BURLEIGH DODDS SERIES IN AGRICULTURAL SCIENCE

Smart farms

Improving data-driven decision making in agriculture

Edited by Professor Claus Grøn Sørensen
Aarhus University, Denmark

burleigh dodds
SCIENCE PUBLISHING

Agri Semantics: developments to improve data interoperability to support farm information management and decision support systems in agriculture

Saba Noor, Jade Bokma and Bart Pardon, Ghent University, Belgium; Gerdien van Schaik, Utrecht University, The Netherlands; and Miel Hostens, Cornell University, USA

1 Introduction
2 Background knowledge and literature review
3 Methodology
4 Case study
5 Conclusion and future trends
6 Acknowledgements
7 References

1 Introduction

1.1 What is Agri Semantics?

In farm animal health management systems (FAHMS), Agri Semantics is part of smart agri systems and is crucial to improving data management and analysis. Semantic web technology is the extension of the World Wide Web and aims to make web content more machine understandable. Ontologies are the fundamental part of developing and implementing a semantic web (Roussaki et al., 2023). Ontologies are the formal and explicit descriptions of data models that define concepts, relationships, and properties within a specific application domain. With the help of Semantic web technologies, Agri Semantics (Khatoon and Ahmed, 2022) enables the structured and semantically enriched representation of farm animal health-related data, leading to several benefits such as improving knowledge representation, data integration, data interoperability, data analysis, and empowering decision support systems that help veterinarians and farmers in making informed decisions related to disease management. By using semantic web technologies, Agri Semantics contributes to more effective disease prevention, early detection, and improved animal health outcomes in the agricultural sector (Niloofar et al., 2021).

http://dx.doi.org/10.19103/AS.2023.0132.05
© The Authors 2024. This is an open access chapter distributed under a Creative Commons Attribution 4.0 License (CC BY).

Chapter taken from: Sørensen, C. G. (ed.), *Smart farms: Improving data-driven decision making in agriculture*, pp. 75–96, Burleigh Dodds Science Publishing, Cambridge, UK, 2024, (ISBN: 978 1 80146 382 9; www.bdspublishing.com)

Data-intensive farming (Bullock et al., 2019) refers to the practice of collecting, analyzing, and utilizing agricultural data for the management and optimization of farming practices. It involves the collection of data from various sources, such as sensors, weather stations, and farm management software to get useful information and make data-driven decisions in agricultural operations. Farmers are encouraged to share their farm-related data in exchange for rewards such as financial rewards, access to specific services, and personalized recommendations, leading to a comprehensive data ecosystem. Through data-driven insights and collaboration, data-intensive farming improves productivity, sustainability, and knowledge sharing in agriculture. On the other hand, precision livestock farming (PLF) is an advanced technology to automate real-time monitoring of farm animal health data, productivity, environmental impact, and welfare (Bahlo et al., 2019). Due to the advancement in this technology, a large amount of data has become available, but the data availability (Villeneuve et al., 2019), confidentiality, accessibility, interoperability, and complexity of such heterogeneous data remain challenging (Hermans et al., 2018). The usefulness of PLF can be increased if the data from different sources are processed together to give more meaningful and unexplored insights (Symeonaki et al., 2022). Although PLF appeals to young people and helps agriculture, it can harm farmers and animals if not tailored to their needs and skills (Fote et al., 2020). From the farmer's point of view, PLF techniques erode their observation skills and increase dependency on tools.

The term 'farm animal health' describes the mental and physical well-being of farm animals. It incorporates various aspects, such as disease prevention, veterinary care, nutrition, and overall welfare. On the other hand, farm management systems incorporate the overall practices and strategies employed by farmers' points of view to raise and care for their animals. While farm animal health and farm animal management systems are interconnected, they are not synonymous. The management system plays a significant role in influencing animal health outcomes. For example, the housing system chosen can impact animal welfare, disease transmission, and stress levels. In this context, the goal of the EU-funded project called DECIDE (European Horizon, 2020 project https://decideproject.eu/) is to develop decision-support tools for improved control of endemic infectious diseases in livestock and aquaculture. DECIDE focuses on respiratory and gastrointestinal syndromes in swine, poultry, cattle, and growth reduction in salmonids. The project involves identifying stakeholder needs, accessing relevant data, determining disease burden and control costs, implementing early warning systems, and understanding disease impact. The early warning system within DECIDE aims to provide timely information to stakeholders, such as farmers and veterinarians, enabling them to make informed decisions on disease-mitigating actions. By detecting signals early,

Published by Burleigh Dodds Science Publishing Limited, 2024.

stakeholders can take proactive measures to prevent or control disease spread and minimize the impact on animal health and welfare (van Schaik et al., 2023).

1.2 Data challenges in traditional farm animal health management systems and the importance of Agri Semantics?

In the context of PLF, traditional farm animal information systems pose significant challenges to data collection, cleaning, integration, and analysis (Rosa, 2021). One of the key data challenges is the lack of standardization of data representation. The collected data comes from various sources such as laboratories, veterinary clinics, and research institutes and often follow a different format and lack a unified structure. This heterogeneity and lack of standardization make it difficult to effectively utilize the data for decision-making. Such systems lack extensibility and are not stable to communicate in heterogeneous systems (Niloofar et al., 2020). Another challenge is the limited interoperability between different disease management systems and data sources. In traditional systems, data exchange and collaboration among stakeholders are hindered by incompatible formats and are difficult to share, therefore limiting the ability to collaborate and exchange data among stakeholders. Additionally, traditional information management systems (Monteiro et al., 2021) face difficulties to ensure data quality and completeness. Outdated, incomplete, and inaccurate data compromise the reliability of disease surveillance, diagnosis, and decision-making. The lack of quality control mechanisms may lead to inconsistencies and errors in the collected data. The lack of advanced analytical capabilities for knowledge discovery and decision support (Newton et al., 2020) is another significant data challenge. This hinders the identification of patterns and trends in disease occurrence, correlation, and transmission. Thus, it limits the ability to develop effective information management strategies and timely interventions.

Addressing these data challenges is crucial for improving the effectiveness of traditional information management systems. The adoption of modern approaches (Newton et al., 2020), such as ontology-based systems supported by Agri Semantics, plays an important role in addressing these data challenges by providing a framework for structuring and harmonizing data in a meaningful way. By implementing ontologies, Agri Semantics enables the creation of a common understanding and representation of agricultural concepts, facilitating data integration across different sources, providing interoperability, and enabling advanced data analysis (Murali and Anouncia, 2022). It allows for the seamless exchange and sharing of data, enabling collaborative efforts in farm animal disease management. By addressing these challenges, stakeholders can enhance disease surveillance, early detection, and control efforts, ultimately leading to improved animal health outcomes and promoting

effective prevention and more efficient disease management practices (Hu et al., 2020).

