable evidence of the agronomic practices followed (e.g. by means of a "farm book" i.e. diary of actions undertaken) and the exact origin of the produced fruits. Equally the collection of this data is important for a number of certifications processes with regards to good agricultural practices (e.g.GLOBALG.A.P).

For the purposes of data monitoring at the field level and the optimisation of the farming practices, a commercial Farm Management Information System (FMIS) installed at the two farms involved in the pilot. This FMIS provides a network of agro-enviromental sensors and data collection mechanisms along with digital tools for decision support and digital recording of cultivation practices undertaken. The deployed FMIS is a proprietary system with its own custom data model and with limited capabilities for data sharing. The smart farming solution is supported by a local contractor (agronomist) that also offers technical support and training services to the farmers. The fruit processing factory already uses an Enterprise Resource Planning (ERP) system combined with a Warehouse Management System (WMS) in order to monitor, track and record the processing activities.

The main actors of this pilot are using legacy systems with limited capabilities for interaction (no standardised data formats and APIs) and they were reluctant to proceed with major changes in their operational ICT systems in order to facilitate data sharing. The actors also expected to have maximum control over who is accessing their data sets while they were unwilling to share their data beforehand with third parties.

Given that the raw and processed peaches are mainly exported to European countries, data modelling and data sharing mechanisms need to address not only the needs of the local stakeholders (e.g. in a regional/national level) but to follow current best practices in order to achieve a maximum level of compliance with dominant data sharing practices.

Competency Questions: Standard competency questions applied to most pilots, including questions, such as *"Who owns farm X?"*, *"What parcels belong to farm X?"*, *"Are peaches fruit?"*, *"Give me the GPS coordinates of farm X?"*. These questions while trivial are still important to validate the common semantic model. Pilot specific competency questions were identified in each case. For Pilot 1, the following specific questions were proposed:

- From what farm does batch X of frozen fruits originate?
- On what measurements is certificate Y for batch X based?

Table 1

Overview of the Ploutos pilots

No	Description	Countries Involved	Sector
1	Supporting a frozen fruit value chain with small farmers, to optimise production, reduce environmental footprint and re-use the data for certification and subsidies	Greece	Frozen Fruit
2	Better food-chain contracts for improved durum wheat production	Italy	Arable/Pasta
3	Empowering consumers through crowdsourcing to take back control over their food and create healthy, sustainable, fair trade products	France, Greece, UK, Germany, Belgium	Cross-sectoral
4	Traceability solutions covering the horticulture greenhouses value chain to improve operations, sustainability performance and brand recognition	Spain	Vegetables
5	Smart Farming on rural farms demonstrating its benefit in the wider agrifood community and co-creating new food products and services	Ireland	Livestock, arable / Food tourism
6	Applying soil-passport approach rewarding land owners/users and a precision farming solution to increase soil health and sustainability	Slovenia	Cross-sectoral
7	Supporting wine producers in taking advantage of the changes in labelling regulations and enhancing their sustainability performance	Cyprus	Wine
8	Carbon Farming: compensating farmers for climate friendly soil management	Netherlands	Cross-sectoral/ Organic
9	Facilitating the transfer of surplus food from farms to socially disadvantaged groups, by aligning logistics and processes	Serbia, N. Macedonia	Cross-sectoral
10	Increase sustainability in the grapevine sector by introducing payments for ecosystem services provision and parametric insurance to support losses from sustainable approaches	Italy	Fruit
11	Improving the sustainability of Balearics agri-food chains with Smart Farming and by using the collected info to organise agri-tourism activities	Spain	Vegetables/ Agri-Tourism

– What activities are (planned) on the farm calendar of farm X?
– How much pesticides, fertilizers and irrigation have been used on farm X in the period Y?

It was this detailed bottom-up analysis of pilot requirements, which has led to the adoption of the three principles described in the introduction above.

4. The Ploutos Conceptual Model

Based on the requirements of the various pilots and their data sharing requirements, a first version of a semantic model with common concepts and relations has been developed, the Ploutos Core Semantic Model (PCSM) [42]. In the next subsection, the guiding principles for the modelling and selection of concepts and relations in the PCSM will be described. Then, the common concepts are briefly listed and the overall design of the PCSM in the form of an ontology is described.

4.1. Guiding Principles

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

1. The PCSM is based on semantic technologies, like RDF and OWL, because this is the best way of defining formal semantics 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, a specific Ploutos namespace for the PCSM has been defined, 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, reuse of proprietary ontologies of different projects is avoided.
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, the concepts and terms of existing ontologies are followed, 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.

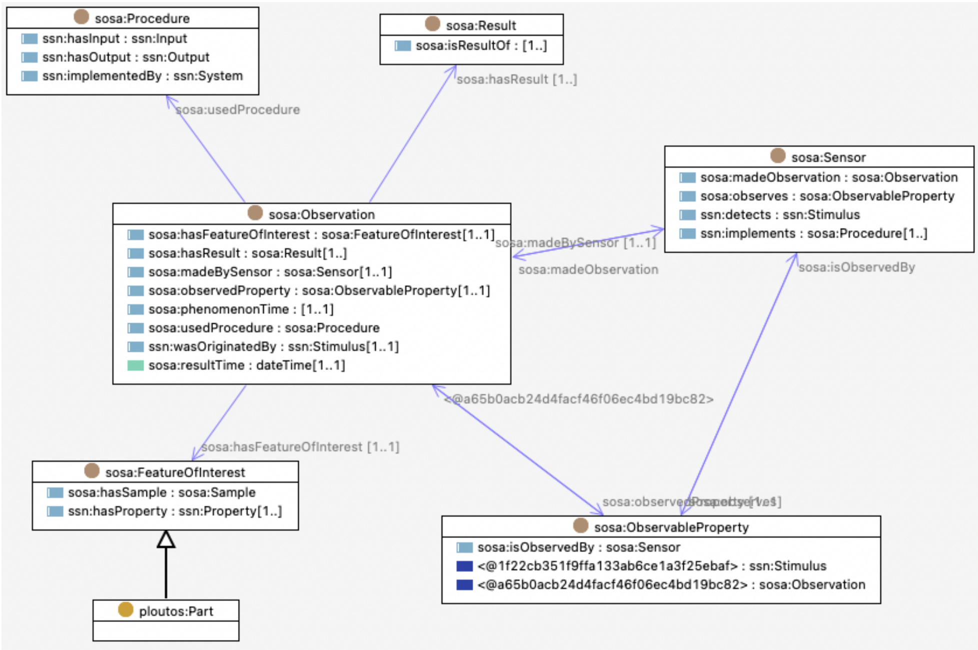

Fig. 1. OWL Diagram of the Part-Observation-Property pattern and the reuse of the SOSA ontology for this.

4.2. Part-Observation-Property design pattern

One of the guiding principles mentioned above is that of the `Part, Observation, Property` design pattern. The goal of this principle is to give structure to the different concepts and properties that appear in the PCSM. A good structure makes it easier to either *find* a concept that you are looking for or *add* a new concept that was still missing. The `Part, Observation, Property` design pattern is based on W3C's SOSA (http://www.w3.org/ns/sosa/) ontology and has been applied before in other ontologies. It can be applied in domains where you want to use different kinds of sensors to measure properties of things on different abstraction levels. For example, on the one hand you might want sensors on a drone to measure the drought of a particular parcel on a farm, while on the other hand a farmer might sample individual plants on the parcel and measure their growth. So, this design pattern can be used for the model to accommodate measurements on both abstraction levels and optionally allow them to be correlated at a later stage.

The pattern divides the concepts of the ontology into three pillars: parts, observations, and properties. Each of the pillars hosts different types of concepts that can be related with each other in predefined ways. The Part pillar consists of part-whole concepts. Within this pillar, concepts can be defined as being a member of the class `Part`, which indicates that the concept can be a part of another, larger concept indicated using the `partOf` relation. For example, both the concepts `Flower` and `Plant` are defined within the `Part` pillar and an instance of the `Flower` concept has a `partOf` relation with instances of the `Plant` concept, because the flower is part of the plant. Another example is that of the concept `Farm` and the concept `Parcel`. An instance of the concept `Parcel` is `partOf` an instance of the concept `Farm`, because parcels belong to a farm. Within the Part pillar a hierarchy of concepts arises that can range from very detailed (i.e. the organs of an animal) to more abstract (i.e. the farm or region).

Note that in this pillar it is important to distinguish between the hierarchy formed using the `subClassOf` relation and the hierarchy formed using the `partOf` relation. The first expresses a taxonomy where children are more specific than their parents and an instance of a child is also always an instance of its parent, while the second hierarchy expresses the structure of reality where a car, for example, consists of four wheels, a chassis and a body. Clearly, the rule that an instance of a wheel is also an instance of a car is not applicable here. In practice this means that wheel and car are related using `partOf` and not using `subClassOf`.

Table 2
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.

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*	

The Property pillar consists of all kinds of properties of the concepts from the Part pillar. For example, the property `Height` is applicable to the `Tree` concept and the `Greenhouse` concept and the property `Temperature` is applicable to the air surrounding the farm (i.e., weather), but also to heating pipes in a greenhouse. Other examples of properties are carbon dioxide levels, humidity, color and radiation. These properties are modelled in the PCSM as concepts in order to define the semantics properly and, even more important, to use them in relations to various other concepts, using for instance the `hasHeight` or `hasTemperature` relation.

The Observation pillar connects the concepts from the Part pillar with the concepts from the Property pillar to model measurements. An observation measures a certain property of a certain part. In principle, every concept in the Part pillar is a `FeatureOfInterest` that can have an `ObservableProperty` in the Property pillar of which measurements can be done. So, measuring the height of a tree is represented as an `Observation` of the `Height ObservableProperty` (property) with the `Tree FeatureOfInterest` (part).

Figure 1 shows the `ploutos:Part` concept as a subclass of `sosa:FeatureOfInterest` concept, which enables the entire `sosa:Observation` structure to be reused for modelling and instantiating measurements of a `sosa:ObservableProperty` of a Ploutos part. Note that this design pattern is a useful way of structuring our model, but not every concept must fit these three pillars. For example, treatments of parcels and packaging of meat do not naturally fit into these three pillars and have their own separate Operation concepts to represent them.

4.3. PSCM concepts and reuse of existing ontologies

The following is a selection of high-level concepts that are required for the pilots and are part of the PCSM. They are selected based on their frequency of occurrences of concepts identified in the analysis of all the pilots (cf. Section 3). In general, if a concept occurs in three or more pilots, it is included here. Different types of concepts to be included are distinguished in the core PCSM. Table 2 gives an overview.

To model these common concepts, the ontologies listed in Table 3 are reused. Unfortunately, not all data models or ontologies considered were compatible, either because the ontology was not in a standard format, or because it was not dereferenceable. Obviously, there are also other ontologies that might become available soon in the necessary format. For instance, the EPPO Global Database (https://gd.eppo.int/) 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.

Table 3
Existing ontologies reused in the PCSM.

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#

4.4. PCSM Design

In this section, the design of the PCSM is described focusing on the most common core concepts. The reuse of concepts of existing ontologies and their properties is also demonstrated.

