

# Experiences with creating a Precision Dairy Farming Ontology (DFO) and a Knowledge Graph for the Data Integration Platform in agriOpenLink

Dana Tomic<sup>1</sup>, Domagoj Drenjanac<sup>2</sup>, Sandra Hoermann<sup>3</sup>, Wolfgang Auer<sup>4</sup>

## INFO

Received

Accepted

Available on-line

Responsible Editor:

## Keywords:

Semantic Web, linked data, ontology, precision dairy farming, ISOagriNET

## ABSTRACT

One of the central problems in creating information management solutions for precision dairy farming is integration and interpretation of heterogeneous data coming from different equipment and data sources. Establishing a unifying data model is recognized as a cornerstone to such solutions. Here, the challenge lies both in agreeing on a common information context, and in selecting appropriate model representations, model query and update techniques, which guarantee model extensibility. While already existing ISOagriNET Data Dictionaries capture broad variety of the livestock and dairy farming concepts, the existing model representation technics cannot support efficient model extension. In this paper, we present our experience with using the representation and query standards and tools of the Semantic Web, endorsed by the World Wide Web Consortium (W3C), in particular the Resource Description Framework (RDF), the Web Ontology Language (OWL), and the SPARQL query standard, to encode and manipulate the dairy farming domain knowledge in a form of the Dairy Farming Ontology (DFO). Within the research project agriOpenLink the DFO has been created in two phases. The first phase focused on using the Semantic Web tools to facilitate easy encoding and manipulation of the dairy farming domain knowledge and the platform operational data models. The second phase focused on translating ISOagriNET Data Dictionaries into their semantic representations. Resulting DFO is maintained in a semantic repository, and it presents a knowledge graph and as integration backbone for the agriOpenLink decision support platform.

## 1. Introduction

### 1.1. Data Integration Problems in Precision Dairy Farming

Precision dairy farming stands for the data-centric dairy farm management, with tools and procedures that extensively rely on a variety of novel data and data processing techniques (Bewley 2013). In the state-of-the-art experimental and commercial dairy farms valuable data is generated by modern farm equipment, such as automated milking and feeding systems, milking robots and parlors, feed mixers, concentrate feeders, animal health monitoring systems such as pedometers, active ear-tags or rumination collars, weather monitoring systems or specialized environment sensors, and not the least, farmers. Data are also available from specialized sources of the dairying knowledge such as information systems of the control organizations and associations, including the animal breeding registers, milk quality databases, or specialized decision support applications providing advices related to different aspects of dairying, e.g., breeding and feeding.

<sup>1</sup> Dana Tomic  
FTW, Donau-City-Straße 1/3, 1220 Vienna, Austria  
[tomic@ftw.at](mailto:tomic@ftw.at)

<sup>2</sup> Domagoj Drenjanac  
FTW, Donau-City-Straße 1/3, 1220 Vienna, Austria  
[drenjanac@ftw.at](mailto:drenjanac@ftw.at)

<sup>3</sup> Sandra Hoermann  
Josephinum Research, Rottenhauser Straße 1, 3250 Wieselburg, Austria  
[sandra.hoermann@josephinum.at](mailto:sandra.hoermann@josephinum.at)

<sup>4</sup> Wolfgang Auer  
MKW electronics GmbH, 4675 Weibern, Jutogasse 3, Austria  
[wolfgang.auer@smartbow.at](mailto:wolfgang.auer@smartbow.at)

The adoption of Precision Dairy Farming technology is strongly motivated with the models discovered and verified within the dairy farming research studies that reveal correlations between parameters of the environment, herds, and individual animals, including animal health parameters, fertility, nutrition needs and milk yield and quality. However a study about the sensor systems for the health monitoring on dairy farms (Rutten et al. 2013) identified low maturity of existing solutions. The existing functionalities are analyzed along a 4-Level functional model, that integrates the measurement level (Level 1), where the individual parameters are assessed, e.g., activity measurement; the interpretation level (Level 2) , where sensor data is translated into status information, e.g., estrus; the integration level (Level 3), at which sensor information is to be integrated with other systems' information to produce advice, e.g., whether to inseminate a cow or not; and the automation / the decision support level (Level 4) where farmers make decisions based on the system output or the system makes the decision autonomously, e.g., to call the inseminator. A review of 126 publications describing 139 sensor systems, revealed that none of the studies reports about the sensor systems that can be used at Levels 3 and Level 4. These results indicate that considerable additional efforts are needed to transition sensor technology together with models exploiting sensor readings into decision support systems (Level 3 and Level 4 systems) that can be used on commercial farms.