1.3 Existing animal livestock ontologies and techniques

In the context of PLF, numerous ontology-based approaches were proposed, such as AGROVOC agriculture thesaurus (Rajbhandari and Keizer, 2012), Animal Trait Ontology for Livestock (ATOL) (Golik et al., 2012), Infectious Disease Ontology (IDO) (Cowell and Smith, 2010), Animal Disease Ontology (ADO) (Malhotra et al., 2014), and Animal Health Surveillance Ontology (AHSO) (Dórea et al., 2019). However, these approaches do not provide a dedicated ontology focusing on disease and pathogen identification in farm animals. Moreover, these ontologies do not provide the level of granularity for our cattle barometer (https://decideproject.eu/barometer/) use case.

Ontologies in farm animal health management systems have drawbacks, including a lack of user knowledge. A practical approach to address these limitations is to develop a cattle barometer use case framework. By developing a species-specific framework, we can help guide the implementation and incorporation of ontologies in the context of cattle health management. This will help us to monitor the implementation process, gather feedback from users, and evaluate the impact of the framework on cattle health and management outcomes. This step will help identify challenges and areas for improvement.

1.4 Problem statement

In the field of FAHMS the transmission of infectious diseases poses a significant challenge, leading to threats to animal health, welfare, and reduced productivity. The datasets related to transmission and contagious diseases often exhibit variations in format, semantics, syntax, and structure. To address the heterogeneity of data, we need to design a cost-effective, accessible, reusable, and lightweight ontology-driven knowledge-based domain-specific framework for farm animal health management systems. This chapter focuses on a cattle barometer use case scenario, which serves as a representative example for the application of the proposed ontology-driven knowledge-based framework in farm animal health management systems. The cattle barometer use case scenario serves as practical and illustrative to represent the effectiveness and applicability of the proposed framework in real-world farm animal health management systems that follows the FAIR (Findable, Accessible, Interoperable, Reusable) guideline principles (van Schaik et al., 2023).

The key contribution of this research include the following:

1. Establishing a standardized framework on FAIR guideline principles to tradeoff between data requirements, data accessibility, availability, data connectivity, and value generation;
2. Developing a species-specific ontology called Livestock Health Ontology (LHO) in which knowledge graphs are enriched by semantic knowledge coming from domain knowledge;
3. Generating Resource Descriptive Framework (RDF) data from the raw data of the cattle use case scenario, leveraging the advantages of RDF for enhanced data sharing, integration, and querying capabilities;
4. Mapping of cleaned RDF data to the species-specific LHO ontology for semantic interoperability; and
5. Utilizing a data visualization tool for visualizing and reporting farm animal disease management data to gain meaningful insight, enhancing data interpretation, and decision-making capabilities.

1.5 Organization of the chapter

The remaining sections of the chapters are organized as follows: In Section 2, we provide background knowledge and a literature review. In Section 3, we present the methodology in detail. In Section 4, we describe a case study integrating and analyzing data from several animal health data sources. In Section 5, we conclude the paper and discuss future research goals and challenges.

2 Background knowledge and literature review

2.1 Ontology-based approaches

In the field of PLF, numerous ontological sources have been developed such as the AGROVOC agriculture thesaurus (Rajbhandari and Keizer, 2012), the most popular one managed by the Food and Agriculture Organization of the United Nations (FAO) which aligns with the FAIR principles (Meyer et al., 2021). It is a linked open data for agriculture consisting of over 3200 concepts and available in 20 different languages. It serves as valuable resource for capturing general concepts regarding agriculture and animal health. However, AGROVOC does not contain a dedicated ontology focusing on disease and pathogen identification in farm animals (Gangadharan and Gupta, 2020).

Other smaller ontologies are IDO (Cowell and Smith, 2010), Coronavirus Infectious Disease Ontology (CIDO), and VIDO named Virus Infectious Disease Ontology (Babcock et al., 2021). These ontologies are the structural representation of infectious diseases and rely on domain-specific extension for example tuberculosis and coronaviruses of various species like cattle, pigs, poultry, and fish. ATOL (Golik et al., 2012) is another ontology that models a multi-species ontology for animal phenotype traits. It focuses on animal

Table 1 Comparative analysis of ontology-based versus non-ontology-based animal health management systems

Ontology-based			Non-ontology based		
Ontology name	Remarks	Reference	Method name	Remarks	Reference
AGROVOC	Constantly updated	https://agrovoc.fao.org/browse/agrovoc/en/index/B	RISKSUR EVA	Non-ontological base but by using EVA tool may in cooperate the elements of ontologies	https://fp7-risksur.eu/results/tools
IDO	Published 2014 and last updated on 08-03-2022	https://bioportal.bioontology.org/ontologies/IDO	EpiTools	Calculating various epidemiological measures	https://epitools.ausvet.com.au/
CIDO	Extension of IDO, updated on August 3, 2020	https://bioportal.bioontology.org/ontologies/IDO-COVID-19	Surveillance Evaluation Tool (SET)	Measure the performance of animal disease surveillance	https://extranet.who.int/sph/index.php/evaluation-action-fao-surveillance-evaluation-tool-set
VIDO	Extension of IDO, updates on November 27, 2022	https://bioportal.bioontology.org/ontologies/VIDO	--------	---------	----------
ATOL	Published on 2014 and updated May 11, 2020	https://bioportal.bioontology.org/ontologies/ATOL	---------	----------	----------
ADO	Last uploaded: November 14, 2018	https://agroportal.lirmm.fr/ontologies/ANDO	--------	---------	----------
AHSO	Last update 2018	Link available on Bioportal and Github: https://bioportal.bioontology.org/ontologies/AHSO/?p=summary https://github.com/SVA-SE/AHSO	--------	---------	----------

Published by Burleigh Dodds Science Publishing Limited, 2024.

products like meat, milk quality, and quantity, reproductivity, and welfare and was last updated in 2014. Similarly, ADO (Malhotra et al., 2014) focuses on representing the knowledge of animal diseases specifically focus 500 pairs of genetic disorders and infectious and metabolic diseases of different animal species. It focuses on the concepts regarding disease treatment, pathogens, diagnostic tests, symptoms, and progression. AHSO (Dórea et al., 2019) focuses on the relation and concepts of animal health surveillance. It provides structural information about disease reporting, monitoring, and surveillance activities. However, these ontologies do not provide the level of granularity that we need for our cattle barometer use case. The conclusion is that the existing ontologies cannot be reused as they are and need to be extended or modified to align with the desired level of detail and compatibility of relations and concepts with the design ontology. An overview of ontology-based versus non-ontology-based animal health management systems is given in Table 1.

2.2 Non-ontology-based approaches (traditional approaches)

Traditional farm animal information systems like Epitools are used for epidemiological analysis, but their usability for non-veterinarians has been questioned (Cocca and Lindberg, 2019), although valuable for decision-making in animal health. Similarly, the Surveillance Evaluation Tool (SET) (Vasconcelos Gioia et al., 2021) is developed by WOAH (World Organisation for Animal Health) to measure the performance and use of animal disease surveillance. To use this tool effectively, it may require a good understanding of surveillance methodologies and principles. RiskSUR (Peyre et al., 2019) is a tool that provides a framework for designing, implementing, analyzing, and reporting surveillance systems. The aforementioned systems have limitations to incorporate advanced modeling technologies consequently providing less accessibility, interoperability, reusability, and availability (Bhat and Huang, 2021). Furthermore, such type of systems are not farmer friendly and requires user expertise to handle and manage.