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 2. As can be seen, the `rdfs:subClassOf` property is used to relate the Ploutos concepts with the concepts of reused existing ontologies. Instances of these Ploutos concepts directly inherit the properties of these reused concepts. However, where needed, additional properties are defined for Ploutos concepts, for 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 well defined by ENVO. For example, the `ploutos:Parcel` concept is an `rdfs:subClassOf`

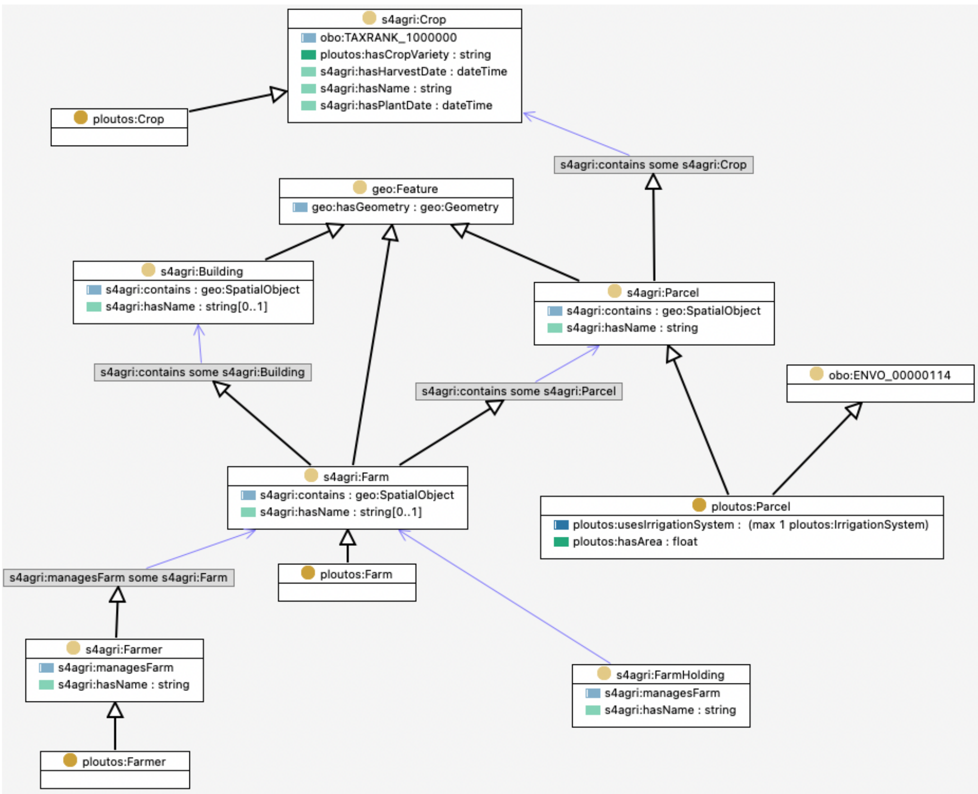

Fig. 2. OWL Diagram of the Ploutos concepts Farmer, Farm, Parcel and Crop.

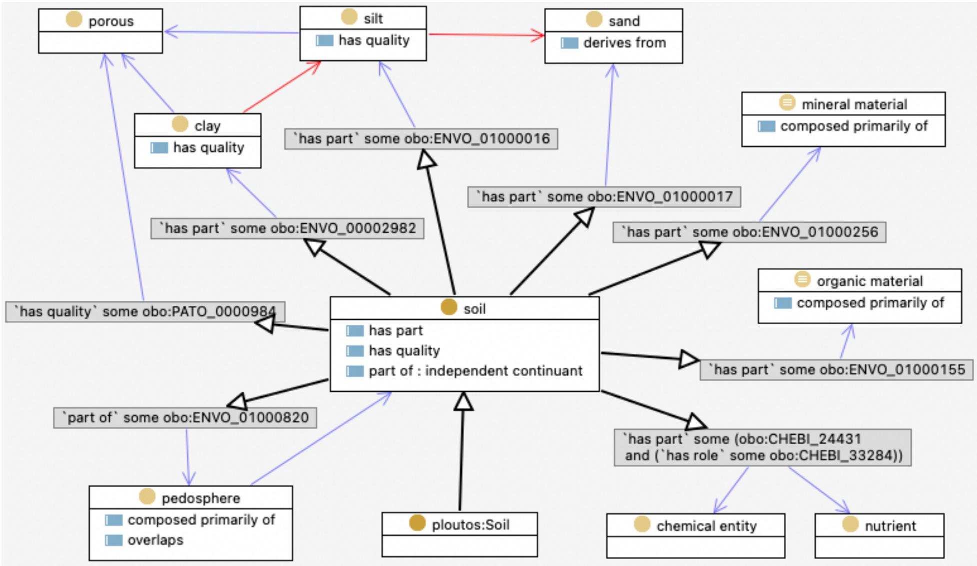

Fig. 3. OWL Diagram of the Ploutos concept Soil and reuse of ENVO concepts.

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".

Soil: In addition to the main common concepts farmer, farm, parcel and crop, the next most common concept is that of "soil" which is part of the parcel and in which the crop grows. For modelling the Ploutos concept `soil`, we reused a similar concept in the ENVO ontology, as shown in Figure 3. As can be seen, the rdfs:subClassOf property is used to relate the ploutos:Soil concept with the ENVO soil concept. Thereby, the Ploutos soil concept inherits the properties of the ENVO soil concept, e.g. the 'has part', 'has quality' and 'part of' properties.These properties show that soil can consist of various other materials, such as 'clay', 'silt', 'sand', 'minerals', 'organic', and that it may contain chemical entities and nutrients.

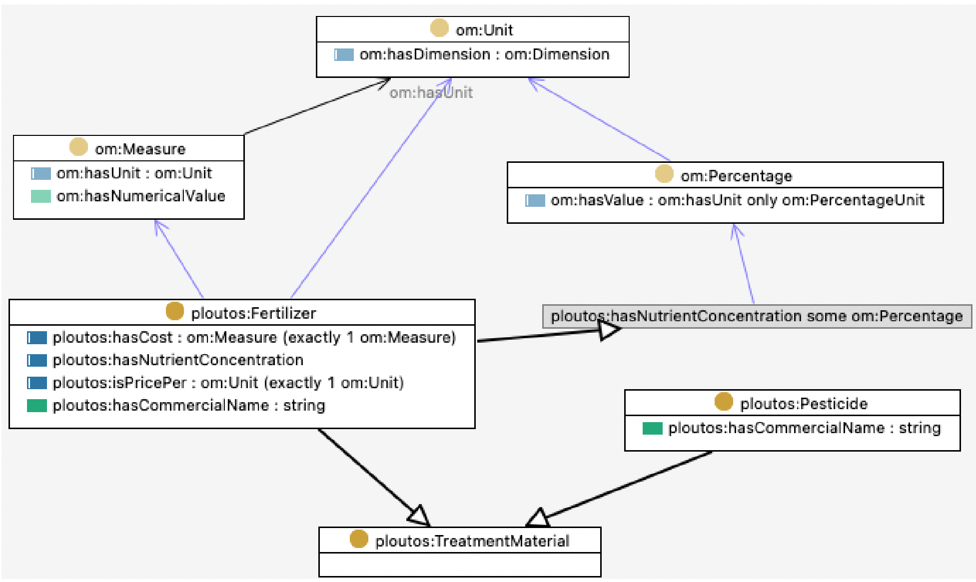

Fig. 4. OWL Diagram of the Ploutos concept TreatmentMaterial containing Fertilisers and Pesticides and reusing OM concepts

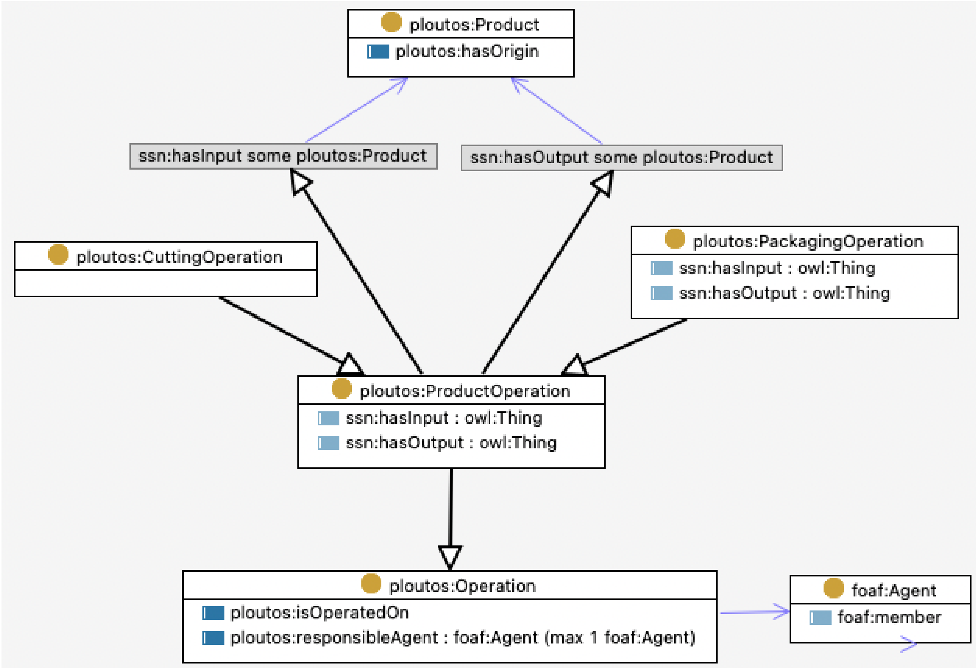

Fig. 5. OWL Diagram of the Ploutos concept Operation, ProductOperation and more specific operations.

Fertiliser and pesticide: Apart from water that is irrigated on the soil that is part of the parcel, also other materials and substances need to be applied to either to the soil or the crop on the parcel. In order to capture such materials, the class `ploutos:TreatmentMaterials` is introduced. This class contains the class of fertilisers with nutrients that are needed to keep the soil healthy as well as the class of pesticides that are used to protect the crop against diseases and dangerous insects. Figure 4 shows the OWL diagram of these classes and shows that the concentration of a nutrient in a certain fertiliser is modelled using the `om:Percentage` class, that the cost of a fertiliser is expressed in terms of an `om:Measure`. The `om:Measure` class contains a numerical value and a unit, in this case the amount in euros to be payed for a unit of the fertiliser.

Operation: In order to model action-related concepts, the class `ploutos:Operation` is introduced. This class can be specialised for various operations on the objects on or at the farm or further down in the agrifood supply chain. The two main properties of a `ploutos:Operation` are the `ploutos:isOperatedOn` property, that designates to the object on which the operation is performed, and the `ploutos:responsibleAgent` property, that indicates the person or organisation that performs the operation. **ProductOperation:** The `ploutos:Operation` class has various subclasses of more specific operations, e.g., the `ploutos:ProductOperation`. This class encapsulates all possible operations on a product produced by for instance the farmer, `ploutos:CuttingOperation`, or the food processing factory, `ploutos:PackagingOperation`. See Figure 5 for an OWL diagram of these classes and their properties. As can be seen, the class `ploutos:ProductOperation` also has an input and output product of the class `ploutos:Product` using the SSN properties `ssn:hasInput` and `ssn:hasOutput`. The `ploutos:Product` concept has additionally a property `ploutos:hasOrigin` that indicates the original farm at which the product is produced. **ParcelOperation:** Besides product operations, specific operations on parcels are distinguished using the class `ploutos:ParcelOperation`, see Figure 6. This class uses the `ploutos:isOperatedOn` property to indicate on which parcel the operation is performed. Furthermore, the class has a `datetime` on which it is started and a `datetime` on which it ended. Finally, the `ploutos:hasAppliedAmount` property indicates the amount of a certain treatment material or substance that is used on the parcel in the operation and the `ploutos:isAppliedPer` property indicates the area of the parcel that is treated with this amount of material or substance. Three different parcel operations are defined in the PCSM as subclasses of the `ploutos:ParcelOperation` concept.

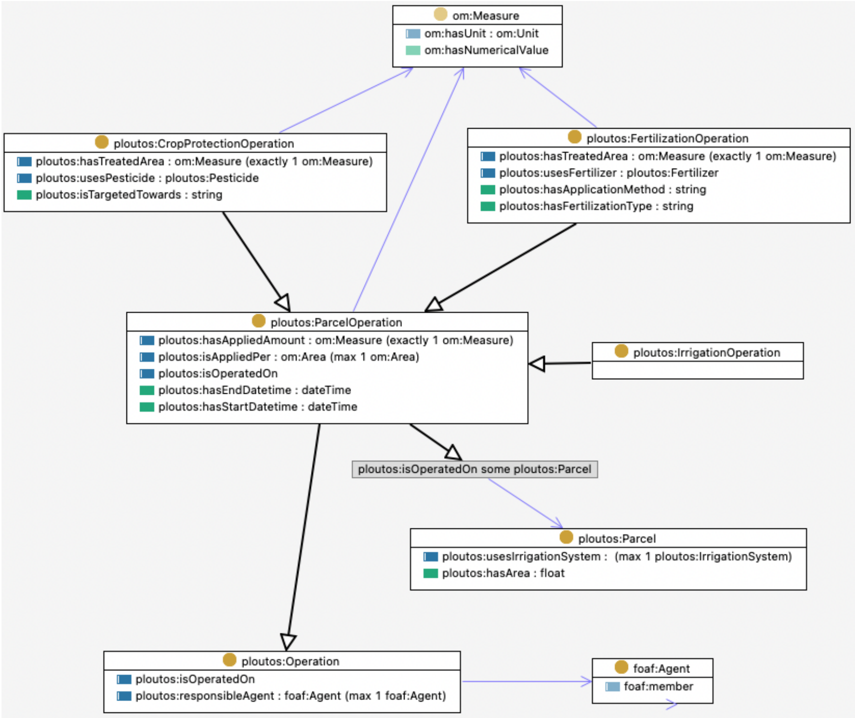

Fig. 6. OWL Diagram of the Ploutos concept Operation, ParcelOperation and more specific operations

1. First, the `ploutos:irrigationOperation` concept represents all operations for irrigating water on the parcel and does not have any additional properties apart from the generic properties of the `ploutos:parcelOperation`.
2. Second, the `ploutos:FertilizationOperation` concept represents all operations for fertilising the soil on the parcel. In addition to the generic properties of the parcel operation, this operation also defines the fertiliser used, `ploutos:usesFertilizer`, the type of fertilization, `ploutos:hasFertilizationType` (e.g., basal fertilization), the application method, `ploutos:hasApplicationMethod` (e.g., broadcasting), and the treated area, `ploutos:hasTreatedArea`, which is an `om:Measure` that indicates the amount of hectares on the parcel that is fertilised.
3. Third, the `ploutos:CropProtectionOperation` concept represents all operations for protecting the crop on the parcel. In addition to the generic properties of the parcel operation, this operation also defines the pesticide used, `ploutos:usesPesticide`, the object to which the pesticide is targeted towards, `ploutos:isTargetedTowards` (e.g., late blight), and the treated area, `ploutos:hasTreatedArea`, which is an om:Measure that indicates the amount of hectares on the parcel that is fertilised.

Possibly, more properties need to be added in the future or other types of parcel operations are common enough to be added to the PCSM.

Weather: The PCSM also contains weather properties that are required by most of the pilots to monitor the various environmental conditions around the crop on a parcel. To model this, the PCSM has the `ploutos:WeatherParameter` concept and reuses the weather properties defined in the BIMERR weather ontology (https://bimerr.iot.linkeddata.es/). The `ploutos:WeatherParameter` is therefore made equivalent to the `weather:WeatherProperty` concept in the BIMERR weather ontology using the `owl:equivalentClass` property. An advantage of using the equivalent class property here is that the `ploutos:WeatherParameter` automatically inherits all sub-classes of the `weather:WeatherProperty` class containing the actual parameters like Temperature, Humidity, Wind, Storm etc.

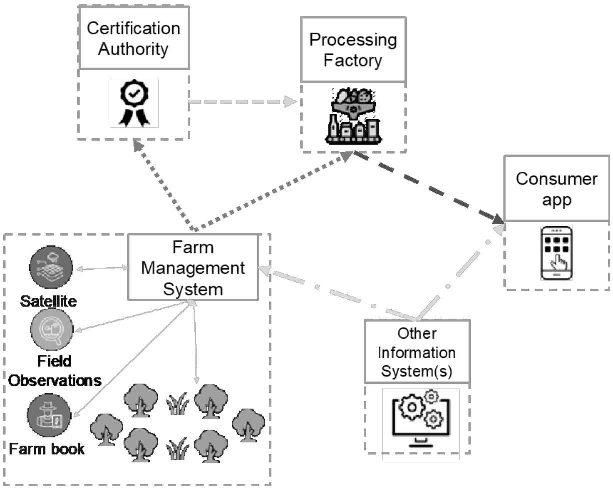

Fig. 7. Current practices for data sharing (if it occurs).

Summarising, the PCSM is a common semantic model that contains the main concepts and their relations for the agrifood domain with a focus on the farmer end. We reused a set of existing ontologies and their concepts to represent the Ploutos concepts. The PCSM is now being used for the pilots in Ploutos and specifically for the traceability scenario in one of the pilots. In the future, the PCSM will be extended with new concepts that are identified as being common for the agrifood supply chain as a whole. The full version of the PCSM is available from a Gitlab repository set-up for the Ploutos project (https://gitlab.com/Ploutos-project/).

5. The Ploutos Architecture

The PSCM that was detailed in the previous sections aims to address the semantic interoperability challenges of the agrifood sector. In the following sections, the PSCM is used along with data management mechanisms in order to facilitate data sharing among the key players. With the digitisation of farm practices and decision support systems (e.g. Farm Management Information Systems), 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 for bench marking or productivity comparisons. 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 7).

5.1. Technical Requirements

This section specifies the core technical requirements that are related to the data collection and data management functionalities that are necessary for the realisation of the PLOUTOS pilots. The following requirements, shown in Table 4 have been extracted based on the analysis of the SIPs combined with the specified high level design principles specified in section 1 of this article. The technical/software requirements drove the design and implementation of the actual functional component of the Ploutos data-sharing framework (cf. [43] for more details)

5.2. Ploutos Data Sharing Architecture

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

Table 4
Technical requirements for the Ploutos architecture

ID	Title	Description
1	Ensure Semantic Interoperability of heterogeneous data items	Heterogeneous data items should be modelled based on if possible standarised data modelling approaches. Use of core ontologies in combination to domain specific ontologies allows the extension of the data model capabilities.
2	Ensure Syntactic Interoperability with various data sources.	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.
4	Support seamless integration with legacy information sources	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 format (e.g. OWL).
5	Ensure interoperability in data provision by legacy data sources	PLOUTOS platform will provide a common query language interface via an interoperable/standarised API that will permit permit data to be retrieved from heterogeneous databases and data sources. The service API should be able to direct the query to a specific database or source, using a query languages like SPARQL (able to provide integrated view over different datasets) or SQL syntax. The API should return 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 format for results.
6	Asynchronous exchange of data	The system should be able to operate in Asynchronous mode. Asynchronous operation is realized based on notification mechanisms to register data consumers for specific events. Often a publish/subscribe process is followed.
7	Support of enhanced descriptions for data items	The data modeling approach should be extendable and to foresee the assignment of additional descriptions in the form of metadata.
8	Make data discoverable/findable	Data provision mechanisms should ensure that the various data items/ data sets will be discoverable with the use of querying mechanisms. This requirement is aligned with the FAIR principles. The use of Linked Data mechanisms will support this functionality.
9	Ensure availability of datasets	The datasets/data items should be available for retrieval through the use of standardised well documented APIs along with robust data provision mechanisms.
10	Ensure confidentiality and integrity of datasets	The confidentiality and integrity of datasets/data items should be ensured through the use of information security mechanisms (e.g. access control, cryptography).
11	Ensure ethical data use	The use made of data sets/data items should conform with current (legal/ethical) regulations. (Code of Conduct, GDPR). This requirement will be addresses via the implementation of both technical and administrative/regulatory means.
12	The data provider must control the data sharing mechanisms and policies	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.

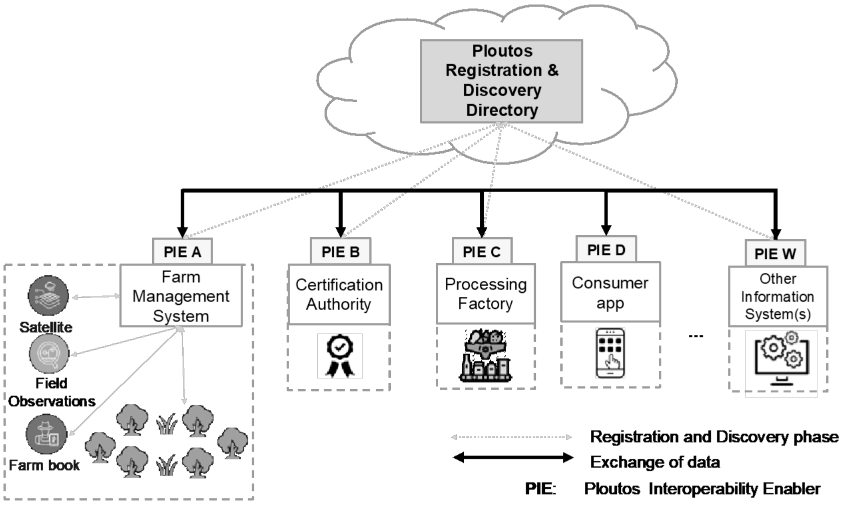

Fig. 8. Conceptual view of Ploutos data sharing approach

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 8 each data source or sink 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 i.e. only the accessibility properties (like hostname and port). Operational properties and capabilities are communicated peer-to-peer. The functionalities of these two core components are elaborated at the following sections.

5.3. Ploutos Interoperability Enable (PIE)

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 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 9 and analysed in the following paragraphs.