3 Methodology

3.1 System model

Figure 1 depicts the data flow of a generic non-ontology based versus ontology-driven knowledge-based framework of Farm Animal Data Management ODKFADM. Non-ontology-based systems are systems that do not use ontologies. Instead they rely on relational databases or other data management and statistical techniques. In such systems, the process of converting raw data into meaningful statistical results is repetitive, time-consuming, and tedious work (Ren et al., 2020). Every time when a new dataset comes in, it requires

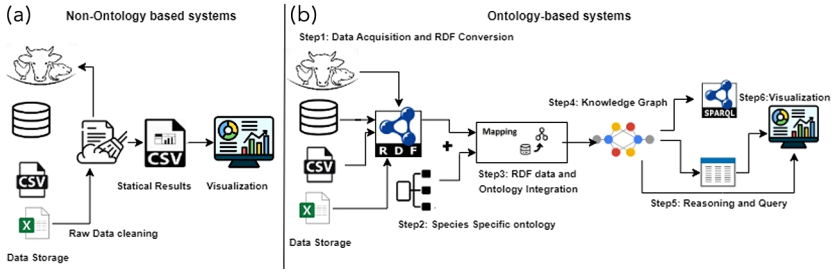


Figure 1 (a) Non-ontology based versus (b) ontology-driven Knowledge-based framework of Farm Animal Data Management (ODKFADM).

rewriting data cleaning code because the naming convention, format, and structure of each dataset are different and require specific adjustments to the data cleaning and processing steps. Therefore, non-ontology-based systems lack standardized semantic structures and reasoning that involve inferring new knowledge from existing datasets and relationships, limiting the system's overall flexibility, adaptability, reusability, and integrity.

However, in ODKFADM the data stored in heterogeneous resources are exposed as a knowledge graph and enriched with semantic knowledge coming from domain knowledge. This ontology mapping results in knowledge graph (Chen et al., 2020), the heterogeneous device data can be expressed in terms of ontologies. Ontologies provide an explicitly structured and standardized formal representation of different concepts, classes, properties, and relationships within the domain of animal farm management. Once the ontology is established, the concepts and relationships remain unchanged nevertheless of the specific dataset being processed. In formal ontology-based approaches (Saha, 2021) that support knowledge graphs, frequent code modification is limited. Therefore, an ontology-driven approach can be reusable promoting interoperability, flexibility, and integrity. It also facilitates querying and reasoning over the data at the ontological level.

4 Case study

4.1 Dataset and test bed

We evaluate the ODKFADM on data focusing on cattle. It involves collection of raw data from different animal health service labs in Europe from different countries. These datasets include information about farms, their geolocation, breed type (beef, dairy, or mixed), diagnostic tests, pathogen identification, and pathogen results for specific bacteria such as *Histophilus somni* (HS), *Mannheimia haemolytica* (MH), *Pasteurella multocida* (PM), *Mycoplasma bovis*

(MB), and viruses such as bovine coronavirus (BCV) and bovine respiratory syncytial virus (BRSV).

The aim is to demonstrate the usefulness of the ODKFADM for analyzing and understanding cattle-related raw data by using the R, Python and Pyspark framework to enhance data integration and advance semantic reasoning capabilities. Moreover, it also improves decisions on disease control and increases animal health and welfare. To achieve this, the ODKFADM framework follows a multistep approach. In addition, the R, Pyspark, and pandas notebook code is uploaded on GitHub, making it publicly available for use. This allows other researchers, data scientists, and stakeholders to access and utilize the code for their own cattle-related data analysis projects. By sharing these resources, we can foster collaboration and knowledge exchange, leading to further advancements in disease control, animal health, and welfare. The code is available at decide-project-eu/LiveStockHealthOntologyWp1 (github.com), and the ontology can be accessed on BioPortal at https://bioportal.bioontology.org/ontologies/LHO version 1.0.0.

4.2 Step 1: data acquisition and resource descriptive frameworks conversions

The code for step 1, as illustrated in Fig. 2, is available on GitHub. In this step, we read the raw data from heterogeneous resources using pandas, Pyspark, and R data frames and then convert the data frames into RDF (resource descriptive frameworks) format. Basically, the RDF (Ali et al., 2022) format is a way to represent the data in an explicit standardized format. To perform these conversions, we utilize the RDFLib library in Python and the RDF packages in R. During RDF conversion, we implement anonymization techniques to protect sensitive information. For this purpose, personal identifiers such as sample numbers, farm identification, and file numbers were anonymized by using the cryptographic SHA 256 hash function. SHA256 hash function ensured that the sensitive information could not be linked back to a specific entity or individual while preserving identifications.

4.3 Step 2: species-specific Livestock Health Ontology

Ontology is the formal and explicit description of a data model within an application domain in which relationships occur between individuals, making it easy to reuse and share knowledge. LHO is an extension of the DECIDE ontology. LHO follows OBO Foundry guidelines and contains information related to the health and well-being of livestock animal species such as cattle. Furthermore, it represents the explicit structured and standardized framework for specific domain knowledge and data integration of livestock species. In the

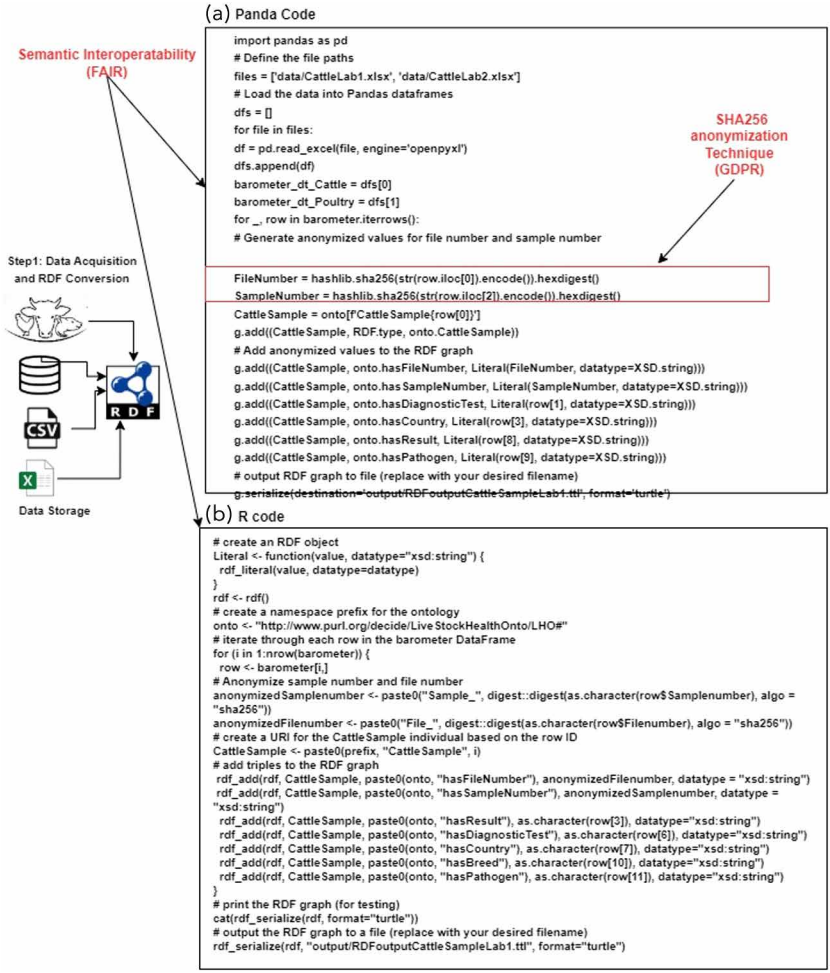


Figure 2 Data acquisition and RDF conversions.