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, time span of data sets). This registration process occurs in an automated and periodical manner. A. PIEs capabilities are distributed directly to other PIEs through a peer-to-peer protocol. The Registration Client periodically contacts the "Ploutos Registry and Discovery Directory" (PRDD) to provide notifications on potential updates on PIEs capabilities. The registration process specifies how exchange of data and knowledge will be realised. For example, the following questions are answered during the registration phase of a PIE:

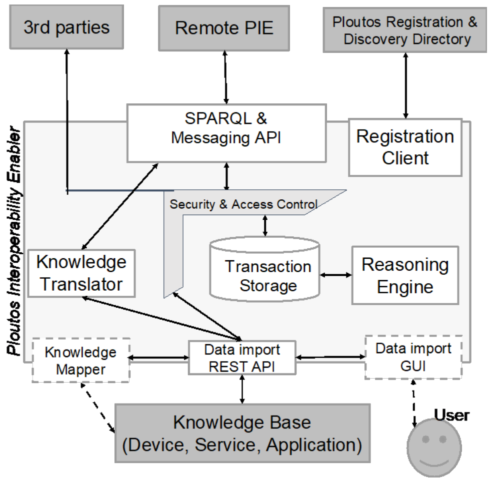

Fig. 9. Functional component diagram of the Ploutos Interoperability Enabler (PIE)

- 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 weather station app might subscribe to knowledge about temperature measurements in a field, or request the current temperature. It might also publish current temperature preferences of a user or a specific crop. A FMIS system might subscribe to both the knowledge about temperature preferences of a crop and temperature measurements to be able to optimally recommend interventions such as irrigation.

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 a Developer REST API using the basic graph pattern elements of SPARQL.

Data import REST API: This component provides the main mechanisms for the PIE to exchange data sets from the underlying system on which it 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 to exchange data with the PIE. However, given the expected heterogeneity of the systems that the PIE will interact additional mechanisms are provided to facilitate the connection with these systems. In case the "Knowledge Base" already provides each own API the optional component "Knowledge Mapper" is used which acts as an intermediate among the "Data Import REST API" and the API of the underlying system. In case the "Knowledge Base" does not 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.

Knowledge Mapper: This enables 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 pilots the translation will be realised according to the concepts and properties of the Ploutos Common Semantic Model (PCSM).

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 the PIE will be based on existing best practices ensuring Confidentiality, Integrity and Availability of data. Moreover, all the interaction between PIEs and between a PIE and its Knowledge Base is supported by secure communication protocols, like SSH and HTTPS. In addition, in the current implementation, a role-based access control mechanism is used to enforce access policies on parts of the data provided by a knowledge base. A PIE capability, which expresses the piece of knowledge or data is provided is defined as the unit of access control. In addition, each knowledge base is owned by an actor or participant that has a certain role in the knowledge network. Access policies are then defined per role as the permission to access a PIEs capability granted to that role. These access policies are maintained by the PRDD and used by the PIEs to enforce them during data exchange phases, e.g., when a remote data query is received by a PIE. The PIE will provide the means for enforcing the dictated authentication, access control and data governance policies during data exchange.

Reasoning engine: Finally, the PIE contains a reasoning engine that supports two types of reasoning 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 described in [44], this is a standard process for the semantic web and a plethora of reasoners are available to be reused. Inferring new data by reasoning is employed in selected cases where the enrichment of the knowledge base is considered useful for the needs of a SIP. The second case of reasoning refers to the orchestration of the process of requesting data from or providing data to knowledge bases and their PIEs in the entire knowledge network. Using this reasoning process, the various PIEs interact with each other and use the Ploutos Registration and Discovery Directory to discover the different knowledge bases.

5.4. 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 to serve one query. In addition, 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.
- Changes in the knowledge network are handled seamlessly by synchronising information about knowledge interactions.
- Established open-source Semantic Web technologies are leveraged to provide knowledge models and reasoning capabilities.

Figure 10 shows the functional components and the potential interactions of the Ploutos Registration and Discovery Directory.

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 authorised 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 if 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. Full access control is still under development but will employ a combination of standard security mechanisms such as rule-based access control and cryptography for protecting the registry of PIEs.

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). Other aspects such as a) the capabilities that it requires from other PIEs, b) the capabilities that it offers, and c) the necessary access policies are communicated peer-to-peer with other PIEs.

6. Worked example

The data sharing solution described above has been implemented, deployed and evaluated in the context of Pilot 1 (cf. Section 3). The core objective of this pilot has been to address the existing data related requirements of a frozen fruit value chain. The two major objectives of this pilot have been:

- Support small farmers in optimising the cultivation practices through a data driven smart farming solution (FMIS).
- Increase production transparency and reduce the administrative burden through data sharing mechanisms.

It should be noted that the deployment and use of the proposed data sharing mechanisms took place in real-world operational environment involving the integration of production systems.

6.1. Demonstration steps

In this section a step-by-step analysis of the demonstrator is presented. The demonstration took place during one cultivation season March - September 2021.

Step 1- Cultivation of peaches with the use of Smart Farming services: In order to improve the cultivation practices and to provide evidence of the applied farming practices the “gaiasense” Smart Farming (FMIS) solution was used (https://www.neuropublic.gr/en/smart-farming-gaiasense/) [45]. The gaiasense solution supports the farmers with data-driven advice on fertilisation, pest management and irrigation based on the collection of data from various sources including IoT-enabled agro-enviromental stations, satellite derived data, the farmer’s digital calendar (farm book) and on-the-field observations (scouting) of the cultivation. All the aforementioned data are collected in a central cloud repository where they are stored, integrated and processed using data analytic techniques. The outcome of the processing, the proposed recommended actions, is confirmed by experts (i.e. agronomists) in order to generate farming advice focusing on the optimisation of inputs (irrigation, pest management, and fertilisation) tailored to the needs of the targeted parcel’s conditions. The actual farming practices that were applied as a response to the advice is then fed back to the gaiasense system and recorded in the farm book. Figure 11 provides a screen-

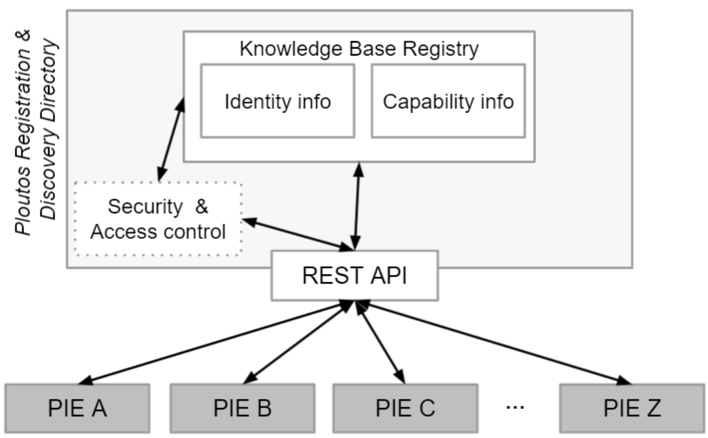

Fig. 10. Functional component diagram of the Ploutos Registration and Discovery Directory

shot of the gaisasense dashboard developed for the needs of Ploutos. The dashboard presents data referring to the selected peach orchard including elements from the farm book (e.g. growth stages, applied pesticides, irrigation), environmental measurements derived by the sensors network and calculated pest infestation risk indexes. Farmers and advisers have access to their respective data collections through web-based dashboards.

The harvesting period for the peaches in this area starts at the end of the summer until beginning of autumn. The farmers in the pilot are collecting the peaches into bins where each bin is assigned a unique identifier expressed with a barcode. The respective records including timestamp, farm-id, farmer-id, bin-id, and each bin's weight are recorded in the farm book and are accessible through the gaiasense FMIS API [1].

The data items that have been identified as relevant with the fruit supply chain and are available through the FMIS webservice include the following:

- Farm's details: Country, area, town, farm size, coexisting cultivated crops in the farm, scale/category (e.g. small family farm or industrial agriculture), employees data, Farm's identifier/name
- Farmer's details: Name, contact details, Farmer's identifier
- Farmers' Cooperative details: Contact details, Number of farmers, financial information, other activities of the cooperative.
- Peach orchard: Parcel's size, parcel's polygon, type of cultivated peaches
- Cultivation details: dates of reached phenological stages (first leaves, blossoming, fruiting, harvest date)
- Farming Operations: Recording of application of pesticides (date, type, dose, target), fertilisers (date, type, dose), irrigation (date, dose)
- Harvesting operation: Date, Farm name (Id), Farmer name (id), Bin id

Step 2- At the processing factory: The bins with the harvested peaches are brought to the farmers' association warehouse and then transferred to the food processing factory. The processing factory is called Alterra (https://www.alterra.gr/index.php/en/) and is located in the same area as the pilot farms and serves the producers for the whole region. Bins received by the processing factory are registered with the factory's Warehouse Management System (WMS) along with the date-time of receipt, peaches type and origin (farmer-id and farm-id). Peaches are then fed to processing lines where they are washed, peeled, sliced, frozen and finally packed into plastic bags that are placed in cardboard boxes of 10-15kg each. Cardboard boxes produced in the same lot are assigned the same lot-id which is

[1] See https://gitlab.com/Ploutos-project for access details.

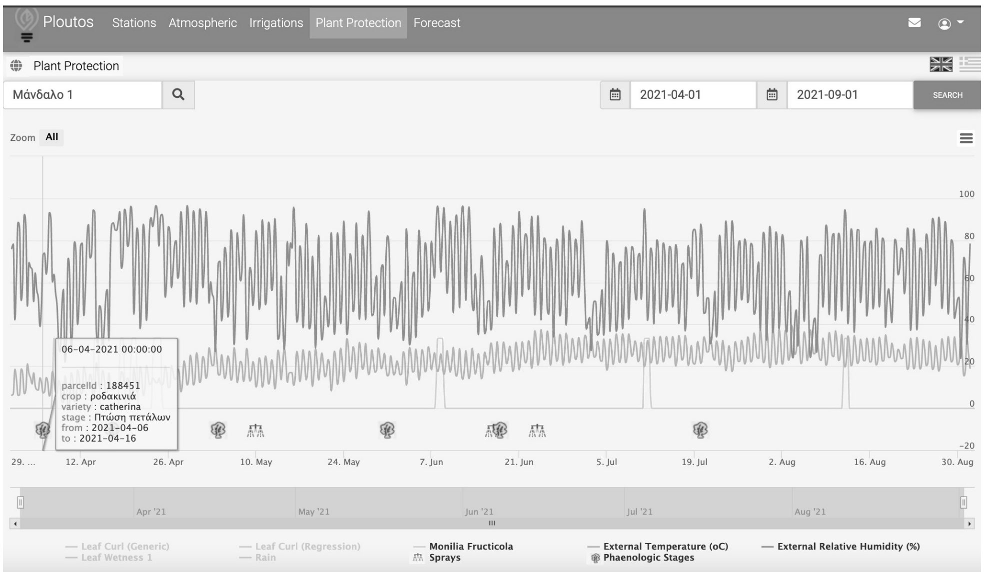

Fig. 11. The gaisense web-based dashboard presents data concerning to cultivation practices

expressed again through a barcode and alphanumeric format. The factory's WMS maintains a registry of the starting and ending times of the process along with the bin-id of the raw peaches and the lot-id of the final product. It should be noted that the specific WMS was not exposing an API but it was feasible to export all necessary records in CSV format. In order to enable connectivity a simple webservice (REST API) was developed that uses as import the CSV file and provides access to data through specific calls.

Relevant data items derived from the fruit processing factory are the following:

- Date of delivery of peaches at the factory
- Location of factory (especially with regards to the farm in order to deduce how far the raw peaches have travelled)
- Processing stages
- Date of packaging
- Other profile details about the factory

All these relevant data items are capable of representation in the PCSM whose design is described in Section 4 above.

Step 3-Second stage of processing: Palettes with cardboard boxes are exported from Greece to various European countries. The recipients of the processed peaches from the pilot farms are mainly big industrial food processing/production companies that further process frozen peaches in order to produce food products (e.g. fruit yogurt). Food processing companies are eager to get more details on the origins, the applied cultivation practices and the processing states of the purchased frozen peaches. Currently various certification schemes are followed (e.g. GlobalGAP) that aim to ensure the high quality of the products but there is a need for additional data that can provide evidence of the agronomic practices applied. For the needs of the Ploutos pilot, a web-based application for mobile devices has been developed. The industrial consumers of the processed fruits can scan the bar code or type the lot-id and according to the access rights grant can retrieve additional information with regards to the cultivation practices and first-stage processing practices. The query is executed through a distribution of queries to registered systems as described in Section 5. The appropriate information describing the processing steps referring to the fruits in the carton box is retrieved and displayed to the end-user's device.

6.2. Operation of data sharing

Figure 12 shows the key information providers and the interactions between them. The main information providers in this demo are the "Smart Farming Service" (gaiasense.gr) and the "Fruit Processing Factory" (alterra.gr). For each of these providers, a Ploutos Interoperability Enabler (PIE) was developed enabling the exchange of information with other Ploutos Interoperability Enablers in the network. Each information systems also needs a customised Knowledge-Mapper that enables the interaction with each PIE. The "Knowledge Mapper" enables interoperability at a semantic level through a data translation service that converts data streams provided by the hosting system to the PCSM and vice-versa. The translation functionality of the "Knowledge Mapper" must be adapted according to the custom data model and API of the hosting system. In general, the "Knowledge Mapper" acts as a translator and API mediator customised to the information system's specifications that it is deployed at.

The Ploutos Registration and Discovery Directory (PRDD) is deployed at a cloud server system accessible through the Internet supporting the overall orchestration of knowledge discovery. The PRDD enables the registration and discovery of the various PIEs along with the information items that they can support. For example, given a specification of the knowledge item that is requested a PIE through the PRDD can figure out the appropriate knowledge base to access it. 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.
- Changes in the knowledge network are handled seamlessly by synchronizing information about knowledge interactions.

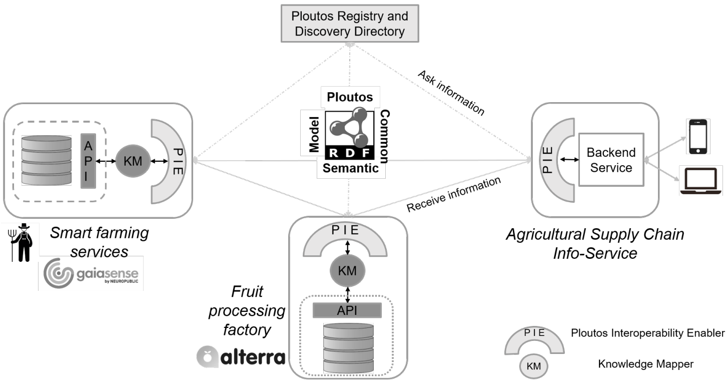

Fig. 12. Conceptual representation of key information providers and interactions of the agri-food data sharing demonstrator.

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

6.3. *Traceability information retrieval scenarios*

The information retrieval through the agricultural supply chain of this demonstrator is initiated with the issuing of a LOT-ID based query with the use of the "Ploutos Traceability app". The overall process is realised in two stages:

a. In the first stage, the supply chain is traced backwards to get an overview of the operations that have been performed and the respective identifiers of the participated stakeholders.
b. In the second stage, more detailed information from a stakeholder or a specific operation is retrieved and presented upon request by the end-user.

Retrieval of overview of processing steps:

In the context of the realised demonstration, the message sequence diagram in Figure 13 corresponds to the first stage interactions between the end user's traceability app, the warehouse management system of the fruit processing factory (Alterra) and the Smart Farming (gaiasense) system. The following sequence of steps are identified:

1. The traceability app receives a LOT-ID from the end-user or the GUI for which an overview of the operations and stakeholders is requested.
2. The traceability app performs a request on its PIE with a graph pattern that expresses an ASK for operations that have directly led to the end-product identified by the LOT-ID. The following graph pattern is defined and used for the knowledge interactions between the PIEs to get an overview of the operations in the supply chain.

```
?endProduct rdf:type ploutos:Product .
?output ploutos:intermediateProductOf ?endProduct .
?operation rdf:type ?operationType .
?operationType rdfs:subClassOf ploutos:ProductOperation .
?operation ploutos:hasResponsibleAgent ?agent .
?agent s4agri:hasName ?name .
?operation ssn:hasInput ?input .
?operation ssn:hasOutput ?output .
?operation ploutos:hasEndDatetime ?timestamp .
```

3. The reasoner in the traceability app PIE uses the PRDD and discovers that the Alterra PIE can provide this information. It then makes a request to the Alterra PIE to ANSWER with its bindings to this graph pattern.
4. The Alterra PIE uses its KM to call the API of the Warehouse Management System to retrieve the list of operations that have led to the LOT-ID together with the bin_ids that were used as input to the LOT-ID. This will usually be a packaging operation where fruit in a bin_id is checked and put into a LOT-ID.

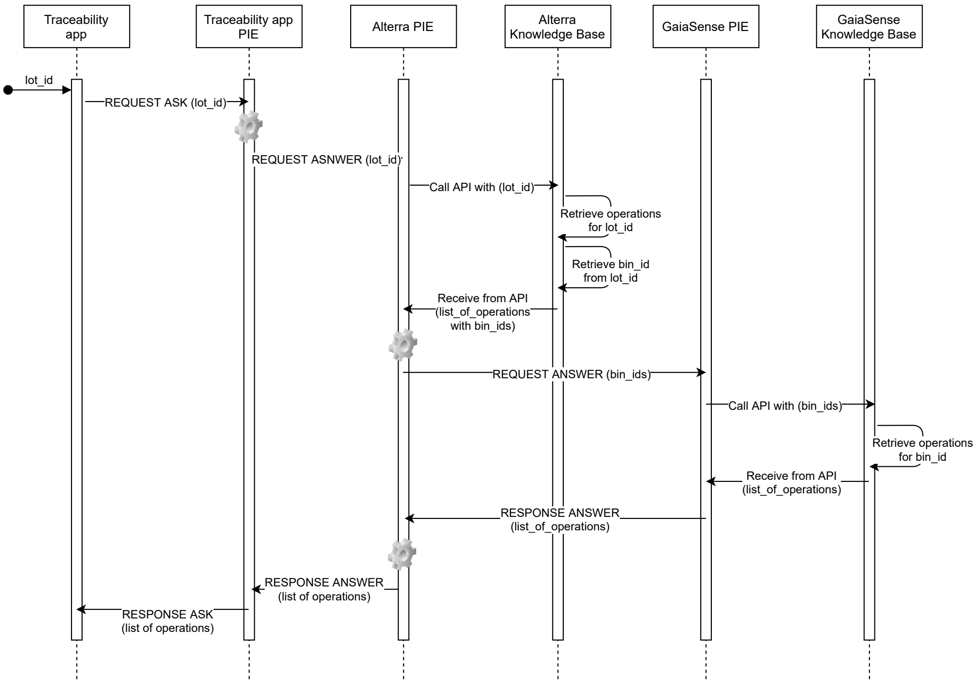

Fig. 13. Sequence diagram of tracing back into food-chain to retrieve operation information.

5. Upon receipt of the operations with bin_ids, the reasoner of the Alterra PIE uses the PRDD again to discover that the gaiasense PIE can provide further information on the operations that have led to the products identified by the bin_ids.
6. The gaiasense PIE uses its KM to call the gaiasense API to retrieve, for each bin_id, the operations that have led to this bin_id. This will usually only be a harvesting operation.
7. The gaiasense PIE will send a response back to the Alterra PIE with the resulting list of operations at the farm.
8. The reasoner of the Alterra PIE will combine the list of operations that it received from the gaiasense PIE with the list of operations that it received from the API of the Alterra Warehouse Management System into a complete list of operations.
9. The Alterra PIE will send a response back to the traceability app PIE with the complete list of operations which will then be responded back to the traceability app.

The result of this sequence of steps is thus a list of “operations” while the actual reply is expressed in JSON format. A reply example follows:


```
[ {
   "endProduct": "<http://alterra.gr/lots/130821A-L4-181011>",
   "operation": "<http://alterra.gr/operations/5488>",
   "input": "<http://alterra.gr/bins/65499>",
   "output": "<http://alterra.gr/lots/130821A-L4-181011>",
   "operationType": "<https://www.tno.nl/agrifood/ontology/ploutos/common#
      PackagingOperation>",
   "agent": "<http://alterra.gr/alterra-org>",
   "agentName": "\"Alterra\"",
   "timestamp": "\"2021-08-24T12:34:56Z\""
  },
```

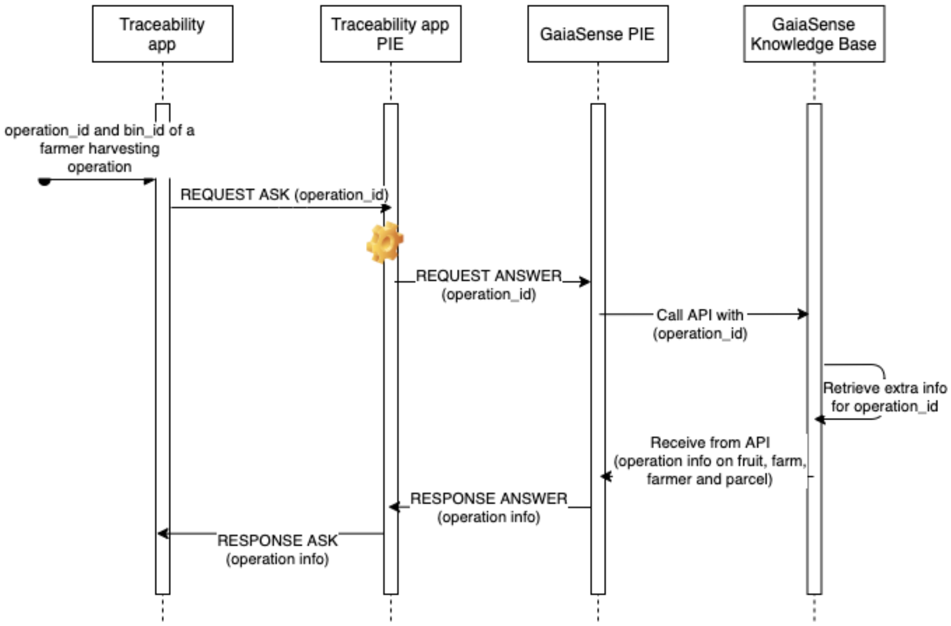

Fig. 14. Sequence diagram for tracing back to retrieve farm and farming information


```
{
 "endProduct": "<http://alterra.gr/lots/130821A-L4-181011>",
 "operation": "<http://sense-web.neuropublic.gr:8585/operations/1337>",
 "input": "<http://sense-web.neuropublic.gr:8585/crops/23904891289>",
 "output": "<http://alterra.gr/bins/65499>",
 "operationType": "<https://www.tno.nl/agrifood/ontology/ploutos/common#
    HarvestingOperation>",
 "agent": "<http://sense-web.neuropublic.gr:8585/farms/7878>",
 "agentName": "\"Mandalo Farmers Association \"",
 "timestamp": "\"2021-08-22T10:22:00Z\""
}]
```


The reply contains two main objects. The first part refers to the 'creation' of product <http://alterra.gr/lots/130821A-L4-181011>. The corresponding operation was of type "PackagingOperation" that happened on "2021-08-24T12:34:56Z" where the responsible organisation for this operation was Alterra. This operation belongs to a chain that results in the end product <http://alterra.gr/lots/130821A-L4-181011>, having as input product of this operation the <http://alterra.gr/bins/65499> and output <http://alterra.gr/lots/130821A-L4-181011>. Given that there is a match between the output and the final product this also denotes that it is the end product of this operation. The second part also refers to the creation of product <http://alterra.gr/lots/130821A-L4-181011>. The corresponding operation was of type "Harvesting" that happened on "2021-08-22T10:22:00Z" where responsible for this operation was the "Mandalo Farmers Association". This operation belongs to a chain that results in the end product <http://alterra.gr/lots/130821A-L4-181011>, having as input product of this operation the <http://sense-web.neuropublic.gr:8585/crops/23904891289> and output <http://alterra.gr/bins/65499>.