development process, knowledge can be acquired from domain experts, and the initial prototype is designed to facilitate reasoning specifically for cattle-related data. Furthermore, LHO contains information on anonymized farm identification, samples, geolocation, breed, diagnostic tests, and pathogen results for bacteria such as HS, MH, PM, and viral pathogen like BCV along with respiratory culture results. This information is represented in the form of concepts of classes, subclasses, and their relationship in the form of object and data properties. The LHO is uploaded on Biopoartal and can be accessed at https://bioportal.bioontology.org/ontologies/LHO. By making LHO available

on Bioportal, we aim to promote its use and encourage community feedback in the field of farm animal health management.

4.3.1 Formal representation of the Livestock Health Ontology

Ontologies belong to the specific domain environment that contains a semantic level, in terms of vocabulary and their relationship (Gómez-Pérez and Corcho, 2002). The LHO model can be formally represented as:

LHO = (C, P, A, I)

Where LHO stands for Livestock Health Ontology model, C is the concepts and instances conceptualization, P is the properties and their relationships, A represents the rules or axioms, and I represent the model interpretation. In LHO, the object properties (P) and relationships (Re) can be formally represented by the following symbols:

P = (D: Data type; O: Object; T: Transitive; F: Functional)

Re = (≡: equivalence; ⊂: subsumed; ∩: Disjoint)

The domain of property is formally represented as follows:

[∀b; b: p (a; b) → Domain (a)]; p∈ P; Domain ∈ C

The range of property is formally represented as follows:

[∀b; b: p (a; b) → Range (b)]; p ∈ P; Range ∈ C

The interpretation of the ontology model is represented as LHO. Classes and properties possess many Re such as subsuming, equivalence, and disjoint relationships. The symbol ≡ shows the equivalence of classes; the symbol ∩ shows conceptually disjoint classes and is symmetric, reflexive, and transitive. Some examples of these relationships in the LHO ontology model are

CattleSample ∩ PoultrySample ∩ PigSample ⊂ *LiveStockHealthOnto*

In the LHO model, all (⊂) subsume relations are asymmetric, irreflexive, and transitive. However, ∩ is conceptually disjoint, and it never violates its property of being transitive (relationships can be inherited), symmetric (relationships are bidirectional), and reflexive (entities can have a relation with themselves). To make the LHO model consistent, it is necessary to fulfill all these properties. Some examples of these hierarchal relationships in the LHO ontology model are

LiveStockHealthOnto ⊂ *LiveStockHealthOntology*

CattleSample ⊂ *LiveStockHealthOnto*

4.3.2 Define classes and subclasses of Livestock Health Ontology

In LHO, *LiveStockHealthOntology* is represented as the top-level concept. *SampleType, SampleNumber, FarmIdentification, Breed, Province, PostalCode, Country, DiagnosticTest, PathogentIdentification, PathogenResults*, etc. are the subclasses of the *LiveStockHealthOnto* class, as shown in Fig. 3.

Figure 3 Define classes and subclasses of LHO.

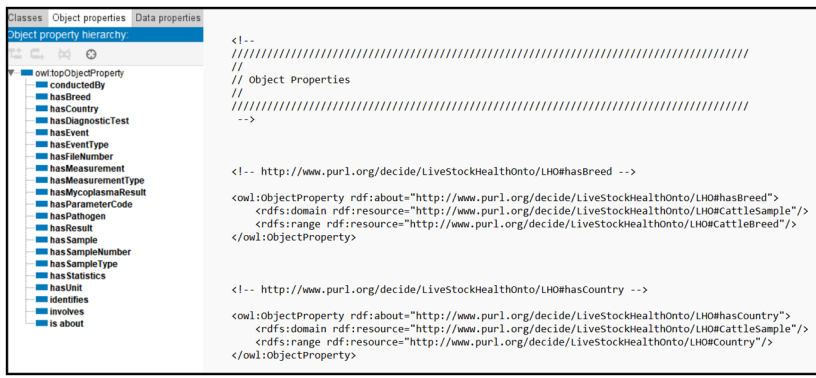


Figure 4 Define object properties of LHO.

4.3.3 Define object properties of Livestock Health Ontology

Figure 4 shows the relationships of object properties in Ontology web language (Owl) and RDF such as *hasBreed*, *hasCountry*, *hasPathogenIdentification*, and *hasPathogenresults*.

4.3.4 Define data properties of Livestock Health Ontology

The proposed ontology also defines data properties to represent the relationship between individuals and data literals. Annotation properties can be used for describing metadata properties. Figure 5 shows the data properties

and its domain and range of particular data properties using semantic web technologies Owl and RDF schemas.

4.3.5 Define Livestock Health Ontology axioms

LHO axioms represent the relationship between classes, instances, and attributes. For example, Fig. 6 shows class axioms describing the semantic structure and meaning such as Owl: disjoint class, rdfs: Subclassof, rdf: Description, and rdf: type. After defining these axioms, we can provide the disjoint and equivalence relationships among classes.