Retrieving information of a specific processing step – Farm and farming practices

The sequence diagram in Figure 14 shows the interactions among the Ploutos traceability application and the Smart Farming service for the retrieval of additional about the farm and the applied farming practices.

The list of data items of interest that are modelled and retrieved for the needs of the traceability demonstration are the following:

– Origin of Peaches - Type and name of peaches.
– Farm's details - farm name, location and parcel size.
– Cooperative details - name of cooperative that the farmer belongs
– Cultivation details- dates of reached phenological stages (first leaves, blossoming, fruiting, harvest date)
– Harvesting operation: Date, Farm name (Id), Farmer name (id), Bin id

To retrieve this information, the following graph pattern is defined and used for the knowledge interaction between the two PIEs.

```
?operation rdf:type ploutos:HarvestingOperation .
?operation ssn:hasOutput ?bin .
?operation ploutos:isOperatedOn ?parcel .
?parcel s4agri:contains ?crop .
?crop rdf:type ?cropType .
?cropType rdfs:label ?cropName .
?parcel ploutos:hasArea ?area .
?parcel ploutos:hasFirstLeavesDate ?leavesDate .
?parcel ploutos:hasBlossomingDate ?blossomDate .
?parcel ploutos:hasFruitingDate ?fruitingDate .
?parcel s4agri:hasHarvestDate ?harvestDate .
?farm rdf:type ploutos:Farm .
?farm wgs84:location ?location .
?location wgs84:lat ?latitude .
?location wgs84:long ?longitude .
?farm s4agri:hasName ?farmName
?farm s4agri:contains ?parcel .
?farmer s4agri:managesFarm ?farm .
?association rdf:type ploutos:FarmAssociation .
?farmer s4agri:isMemberOf ?association .
?association s4agri:hasName ?associationName.
```

The scoping on the `HarvestingOperation` is done because this is the final operation of the farmer that has as output the harvested peaches put into a bin_id. In addition, when the request is done, the `?operation` variable is bound to the value of the harvesting operation_id and the `?bin` variable is bound to one of the bin_ids that have led to the original LOT-ID. The traceability app PIE then has the capability that the traceability app can ASK for knowledge in terms of this graph pattern and the gaiasense PIE can ANSWER about their knowledge in terms of this graph pattern. Bearing this in mind, the sequence diagram can be explained as the following sequence of steps:

1. The traceability app receives a request from the end-user or the GUI to get more information about a specific operation with one of the bin_ids as output, in this case a harvesting operation at the farm that produced this bin_id.
2. The traceability app performs a request on its PIE with a graph pattern that expresses an ASK for the list of crop, parcel, farm and farmer information responsible for this operation.
3. The reasoner in the traceability app PIE uses the PRDD to discover that the gaiasense PIE can provide this information and makes a request to the gaiasense PIE to ANSWER with its bindings to this graph pattern.
4. The gaiasense PIE uses its KM to call the gaiasense API to retrieve the requested crop, parcel, farm and farmer information for the harvesting operation of the peaches of the bin_id.
5. Upon receipt of this information, the gaiasense PIE will send a response back to the traceability app PIE with the information which will then be responded back to the traceability app.

The result of this sequence of steps is thus a set of bindings for the harvesting operation, crop, parcel, farm and farmer information. In JSON format this looks like the following list:


```
[{
   "operation": "<http://sense-web.neuropublic.gr:8585/operations/1337>",
   "bin": "<http://sense-web.neuropublic.gr:8585/bins/65499>",
   "parcel": "<http://sense-web.neuropublic.gr:8585/locations/434>",
   "cropType": "<http://purl.obolibrary.org/obo/NCBITaxon_3760>",
   "cropName": "peach tree",
   "area": "97 hectare",
   "leavesDate": "\"2021-04-22T00:00:00Z\"",
   "blossomDate": "\"2021-05-22T00:00:00Z\"",
   "fruitingDate": "\"2021-06-22T00:00:00Z\"",
   "harvestDate": "\"2021-08-22T00:00:00Z\"",
   "farm": "<http://sense-web.neuropublic.gr:8585/farms/7878>",
   "farmName": "\"Happy Peaches Farm Ltd.\"",
   "latitude": "22.2170236",
   "longitude": "40.7856276"
   "associationName": "\"Mandalo-Proodos\"",
 }]
```


In addition, a list of data items of interest to be retrieved for giving insight into the operations on the parcel, soil and plant are the following:

- Farming Operations: Recording of application of pesticides (date, type, dose, target), fertilisers (date, type, dose), irrigation (date, dose)

To retrieve this information, the following graph pattern is defined and used for the knowledge interaction between the two PIEs.

```
?operation rdf:type ?parcelOperationType .
?parcelOperationType rdfs:subClassOf ploutos:ParcelOperation .
?operation ploutos:isOperatedOn ?parcel .
?operation ploutos:hasAppliedAmount ?applied .
?applied om:hasUnit ?appliedUnit .
?applied om:hasNumericalValue ?appliedValue .
?operation ploutos:isAppliedPer ?area .
?area om:hasUnit ?areaUnit .
?area om:hasNumericalValue ?areaValue .
?operation ploutos:hasStartDatetime ?start .
?operation ploutos:hasEndDatetime ?end .
?operation ploutos:responsibleAgent ?agent .
?operation ploutos:usesMaterial ?material .
?material ploutos:hasActiveSubstance ?activeSubstance .
```

In a similar way, a sequence of interaction steps between the PIEs leads to a set of bindings for the parcel operation with more detailed information on fertilizers, pesticides and irrigation. In JSON format this looks like the following list:


```
[{
      "areaUnit": "<http://www.ontology-of-units-of-measure.org/resource/om
         -2/squareMetre>",
```

```
        "area": "<http://sense-web.neuropublic.gr:8585/.well-known/genid/35146
           dde-efe6-4885-bcfa-28fb67fde45a>",
        "agent": "<http://sense-web.neuropublic.gr:8585/agents/
           f7b0c242bebc26a65ccaf033b7798298>",
        "parcel": "<http://sense-web.neuropublic.gr:8585/parcels/188451>",
        "parcelOperationType": "<https://www.tno.nl/agrifood/ontology/ploutos/
           common#IrrigationOperation>",
        "applied": "<http://sense-web.neuropublic.gr:8585/.well-known/genid
           /471f75bc-a8b8-4f2a-84e3-3dabf4fa5785>",
        "start": "\"2021-06-04T18:00:00\"^^<http://www.w3.org/2001/XMLSchema#
           dateTime>",
        "appliedUnit": "<http://www.ontology-of-units-of-measure.org/resource/
           om-2/cubicMetre>",
        "appliedValue": "241.26663906946777",
        "material": "<https://www.tno.nl/agrifood/ontology/ploutos/common#
           water>",
        "end": "\"2021-06-04T22:00:00\"^^<http://www.w3.org/2001/XMLSchema#
           dateTime>",
        "activeSubstance": "\"Water\"",
        "areaValue": "4021.1106511577964",
        "operation": "<http://sense-web.neuropublic.gr:8585/.well-known/genid
           /8cf6ddf2-6e25-44b6-bcc8-c852a3017cad>"
      },
      {
        "areaUnit": "<http://www.ontology-of-units-of-measure.org/resource/om
           -2/squareMetre>",
        "area": "<http://sense-web.neuropublic.gr:8585/.well-known/genid/
           c41db423-febe-4456-b7a9-075d52182457>",
        "agent": "<http://sense-web.neuropublic.gr:8585/agents/
           f7b0c242bebc26a65ccaf033b7798298>",
        "parcel": "<http://sense-web.neuropublic.gr:8585/parcels/188451>",
        "parcelOperationType": "<https://www.tno.nl/agrifood/ontology/ploutos/
           common#CropProtectionOperation>",
        "applied": "<http://sense-web.neuropublic.gr:8585/.well-known/genid/
           dc75a7d4-14e3-4374-b2eb-9345a47fbb5b>",
        "start": "\"2021-03-12\"^^<http://www.w3.org/2001/XMLSchema#date>",
        "appliedUnit": "<http://www.ontology-of-units-of-measure.org/resource/
           om-2/gram>",
        "appliedValue": "60.0",
        "material": "<http://sense-web.neuropublic.gr:8585/.well-known/genid
           /28a58e3cfaa3623888fe64c6da3192b4>",
        "end": "\"2021-03-12\"^^<http://www.w3.org/2001/XMLSchema#date>",
        "activeSubstance": "\"Dithianon\"",
        "areaValue": "4021.1106511577964",
        "operation": "<http://sense-web.neuropublic.gr:8585/.well-known/genid
           /9bb44bce-3541-4dba-a767-191433374c37>"
      }
    ]
```


A similar approach is taken for retrieving additional data from the food processing factory and for potentially any other stakeholder connected via a PIE.

7. Discussion

The Ploutos data sharing platform enables open and semantically standardised interoperability between existing systems that keep their data at the source. In addition, our approach enables transparent but secure traceability backward up the chain to obtain sustainability information by the retailer or consumer. For the peach example in the previous section, we have worked out an implementation of the Ploutos data sharing and traceability architecture for the agrifood chain from peach farmer via fruit processor to the food company and consumer. For this purpose, we have developed knowledge mappers and PIEs as well as a traceability app as shown in Figure 12. A key achievement has been to enable semantic/syntactic interoperability at a high level (e.g. using formal semantic expressions) with existing operational systems without requesting from them to change their mode of operation at all.