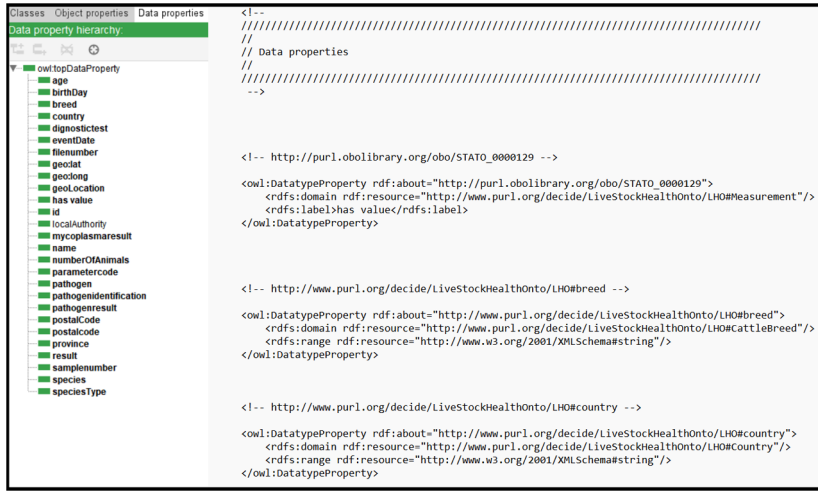


Figure 5 Define data properties of LHO.

```
<!--
////////////////////////////////////////////////////////////////////////////////
//
// General axioms
//
////////////////////////////////////////////////////////////////////////////////
 -->

<rdf:Description>
    <rdf:type rdf:resource="http://www.w3.org/2002/07/owl#AllDisjointClasses"/>
    <owl:members rdf:parseType="Collection">
        <rdf:Description rdf:about="http://www.purl.org/decide/LiveStockHealthOnto/LHO#CattleSample"/>
        <rdf:Description rdf:about="http://www.purl.org/decide/LiveStockHealthOnto/LHO#PigSample"/>
        <rdf:Description rdf:about="http://www.purl.org/decide/LiveStockHealthOnto/LHO#PoultrySample"/>
    </owl:members>
</rdf:Description>
<rdf:Description>
    <rdf:type rdf:resource="http://www.w3.org/2002/07/owl#AllDisjointClasses"/>
    <owl:members rdf:parseType="Collection">
        <rdf:Description rdf:about="http://www.purl.org/decide/LiveStockHealthOnto/LHO#CattleSampleResult"/>
        <rdf:Description rdf:about="http://www.purl.org/decide/LiveStockHealthOnto/LHO#PigSampleResult"/>
        <rdf:Description rdf:about="http://www.purl.org/decide/LiveStockHealthOnto/LHO#PoultrySampleResult"/>
    </owl:members>
</rdf:Description>
</rdf:RDF>
```

Figure 6 Define axioms of LHO.

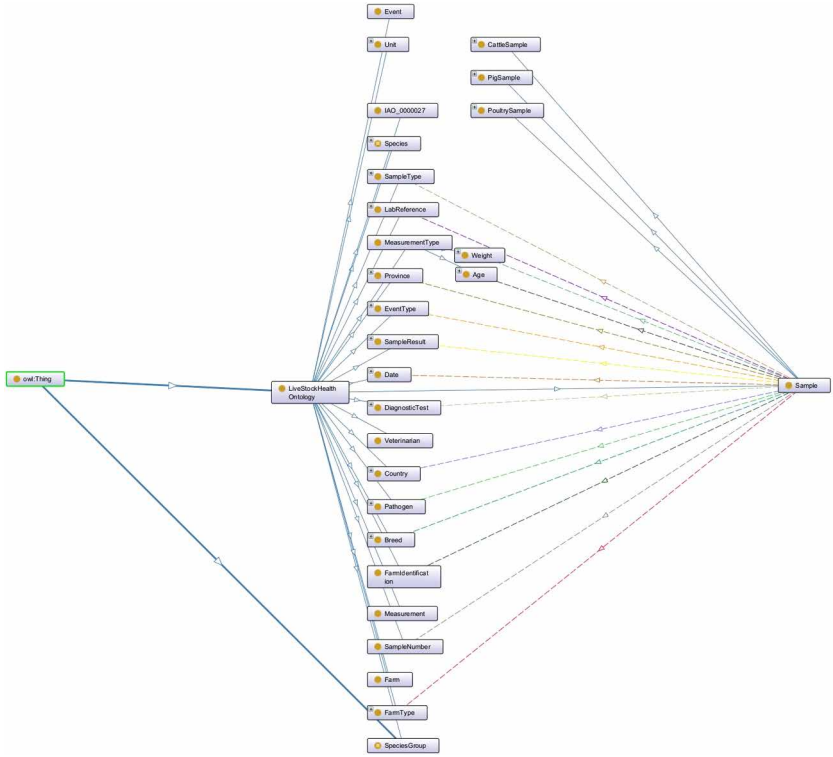


Figure 7 Graphical representation of LHO.

4.3.6 Graphical representation of ontology

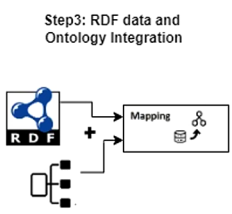
Figure 7 shows the graphical representation of ontology in Onto Graph representing classes, subclasses, and their relationships.

4.4 Step 3: Resource Descriptive Framework data and ontology integration (mapping)

With the help of RDF data representation, we map the RDF data explained in step 1 with LHO in step 2. This mapping process (Sahoo et al., 2009) enriches the data with semantic meanings, thus enabling advanced reasoning and querying as shown in Fig. 8.

4.5 Step 4: knowledge graph (ontology update)

The knowledge graph is a well-unified, structured, and integrated knowledge repository. When the RDF data are integrated with the ontology, it may require updating existing ontology consequently contributing to the development




```
# Step 3: Load the RDF data and ontology into a Panda DataFrame:
# Define the original path
path_to_RDF = "output/RDFoutputCattleSampleLab1.ttl"
   # Create a new graph
   g = Graph()
   # Parse the RDF file in Turtle format
   g.parse(path_to_RDF, format='ttl')

   # Parse the ontology file in OWL format and add it to the graph
   path_to_ontology = "Ontology/LivestockHealthOnto1.0.owl"
   g.parse(path_to_ontology, format="xml")

#Use RDFS or OWL reasoning to infer additional knowledge
g.bind('rdfs', RDFS)
g.bind('owl', OWL)
g.bind('onto', Namespace("http://www.purl.org/decide/LivestockHealthOnto"))
```

Figure 8 RDF data and ontology integration (mapping).

	FileNumber	DiagnosticTest	SampleNumber	Breed	LabReference	SampleType	PostalCode	Province	PathogenIdentification	PathogenResult	MycoplasmaResult
0	381ab40916b42502cd14b9de7201d798d31c35255ccd2a...	Culture	854116431e19bdb205ca2a76204cda6aba4069bb500bdc...	Beef	1	Autopsy	9800	nan	Histophilus somni	POS++	POS++
1	381ab40916b42502cd14b9de7201d798d31c35255ccd2a...	Culture	854116431e19bdb205ca2a76204cda6aba4069bb500bdc...	Beef	1	Autopsy	9800	nan	Histophilus somni	POS++	POS++
2	381ab40916b42502cd14b9de7201d798d31c35255ccd2a...	Culture	854116431e19bdb205ca2a76204cda6aba4069bb500bdc...	Beef	1	Autopsy	9800	nan	Mannheimia haemolytica	POS++	POS++
3	381ab40916b42502cd14b9de7201d798d31c35255ccd2a...	Culture	854116431e19bdb205ca2a76204cda6aba4069bb500bdc...	Beef	1	Autopsy	9800	nan	Mannheimia haemolytica	POS++	POS++
4	381ab40916b42502cd14b9de7201d798d31c35255ccd2a...	Culture	854116431e19bdb205ca2a76204cda6aba4069bb500bdc...	Beef	1	Autopsy	9800	nan	Mycoplasma bovis	POS++	POS++

Figure 9 Query 1: filtering positive PathogenResults and MycoplasmaResults associated with CattleSample.

of a knowledge graph. Such updating could be done by adding new classes, relationships, and properties concepts for a better understanding of integrated data, thus resulting in a knowledge graph that enhances more advanced reasoning and querying capabilities (Hogan et al., 2021).