7.1. Feasibility

We used the PCSM to define graph patterns for the knowledge interaction between the traceability app and the farmer's and fruit processor's systems. One of the graph patterns expresses the chain of operations that the peaches go through from harvesting via processing to final packaging and consumption. In addition, a graph pattern is used that expresses more specific sustainability information from the farm, including the possible use of fertilisers and pesticides. This graph pattern at the farmer's side is made only accessible to the fruit processing company and not to the consumer as this is too much detail.

The implemented PIEs automatically discover and connect to each other based on the knowledge interactions specified. Thereby, the collection of PIEs form a knowledge network that has the advantage that it releases the partners in the chain from implementing 1-to-1 connections with each different partner. The knowledge mappers are used to transform specific information available via an API at the farmer and fruit processor to the PCSM graph patterns. This is the only implementation activity in applying the Ploutos data sharing platform to this use case. A knowledge mapper also has a role based access control mechanism that permits or denies access to specific graph patterns of the PCSM. As a result, the consumer cannot get access to specific parcel operations on the farm, while the fruit processing company can. As a result, the entire traceability scenario shows that it is feasible with limited effort to apply the Ploutos data sharing platform to provide sustainability information via the Ploutos traceability app.

7.2. Performance

From a scalability perspective, the peach traceability example is fairly small. It only considers 3 partners with a few knowledge interactions. When scaling up to multiple users of the Ploutos app that request information about a product, the knowledge network should be efficient enough to handle them in parallel. Caching might be one of the approaches here, which is already part of the knowledge mapper implementation. How to increase the level of parallelism and caching in the knowledge network itself is future work.

With respect to response time, a few aspects need to be considered. First, it is important to find the optimum between user actions in the Ploutos app and the number of times knowledge is being collected via the knowledge network. This is fine-tuning work for each application. For example, upon start-up the app can collect all the information that it might show to the user at once , but this might take unacceptably long. An alternative is to only collect knowledge when necessary, e.g. when clicking a button or changing to another tab. Thereby, the total response time is divided over multiple different knowledge interactions.

Second, the number of calls from a Knowledge Mapper to the API of the corresponding knowledge base needs to be limited. In our scenario, we encountered that a lot (pallet) with peaches can consist of peaches coming from 26 bins from the farm. In order to get more info on each of these bins from the farm, 26 calls had to be made to the API. This adds up to the total response time of the traceability app backwards into the supply chain. Splitting up the knowledge interactions might be a way to deal with that, which is an application specific fine-tuning step.

Finally, the number of PIEs (and thus knowledge bases) connected to the knowledge network is a factor that determines the response time. At the moment, a PIE checks upon request of a knowledge interaction which other PIEs exactly match with this knowledge interaction. Subsequently, only matching PIEs will further process the request. As a result, the amount of interactions in the knowledge network is small and the response time depends linearly on the amount of PIEs.

8. Conclusions and Future Work

There has been growing pressure to address the environmental impact of the agrifood sector and with this the recognition that financial, social and environmental dimensions all need to be considered when trying to make the food system less environmentally damaging. The search is on both to incentivise better agronomic practices and to have suitable mechanisms for the monitoring and evaluation of such practices. Technology has been seen both by researchers and policy makers as playing a key role. For example, IoT devices as well as earth observation are seen as providing potential sources of monitoring data. More generally, sensors across the food system in combination with machine learning/AI are seen as providing opportunities not just for monitoring but also predicting and decision support. None of this will be possible without extensive effort on enabling data integration both between devices and systems, and among actors and stakeholders. Heterogeneity of actors makes data sharing a key enabling technology.

This paper has described 1) a bottom-up end-user driven process of requirements collection and technological design in the context of the Ploutos project; 2) the Ploutos Core Semantic Model; and 3) the data sharing technology by providing an example in a fruit production supply chain. The requirements collection stage made clear the need for stakeholders to have control over their own data and the uses to which the data is put. Secondly the importance of integrating any new systems with existing systems whether they are legacy systems that a participant already uses or whether they are future systems. Obviously one of the consequences of designing and making available the PCSM is the hope that some future systems may treat such a data model as a native format or may build in *ex initio* compatible APIs but for this there are no guarantees. The third core requirement concerning the avoidance of vendor lock-in also leads a system designer to choose to use semantic technologies, explicit semantic standards, and as much of the semantic web layer cake of technologies as is useful. The technological components and architecture have been designed to make it possible for a variety of data sharing applications or services to be provided i) along the supply chain ii) between actors in a particular stage e.g. among food producers and advisers; iii) and off-chain actors such as food safety agencies or similar. An open standards based approach as proposed here would enable a variety of actors to offer different services to the network while still ensuring the openness enables innovations to occur.

Future work will initially focus on a) adapting the technical solutions to the needs of the other pilots within the project, including the modular extension of the PCSM for additional agrifood domains; and b) understanding how the technical solutions provide opportunities for new or different business models or impose or encourage behavioural changes in order to be successfully adopted, or in contrast provide other new technical requirements. Refining and extending the access control mechanisms by integrating previous work on ontology based access control (OBAC) [46] is also part of planned development. More generally beyond the specific requirements of the project, it will be important to explore the scalability of the approach proposed, and from an adoption perspective to understand how and whether the overall approach can be taken up by the wider agrifood technological community.

Acknowledgements

Parts of the work presented here were funded by the European Commission within the H2020 Programme project "PLOUTOS - A Sustainable Innovation Framework to re-balance agri-food value chains" under grant agreement no. 101000594. The authors are grateful to the farmers, agronomists, fruit processors, and other stakeholders involved in the pilots.

References

[1] M. Crippa, E. Solazzo, D. Guizzardi, F. Monforti-Ferrario, F.N. Tubiello and A. Leip, Food Systems Are Responsible for a Third of Global Anthropogenic GHG Emissions, *Nature Food* **2**(3) (2021), 198–209. doi:10.1038/s43016-021-00225-9. https://www.nature.com/articles/s43016-021-00225-9.

[2] S. Wang, Agrobiodiversity and Agroecosystem Stability, in: *Introduction to Agroecology*, C.D. Caldwell and S. Wang, eds, Springer, Singapore, 2020, pp. 137–154. ISBN 9789811588365. https://doi.org/10.1007/978-981-15-8836-5_10.

[3] S. Diaz, J. Settele, E. Brondizio, H.T. Ngo, M. Gueze, J. Agard and A. Arneth, Summary for Policymakers of the Global Assessment Report on Biodiversity and Ecosystem Services of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services, Technical Report, IPBES, 2019. https://ipbes.net/sites/default/files/2020-02/ipbes_global_assessment_report_summary_for_policymakers_en.pdf.

[4] J. Mateo-Sagasta, S.M. Zadeh and H. Turral, More People, More Food, Worse Water? A Global Review of Water Pollution from Agriculture, Technical Report, FAO and IWMI, 2018. https://www.unwater.org/water-pollution-from-agriculture-a-global-review/.

[5] C. Brewster, I. Roussaki, N. Kalatzis, K. Doolin and K. Ellis, IoT in Agriculture: Designing a Europe-Wide Large-Scale Pilot, *IEEE Communications Magazine* **55**(9) (2017), 26–33. doi:10.1109/MCOM.2017.1600528. http://ieeexplore.ieee.org/document/8030481/.

[6] CEMA, Full Deployment of Agricultural Machinery Data-Sharing: Technical Challenges & Solutions., Technical Report, CEMA, 2020. https://cema-agri.org/images/publications/position-papers/2020_02_05_CEMA_PT3_PP_Strategy_paper_agricultural_machinery_data_sharing.pdf.

[7] Copa-Cogeca, CEMA, CEETTAR, ESA, F. Fertiliser Europe, FEFAC, ECPA, EFFAB and CEJA, EU Code of Conduct on Agricultural Data Sharing by Contractual Agreement, Technical Report, Copa-Cogeca, 2018. http://www.ecpa.eu/sites/default/files/documents/AgriDataSharingCoC_2018.pdf.

[8] A. Durrant, M. Markovic, D. Matthews, D. May, G. Leontidis and J. Enright, How Might Technology Rise to the Challenge of Data Sharing in Agri-Food?, *Global Food Security* **28** (2021), 100493. doi:10.1016/j.gfs.2021.100493. https://www.sciencedirect.com/science/article/pii/S2211912421000031.

[9] K. Spanaki, E. Karafili and S. Despoudi, AI Applications of Data Sharing in Agriculture 4.0: A Framework for Role-Based Data Access Control, *International Journal of Information Management* **59** (2021), 102350. doi:10.1016/j.ijinfomgt.2021.102350. https://www.sciencedirect.com/science/article/pii/S0268401221000438.

[10] M. Wysel, D. Baker and W. Billingsley, Data Sharing Platforms: How Value Is Created from Agricultural Data, *Agricultural Systems* **193** (2021), 103241. doi:10.1016/j.agsy.2021.103241. https://www.sciencedirect.com/science/article/pii/S0308521X21001943.

[11] J. Astill, R.A. Dara, M. Campbell, J.M. Farber, E.D.G. Fraser, S. Sharif and R.Y. Yada, Transparency in Food Supply Chains: A Review of Enabling Technology Solutions, *Trends in Food Science & Technology* **91** (2019), 240–247. doi:10.1016/j.tifs.2019.07.024. http://www.sciencedirect.com/science/article/pii/S0924224418309178.

[12] H. Scholten, C.N. Verdouw, A. Beulens and J.G.A.J. van der Vorst, Defining and Analyzing Traceability Systems in Food Supply Chains, in: *Advances in Food Traceability Techniques and Technologies*, Elsevier, 2016, pp. 9–33. ISBN 978-0-08-100310-7. http://linkinghub.elsevier.com/retrieve/pii/B9780081003107000028.

[13] F. Minnens, N. Lucas Luijckx and W. Verbeke, Food Supply Chain Stakeholders' Perspectives on Sharing Information to Detect and Prevent Food Integrity Issues, *Foods* **8**(6) (2019), 225. doi:10.3390/foods8060225. https://www.mdpi.com/2304-8158/8/6/225.

[14] M.D. Wilkinson, M. Dumontier, I.J.J. Aalbersberg, G. Appleton, M. Axton, A. Baak, N. Blomberg, J.-W. Boiten, L.B. da Silva Santos, P.E. Bourne, J. Bouwman, A.J. Brookes, T. Clark, M. Crosas, I. Dillo, O. Dumon, S. Edmunds, C.T. Evelo, R. Finkers, A. Gonzalez-Beltran, A.J.G. Gray, P. Groth, C. Goble, J.S. Grethe, J. Heringa, P.A.C. 't Hoen, R. Hooft, T. Kuhn, R. Kok, J. Kok, S.J. Lusher, M.E. Martone, A. Mons, A.L. Packer, B. Persson, P. Rocca-Serra, M. Roos, R. van Schaik, S.-A. Sansone, E. Schultes, T. Sengstag, T. Slater, G. Strawn, M.A. Swertz, M. Thompson, J. van der Lei, E. van Mulligen, J. Velterop, A. Waagmeester, P. Wittenburg, K. Wolstencroft, J. Zhao and B. Mons, The FAIR Guiding Principles for Scientific Data Management and Stewardship, *Scientific data* **3** (2016), 160018. doi:10.1038/sdata.2016.18.