4.6 Step 5: reasoning and query

To explore the knowledge graph, we have used reasoning and SPARQL queries (Kollia et al., 2011). The following section elaborates on different SPARQL queries on LHO.

4.6.1 Query 1: filtering positive PathogenResults and MycoplasmaResults associated with CattleSample

Figure 9 shows the query that filters positive PathogenResult and MycoplasmaResult which is associated with CattleSample-related data. For each CattleSample, the results show the information of file number, diagnostic tests, breed, sample type, sample number, lab reference, geolocation, and pathogen information.

```
PREFIX decide: <http://www.purl.org/decide#>
PREFIX LHO: <https://www.purl.org/decide/LiveStockHealthOnto/LHO#>
```

SELECT ?FileNumber ?DiagnosticTest ?SampleNumber ?Breed ?LabReference ?SampleType ?Result ?Pathogen ?PostalCode ?Province ?PathogenIdentification ?PathogenResult ?MycoplasmaResult

WHERE {
?CattleSample LHO:hasFileNumber ?FileNumber .
?CattleSample LHO:hasDiagnosticTest ?DiagnosticTest .
?CattleSample LHO:hasSampleNumber ?SampleNumber .
?CattleSample LHO:hasBreed ?Breed .
?CattleSample LHO:hasLabReference ?LabReference .
?CattleSample LHO:hasSampleType ?SampleType .
?CattleSample LHO:hasResult ?Result .
?CattleSample LHO:hasPathogen ?Pathogen .
?CattleSample LHO:hasPostalCode ?PostalCode .
?CattleSample LHO:hasProvince ?Province .
?CattleSample LHO:hasPathogenIdentification ?PathogenIdentification .
?CattleSample LHO:hasPathogenResult ?PathogenResult .
FILTER (?PathogenResult = "POS++")
?CattleSample LHO:hasMicoplasmaResult ?MycoplasmaResult .
FILTER (?MycoplasmaResult = "POS++")

4.6.2 Query 2

Breed has three types: beef, meat, or mixed. The SPARQL query 2 retrieves the results based on Breed begin Beef as shown in Fig. 10

PREFIX decide: <http://www.purl.org/decide#>
PREFIX LHO: <https://www.purl.org/decide/LiveStockHealthOnto/LHO#>
SELECT ?FileNumber ?DiagnosticTest ?SampleNumber ?Breed ?LabReference ?SampleType ?Result ?Pathogen ?PostalCode ?Province ?PathogenIdentification ?PathogenResult ?MycoplasmaResult

WHERE {
?CattleSample LHO:hasFileNumber ?FileNumber .
?CattleSample LHO:hasDiagnosticTest ?DiagnosticTest .
?CattleSample LHO:hasSampleNumber ?SampleNumber .
?CattleSample LHO:hasBreed ?Breed .

	FileNumber	DiagnosticTest	SampleNumber	Breed	LabReference	SampleType	Result	Pathogen	PostalCode	Province	PathogenIdentification	PathogenResult	MycoplasmaResult
0	952fe97b86661dc65d3c1bae7e87dcac57f0cd96cc92b2...	PCR	22a35575c079e307143b9fcfb252d27eefea042fbbc958...	Beef	1	BAL	0	AD Corona Ag (PCR)	8700	West Flanders	nan	nan	nan
1	952fe97b86661dc65d3c1bae7e87dcac57f0cd96cc92b2...	PCR	6e116f0e1b58751c4ad0d6144618e77ffe8f1792437291...	Beef	1	BAL	0	AD Corona Ag (PCR)	8700	West Flanders	nan	nan	nan
2	952fe97b86661dc65d3c1bae7e87dcac57f0cd96cc92b2...	PCR	f81c48f81ec2fd83f7485884ec165b48241202fd2430b3...	Beef	1	BAL	0	AD Corona Ag (PCR)	8700	West Flanders	nan	nan	nan
3	952fe97b86661dc65d3c1bae7e87dcac57f0cd96cc92b2...	PCR	22a35575c079e307143b9fcfb252d27eefea042fbbc958...	Beef	1	BAL	1	AD Corona Ag (PCR)	8700	West Flanders	nan	nan	nan
4	952fe97b86661dc65d3c1bae7e87dcac57f0cd96cc92b2...	PCR	6e116f0e1b58751c4ad0d6144618e77ffe8f1792437291...	Beef	1	BAL	1	AD Corona Ag (PCR)	8700	West Flanders	nan	nan	nan

Figure 10 Query 2: filtering the results based on the breed being 'Beef' associated with CattleSample.

```
FILTER (?Breed = "Beef")
?CattleSample LHO:hasLabReference ?LabReference .
?CattleSample LHO:hasSampleType ?SampleType .
?CattleSample LHOhasResult ?Result .
?CattleSample LHO:hasPathogen ?Pathogen .
?CattleSample LHO:hasPostalCode ?PostalCode .
?CattleSample LHO:hasProvince ?Province .
?CattleSample LHO:hasPathogenIdentification ?PathogenIdentification .
?CattleSample LHO:hasPathogenResult ?PathogenResult .
?CattleSample LHO:hasMicoplasmaResult ?MycoplasmaResult .
}
```

4.7 Step 6: visualization and analysis

On the basis of our knowledge graph complexity, tools, and techniques, we choose the tool Tableau for visualization. Tableau provides meaningful insights to explore the knowledge graph. For this, we need a working ODBC connection to a Virtuoso Instance and ODBC or JDBC Compliant version of Tableau or Tableau Server (Heward-Mills, 2017). We can create a visualization specifically tailored for the cattle barometer use case, focusing on respiratory infectious diseases. This visualization, known as the cattle barometer, is a powerful visualization tool designed to provide meaningful insights and understanding

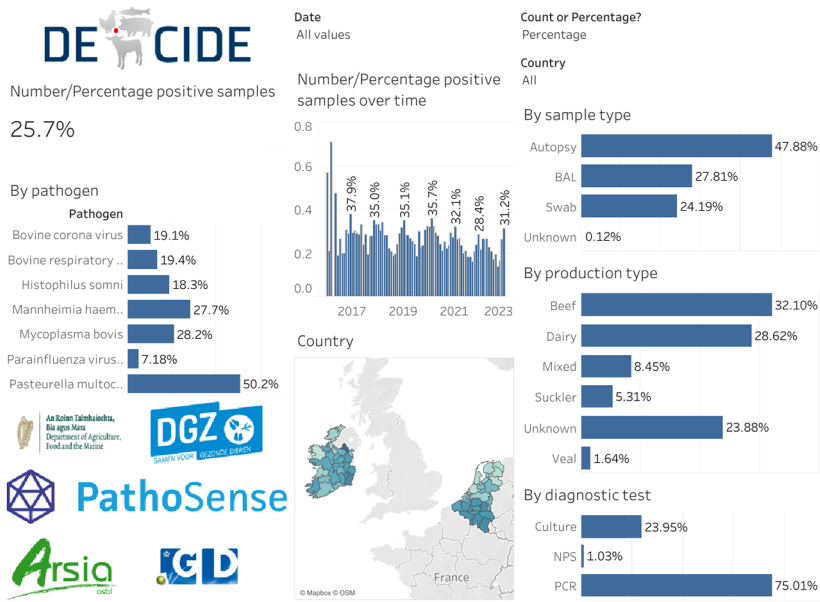


Figure 11 Cattle barometer use case, focusing on respiratory infectious diseases.

from the data related to respiratory infectious diseases in cattle. This tool is intended to be used by veterinary professionals, researchers, and livestock industry stakeholders such as farmers and producers keen on monitoring and mitigating the impact of diseases on cattle population, as shown in Fig. 11. The following are the different types of users who interact with a system or tool to achieve specific goals in future.

Veterinarians: Veterinary professionals can use the cattle barometer to monitor disease trends in specific regions, make informed decisions about vaccination strategies, and allocate resources to high-risk areas.

Researchers: Researchers can leverage the cattle barometer to identify patterns and contributing factors to disease outbreaks, facilitating the development of targeted interventions and prevention strategies.

Stakeholders: Stakeholders in the livestock industry, including farmers and producers, can access actionable insights to optimize herd management practices, minimize disease-related losses, and ensure sustainable production.

5 Conclusion and future trends

In this chapter, we have presented an ontology-driven, knowledge-based framework for farm animal health management systems, with a specific focus on a cattle barometer use case. We recognize the significant challenges faced in data acquisition, integration, and analysis within this domain and highlight the importance of agricultural semantics in addressing these challenges. By leveraging semantic web technologies, our framework aims to improve data management, enable informed decision-making, and ultimately enhance animal health outcomes.

One of the key contributions of our framework is the development of the LHO, which serves as a standardized framework for representing and organizing cattle health-related knowledge. LHO extends the existing DECIDE ontology and provides a species-specific ontology tailored specifically for cattle. This ontology-driven approach ensures standardization, integrity, interoperability, and reusability of data within farm animal health management systems. By making LHO available on Bioportal, we aim to promote its use and encourage community feedback in the field of farm animal health management.

To demonstrate the effectiveness of our ODKFADM, we have conducted a comprehensive case study focused on disease surveillance in cattle. The case study shows the practical application of our framework in improving disease monitoring, gathering feedback from users, and evaluating the impact of the framework on cattle health and management outcomes. The SPARQL query results highlight the potential of our approach in the field of farm animal health management.

Published by Burleigh Dodds Science Publishing Limited, 2024.

In the future, we are extending the LHO to include species-specific ontologies for other farm animals such as salmon, pigs, and poultry. By doing so, we will further enhance the semantic interoperability and applicability of our framework across different livestock sectors. This expansion will enable a broader range of stakeholders to leverage the benefits of our ontology-driven approach, promoting data standardization and advanced analytical reasoning in the management of various farm animal health systems.

In conclusion, ontology-driven, knowledge-based framework for farm animal health management systems stands out from other animal livestock ontologies. Through practical use case applications and a farmer-friendly approach, it focuses on a cattle scenario, providing a standardized framework, species-specific LHO ontology, RDF data generation, and semantic interoperability. By adhering to FAIR guidelines and leveraging semantic web technologies, it enhances data accessibility, connectivity, and advanced analytical reasoning. Additionally, the framework prioritizes the protection and privacy of sensitive data in accordance with GDPR principles. The integration of data visualization tools facilitates data interpretation, empowering farmers and their veterinarians with valuable insights. Ultimately, this framework tackles data acquisition, integration, and analysis challenges, resulting in improved disease surveillance, early detection, and enhanced animal health outcomes.

However, we acknowledge that there are several challenges that need to be addressed in future research. These challenges include refining the ontology and improving its coverage, developing efficient data acquisition and integration mechanisms, enhancing analytical reasoning capabilities, and ensuring the scalability and sustainability of the framework. By tackling these challenges, we aim to continuously improve the effectiveness and practicality of our framework, advancing the field of farm animal health management and contributing to the overall well-being of livestock. Integration of federated learning techniques is another exciting future research within our framework. By incorporating federated learning, we can tap into the collective intelligence and data resources of multiple stakeholders while maintaining data privacy and security. This approach has the potential to further enhance disease surveillance, early detection, and decision-making in farm animal health management systems, ultimately leading to improved animal health outcomes in the agriculture.

6 Acknowledgements

The DECIDE project (https://decideproject.eu/) has received funding from the European Unions Horizon 2020 research and innovation program under grant agreement no. 101000494.

We would like to thank DGZ (https://www.dgz.be/mijn-dgz), The Royal GD (https://www.gddiergezondheid.nl/Diergezondheid/Onderzoek),Pathosen (https://www.pathosense.com/en/howpathosenseworks), ARSIA (https://www.arsia.be/), and Irish Laboratory (https://www.gov.ie/ga/eagraiocht/an-roinn-talmhaiochta-bia-agus-mara/) for their valuable data contributions.

7 References

Ali, W., Saleem, M., Yao, B., Hogan, A. and Ngomo, A.-C. N. 2022. A survey of RDF stores & SPARQL engines for querying knowledge graphs. *The VLDB Journal*, 31, 1–26.

Babcock, S., Beverley, J., Cowell, L. G. and Smith, B. 2021. The infectious disease ontology in the age of COVID-19. *Journal of Biomedical Semantics* 12(1), 1–20.

Bahlo, C., Dahlhaus, P., Thompson, H. and Trotter, M. 2019. The role of interoperable data standards in precision livestock farming in extensive livestock systems: a review. *Computers and Electronics in Agriculture* 156, 459–466.

Bhat, S. A. and Huang, N.-F. 2021. Big data and ai revolution in precision agriculture: survey and challenges. *IEEE Access* 9, 110209–110222.

Bullock, D. S., Boerngen, M., Tao, H., Maxwell, B., Luck, J. D., Shiratsuchi, L., Puntel, L. and Martin, N. F. 2019. The data-intensive farm management project: changing agronomic research through on-farm precision experimentation. *Agronomy Journal* 111(6), 2736–2746.

Chen, X., Jia, S. and Xiang, Y. 2020. A review: knowledge reasoning over knowledge graph. *Expert Systems with Applications* 141, 112948.

Cocca, G. and Lindberg, A. 2019. Overview of data analysis and reporting tools in the context of the SIGMA project. *EFSA Supporting Publications* 16(8), 1690E.

Cowell, L. G. and Smith, B. 2010. Infectious disease ontology,Infectious Disease Informatics. 373–395,Springer.https://doi.org/10.1007/978-1-4419-1327-2_19

Dórea, F. C., Vial, F., Hammar, K., Lindberg, A., Lambrix, P., Blomqvist, E. and Revie, C. W. 2019. Drivers for the development of an Animal Health Surveillance Ontology (AHSO). *Preventive Veterinary Medicine* 166, 39–48.

Fote, F. N., Roukh, A., Mahmoudi, S., Mahmoudi, S. A. and Debauche, O. 2020. Toward a big data knowledge-base management system for precision livestock farming. *Procedia Computer Science* 177, 136–142.

Gangadharan, V. and Gupta, D. 2020. Recognizing named entities in agriculture documents using LDA based topic modelling techniques. *Procedia Computer Science* 171, 1337–1345.