[15] S. Rajbhandari and J. Keizer, The AGROVOC Concept Scheme – A Walkthrough, *Journal of Integrative Agriculture* **11**(5) (2012), 694–699. doi:10.1016/S2095-3119(12)60058-6. https://www.sciencedirect.com/science/article/pii/S2095311912600586.

[16] T. Baker, B. Whitehead, R. Musker and J. Keizer, Global Agricultural Concept Space: Lightweight Semantics for Pragmatic Interoperability, *npj Science of Food* **3**(1) (2019), 16. doi:10.1038/s41538-019-0048-6. https://www.nature.com/articles/s41538-019-0048-6.

[17] E. Arnaud, M.-A. Laporte, S. Kim, C. Aubert, S. Leonelli, B. Miro, L. Cooper, P. Jaiswal, G. Kruseman, R. Shrestha, P.L. Buttigieg, C.J. Mungall, J. Pietragalla, A. Agbona, J. Muliro, J. Detras, V. Hualla, A. Rathore, R.R. Das, I. Dieng, G. Bauchet, N. Menda, C. Pommier, F. Shaw, D. Lyon, L. Mwanzia, H. Juarez, E. Bonaiuti, B. Chiputwa, O. Obileye, S. Auzoux, E.D. Yeumo, L.A. Mueller, K. Silverstein, A. Lafargue, E. Antezana, M. Devare and B. King, The Ontologies Community of Practice: A CGIAR Initiative for Big Data in Agrifood Systems, *Patterns* **1**(7) (2020), 100105. doi:10.1016/j.patter.2020.100105. https://www.sciencedirect.com/science/article/pii/S2666389920301392.

[18] EFSA, The Food Classification and Description System FoodEx2 (Revision 2), Technical Report, European Food Standards Agency, 2015. https://efsa.onlinelibrary.wiley.com/doi/epdf/10.2903/sp.efsa.2015.EN-804.

[19] D.M. Dooley, E.J. Griffiths, G.S. Gosal, P.L. Buttigieg, R. Hoehndorf, M.C. Lange, L.M. Schriml, F.S.L. Brinkman and W.W.L. Hsiao, FoodOn: A Harmonized Food Ontology to Increase Global Food Traceability, Quality Control and Data Integration, *npj Science of Food* **2**(1) (2018), 23. doi:10.1038/s41538-018-0032-6.

[20] L. Magliulo, L. Genovese, V. Peretti and N. Murru, Application of Ontologies to Traceability in the Dairy Supply Chain, *Agricultural Sciences* **04**(05) (2013), 41–45. doi:10.4236/as.2013.45B008.

[21] J.P.C. Verhoosel and J. Spek, Applying Ontologies in the Dairy Farming Domain for Big Data Analysis, in: *Proceedings of the Joint Proceedings of the 3rd Stream Reasoning (SR 2016) and the 1st Semantic Web Technologies for the Internet of Thing (SWIT 2016) Workshops (SR+SWIT)*, Kobe, Japan, 2016, pp. 91–100. http://ceur-ws.org/Vol-1783/paper-09.pdf.

[22] J. Verhoosel, M. Van Bekkum and T. Verwaart, Semantic Interoperability for Data Analysis in the Food Supply Chain, *International Journal on Food System Dynamics* **Vol 9** (2018), 101–111 Pages. doi:10.18461/IJFSD.V9I1.917. http://centmapress.ilb.uni-bonn.de/ojs/index.php/fsd/article/view/917.

[23] D. Project, D2.3 Common Data Models and Semantic Interoperability Mechanisms - Release 2, Deliverable, Deliverable D2.3, Demeter Project, 2021.

[24] F.K. van Evert, S. Fountas, D. Jakovetic, V. Crnojevic, I. Travlos and C. Kempenaar, Big Data for Weed Control and Crop Protection, *Weed Research* **57**(4) (2017), 218–233. doi:10.1111/wre.12255.

[25] J. Verhoosel, M. van Bekkum and T. Verwaart, Semantic Interoperability for Data Analysis in the Food Supply Chain, *International Journal on Food System Dynamics* **9**(1) (2018), 101–111. doi:10.18461/ijfsd.v9i1.917. http://centmapress.ilb.uni-bonn.de/ojs/index.php/fsd/article/view/917.

[26] M. Solanki and C. Brewster, Enhancing Visibility in EPCIS Governing Agri-Food Supply Chains via Linked Pedigrees:, *International Journal on Semantic Web and Information Systems* **10**(3) (2014), 45–73. doi:10.4018/IJSWIS.2014070102. http://cbrewster.com/papers/Solanki_IJSWIS15.pdf.

[27] M. Solanki and C. Brewster, OntoPedigree: Modelling Pedigrees for Traceability in Supply Chains, *Semantic Web* **7**(5) (2016), 483–491. doi:10.3233/SW-150179. http://www.semantic-web-journal.net/system/files/swj980.pdf.

[28] A.Y. Halevy, A. Rajaraman and J.J. Ordille, Data Integration: The Teenage Years, in: *Proceedings of the 32nd International Conference on Very Large Data Bases, Seoul, Korea, September 12-15, 2006*, U. Dayal, K.-Y. Whang, D.B. Lomet, G. Alonso, G.M. Lohman, M.L. Kersten, S.K. Cha and Y.-K. Kim, eds, ACM, 2006, pp. 9–16. http://dl.acm.org/citation.cfm?id=1164130.

[29] I. Horrocks, M. Giese, E. Kharlamov and A. Waaler, Using Semantic Technology to Tame the Data Variety Challenge, *IEEE Internet Computing* **20**(6) (2016), 62–66. doi:10.1109/MIC.2016.121. http://ieeexplore.ieee.org/document/7781544/.

[30] M. Klusch, P. Kapahnke, S. Schulte, F. Lecue and A. Bernstein, Semantic Web Service Search: A Brief Survey, *KI - Künstliche Intelligenz* **30**(2) (2016), 139–147. doi:10.1007/s13218-015-0415-7.

[31] A.L. Lemos, F. Daniel and B. Benatallah, Web Service Composition: A Survey of Techniques and Tools, *ACM Computing Surveys* **48**(3) (2015), 33:1–33:41. doi:10.1145/2831270.

[32] J. Domingue, D. Roman and M. Stollberg, Web Service Modeling Ontology (WSMO): An Ontology for Semantic Web Services, in: *W3C Workshop on Frameworks for Semantics in Web Services*, Innsbruck, Austria, 2005. https://www.w3.org/2005/04/FSWS/Submissions/1/wsmo_position_paper.html.

[33] D. Martin, M. Paolucci, S. McIlraith, M. Burstein, D. McDermott, D. McGuinness, B. Parsia, T. Payne, M. Sabou, M. Solanki, N. Srinivasan and K. Sycara, Bringing Semantics to Web Services: The OWL-S Approach, in: *Semantic Web Services and Web Process Composition*, J. Cardoso and A. Sheth, eds, Lecture Notes in Computer Science, Springer, Berlin, Heidelberg, 2005, pp. 26–42. ISBN 978-3-540-30581-1.

[34] J. Miller, K. Verma, P. Rajasekaran, A. Sheth, R. Aggarwal and K. Sivashanmugam, WSDL-S: Adding Semantics to WSDL, *Kno.e.sis Publications* (2005). https://corescholar.libraries.wright.edu/knoesis/199.

[35] S.A. McIlraith and T.C. Son, Adapting Golog for Composition of Semantic Web Services, in: *Proceedings of the Eights International Conference on Principles of Knowledge Representation and Reasoning*, KR'02, Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 2002, pp. 482–496. ISBN 978-1-55860-554-1.

[36] D. Redavid, L. Iannone, T.R. Payne and G. Semeraro, OWL-S Atomic Services Composition with SWRL Rules, in: *Foundations of Intelligent Systems, 17th International Symposium, ISMIS 2008, Toronto, Canada, May 20-23, 2008, Proceedings*, A. An, S. Matwin, Z.W. Ras and D. Slezak, eds, Lecture Notes in Computer Science, Vol. 4994, Springer, 2008, pp. 605–611. https://eprints.soton.ac.uk/265132/1/ISMIS08.pdf.

[37] E. Sirin and B. Parsia, Planning for Semantic Web Services, in: *Proceedings of the ISWC 2004 Workshop on Semantic Web Services: Preparing to Meet the World of Business Applications, SWS@ISWC 2004, Hiroshima, Japan, November 8, 2004*, D. Martin, R. Lara and T. Yamaguchi, eds, CEUR Workshop Proceedings, Vol. 119, CEUR-WS.org, 2004. http://ceur-ws.org/Vol-119/paper1.pdf.

[38] D. Andročec, M. Novak and D. Oreški, Using Semantic Web for Internet of Things Interoperability: A Systematic Review, *International Journal on Semantic Web and Information Systems (IJSWIS)* **14**(4) (2018), 147–171. doi:10.4018/IJSWIS.2018100108. https://www.igi-global.com/article/using-semantic-web-for-internet-of-things-interoperability/210657.

[39] M. Noura, M. Atiquzzaman and M. Gaedke, Interoperability in Internet of Things: Taxonomies and Open Challenges, *Mobile Networks and Applications* **24**(3) (2019), 796–809. doi:10.1007/s11036-018-1089-9.

[40] M. Bauer, E. Kovacs, A. Schülke, N. Ito, C. Criminisi, L.-W. Goix and M. Valla, The Context API in the OMA Next Generation Service Interface, in: *2010 14th International Conference on Intelligence in Next Generation Networks*, 2010, pp. 1–5. doi:10.1109/ICIN.2010.5640931.

[41] N. Kalatzis, G. Routis, Y. Marinellis, M. Avgeris, I. Roussaki, S. Papavassiliou and M. Anagnostou, Semantic Interoperability for IoT Platforms in Support of Decision Making: An Experiment on Early Wildfire Detection, *Sensors* **19**(3) (2019), 528. doi:10.3390/s19030528. https://www.mdpi.com/1424-8220/19/3/528.

[42] J. Verhoosel, C. Brewster, H. Kruiger and B. Nouwt, D4.2 Data Interoperability for the Agri-Food Sector, Deliverable, Ploutos Project, 2021. http://ploutos-h2020.eu/wp-content/uploads/2021/09/Ploutos_D4.2-Data-interoperability-for-the-agri-food-sector_v1.0_2021-06-11.pdf.

[43] N. Kalatzis, C. Brewster, J. Verhoosel, B. Nouwt, H. Kruiger and A. Georgousis, D4.4: Initial Agri-Food Data-Sharing Reference Architecture, Deliverable, Ploutos Project, 2021. http://ploutos-h2020.eu/wp-content/uploads/2021/09/Ploutos_D4.4-Initial-agri-food-data-sharing-reference-architecture.pdf.

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

[45] N. Kalatzis, N. Marianos and F. Chatzipapadopoulos, IoT and Data Interoperability in Agriculture: A Case Study on the gaiasenseTM Smart Farming Solution, in: *2019 Global IoT Summit (GIoTS)*, 2019, pp. 1–6. doi:10.1109/GIOTS.2019.8766423.

[46] C. Brewster, B. Nouwt, S. Raaijmakers and J. Verhoosel, Ontology-Based Access Control for FAIR Data, *Data Intelligence* **2**(1) (2020), 66–77. http://www.data-intelligence-journal.org/p/36/.