Golik, W., Dameron, O., Bugeon, J., Fatet, A., Hue, I., Hurtaud, C., Reichstadt, M., Salaün, M.-C., Vernet, J. and Joret, L. 2012. ATOL: the multi-species livestock trait ontology. In: Metadata and Semantics Research: 6th Research Conference, MTSR 2012, Cádiz, Spain, November 28–30, 2012. Proceedings 6. Springer, 289–300.

Gómez-Pérez, A. and Corcho, O. 2002. Ontology languages for the semantic web. *IEEE Intelligent Systems* 17(1), 54–60.

Hermans, K., Opsomer, G., Ranst, B. V. and Hostens, M. 2018. Promises and challenges of big data associated with automated dairy cow welfare assessment. In: *Animal Welfare in a Changing World*. CAB International Wallingford, A. Butterworth.

Heward-Mills, D. 2017. Visualizing SPARQL results in tableau. Available at: https://medium.com/virtuoso-blog/virtuoso-tableau-sparql-f9411852a87d (Accessed on 22 June 2023).

Hogan, A., Blomqvist, E., Cochez, M., D'amato, C., Melo, G. D., Gutierrez, C., Kirrane, S., Gayo, J. E. L., Navigli, R. and Neumaier, S. 2021. Knowledge graphs. *ACM Computing Surveys (CSUR)* 54, 1–37.

Hu, G., Do, D. N., Gray, J. and Miar, Y. 2020. Selection for favorable health traits: a potential approach to cope with diseases in farm animals. *Animals* 10(9), 1717.

Khatoon, P. S. and Ahmed, M. 2022. Importance of semantic interoperability in smart agriculture systems. *Transactions on Emerging Telecommunications Technologies* 33(5), e4448.

Kollia, I., Glimm, B. and Horrocks, I. 2011. SPARQL query answering over OWL ontologies. In: The Semantic Web: Research and Applications: 8th Extended Semantic Web Conference, ESWC 2011, Heraklion, Crete, Greece, May 29–June 2, 2011. Proceedings, Part I 8. Springer, 382–396.

Malhotra, A., Younesi, E., Gündel, M., Müller, B., Heneka, M. T. and Hofmann-Apitius, M. 2014. ADO: a disease ontology representing the domain knowledge specific to Alzheimer's disease. *Alzheimer's and Dementia* 10(2), 238–246.

Meyer, A., Faverjon, C., Hostens, M., Stegeman, A. and Cameron, A. 2021. Systematic review of the status of veterinary epidemiological research in two species regarding the FAIR guiding principles. *BMC Veterinary Research* 17(1), 270.

Monteiro, A., Santos, S. and Gonçalves, P. 2021. Precision agriculture for crop and livestock farming—brief review. *Animals* 11(8), 2345.

Murali, E. and Anouncia, S. M. 2022. An ontology-based knowledge mining model for effective exploitation of agro information. *IETE Journal of Research*, 68,1–18.

Newton, J. E., Nettle, R. and Pryce, J. E. 2020. Farming smarter with big data: insights from the case of Australia's national dairy herd milk recording scheme. *Agricultural Systems* 181, 102811.

Niloofar, P., Francis, D. P., Lazarova-Molnar, S., Vulpe, A., Vochin, M.-C., Suciu, G., Balanescu, M., Anestis, V. and Bartzanas, T. 2021. Data-driven decision support in livestock farming for improved animal health, welfare and greenhouse gas emissions: overview and challenges. *Computers and Electronics in Agriculture* 190, 106406.

Niloofar, P., Lazarova-Molnar, S., Francis, D. P., Vulpe, A., Suciu, G. and Balanescu, M. 2020. Modeling and simulation for decision support in precision livestock farming. *Winter Simulation Conference (WSC)*. IEEE, 2601–2612.

Peyre, M., Hoinville, L., Njoroge, J., Cameron, A., Traon, D., Goutard, F., Calba, C., Grosbois, V., Delabouglise, A., Varant, V., Drewe, J., Pfeiffer, D. and Häsler, B. 2019. The RISKSUR EVA Tool (Survtool): a tool for the integrated evaluation of animal health surveillance systems. *Preventive Veterinary Medicine* 173, 104777.

Rajbhandari, S. and Keizer, J. 2012. The AGROVOC concept scheme–a walkthrough. *Journal of Integrative Agriculture* 11(5), 694–699.

Ren, G., Lin, T., Ying, Y., Chowdhary, G. and Ting, K. C. 2020. Agricultural robotics research applicable to poultry production: a review. *Computers and Electronics in Agriculture* 169, 105216.

Rosa, G. J. 2021. *Grand Challenge in Precision Livestock Farming*. Frontiers in Animal Science. 2, 650324. doi: 10.3389/fanim.2021.650324

Roussaki, I., Doolin, K., Skarmeta, A., Routis, G., Lopez-Morales, J. A., Claffey, E., Mora, M. and Martinez, J. A. 2023. Building an interoperable space for smart agriculture. *Digital Communications and Networks* 9(1), 183–193.

Saha, S. 2021. *Towards a Formal Ontology to Support Knowledge Sharing across Product Design and Manufacturing*. Doctoral dissertation, Coventry University.

Sahoo, S. S., Halb, W., Hellmann, S., Idehen, K., Thibodeau Jr., T., Auer, S., Sequeda, J. and Ezzat, A. 2009. A survey of current approaches for mapping of relational databases to RDF. *W3C RDB2RDF Incubator Group Report* 1, 113–130.

Symeonaki, E., Arvanitis, K. G., Piromalis, D., Tseles, D. and Balafoutis, A. T. 2022. Ontology-based IoT middleware approach for smart livestock farming toward agriculture 4.0: a case study for controlling thermal environment in a pig facility. *Agronomy* 12(3), 750.

Van Schaik, G., Hostens, M., Faverjon, C., Jensen, D. B., Kristensen, A. R., Ezanno, P., Frössling, J., Dórea, F., Jensen, B.-B., Carmo, L. P., Steeneveld, W., Rushton, J., Gilbert, W., Bearth, A., Siegrist, M., Kaler, J., Ripperger, J., Siehler, J., de Wit, S., Garcia-Morante, B., Segalés, J., Pardon, B., Bokma, J. and Nielen, M. 2023. The DECIDE Project: from surveillance data to decision-support for farmers and veterinarians. *Open Research Europe* 3, 82.

Vasconcelos Gioia, G., Lamielle, G., Aguanno, R., Elmasry, I., Mouillé, B., De Battisti, C., Angot, A., Ewann, F., Sivignon, A. and Donachie, D. 2021. Informing resilience building: FAO's Surveillance Evaluation Tool (SET) biothreat detection module will help assess national capacities to detect agro-terrorism and agro-crime. *One Health Outlook* 3, 1–13.

Villeneuve, É., Abi Akle, A. A., Merlo, C., Masson, D., Terrasson, G. and Llaria, A. 2019. Decision support in precision sheep farming. *IFAC-PapersOnLine* 51(34), 236–241.

Published by Burleigh Dodds Science Publishing Limited, 2024.