On the other hand, the dairy producers already need to interact with an increasing number of data sources on a regular daily, weekly or monthly bases. This is often perceived as having a negative effect on productivity. For example, the time that farmers spend in inserting, accessing and interpreting information, i.e., in interacting with their systems over many different interfaces, is often perceived as too extensive. Furthermore, in multi-vendor production environments where different product lines support different aspects of the dairy production and the herd and animal monitoring management, the information is often available only for visual inspection and manual integration. Manual interaction is prone to errors, as often the same information has to be entered into several different systems. Consequently, farmers often perceive information overload, and use only a small fraction of the available data (Bewley 2013). For example, from a huge variety of possible reports, only a small number is regularly being checked. The option to manually save reports of interest in formats suitable for further processing, such as comma separated value format (CSV), in order to share them with advisers or process them in the standard spread sheet tools is rarely used.

There is therefore a pressing need to automatically integrate available data within a decision support (Level 4) system that can then provide holistic advices to farmers leading to more efficient herd management, including production and workforce organization, time saving, improving precision of insemination, detection and prevention of health problems, feed rationing adjustment to the reproduction cycle and animal need, and compliance with the milk quality standards and the animal and medication tracking regulations.

The Herd Management Systems aim at reducing the information overload, by providing the farmers with a single point of access to the farm information, typically over a unified graphical user interface (GUI), of a browser- or a desktop-application, or of an App for a Smart Phone a Tablet, or even Google Glasses (Elite 2014). However, there is a huge potential and need for improvement of Herd Management Systems in respect to 1) how they approach integration of different devices and external databases, 2) how they support integration of new knowledge, and 3) what level of customization of the provided advice to the farmer, do they offer. To integrate different systems and knowledge, the HMSs often use a “hard-wiring” approach: here the interfaces of different systems are integrated based on a proprietary data model, and the knowledge, such as management practices, advices, rules, etc., are fix programmed within the software. The practice of “hard wiring” is partially a result of the fact that the existing standards for data exchange on farms are not well adopted.

Indeed, while International Standardization Organization (ISO) offers standards for data exchange at livestock farms, which are currently being also endorsed and popularized by the ISOagriNET<sup>5</sup> industrial association, there are still very few existing implementations. The standards commonly referred to as ADIS/ADED cover both the communication protocol, the syntax and the semantics of the information component. The Data Dictionaries reflect the relational model between the knowledge modelling

---

<sup>5</sup> [www.isoagrinet.org](http://www.isoagrinet.org)

concepts, which encoding is optimized for the efficiency of transmission. It therefore does not account for the need to use unique knowledge representations in order to guarantee reliable disambiguation of the used terms. Further standard ambiguities lead to an unsatisfying level of interoperability of existing implementations. Another important problem is related to the extensibility of the standard knowledge model. In the rapidly growing market for dairy farming equipment, considerable time needed to extend Data Dictionaries with new terms hinders innovation adoption. In case of a new equipment, e.g., a new sensor-based monitoring system, a necessary update to the standard typically takes several years to be accepted. To avoid this, the integration is attempted based on proprietary models and interfaces. Moreover, even the equipment which data model is already accounted for within the current Data Dictionary, often does not implement standardized protocol on its interface. This is typically the case for the equipment integrated within single-vendor monolithic solutions.

The project agriOpenLink (Adaptive Agricultural Processes via Open Interfaces and Linked Services) addresses the problems mentioned here by attempting to close the gap between the implementation-agnostic transmission-centric information representation of the existing ISOagriNET standards (ADIS/ADED), and a vision of a dairy farming data integration platform based on broadly adopted Web standards, capable to support life-cycle management of the common data integration model including its continuous dynamic extension and adaptation. To this end, agriOpenLink exploits the benefits of the semantic extension of the Web, the so called Semantic Web, which common data formats and exchange protocols, transform the Web in a broadly available information management platform (Jepsen 2009). Accordingly we use the standardized model representation techniques, information management paradigms, and existing tools and stacks of the Semantic Web to efficiently cover all aspects of data and knowledge integration, interlinking and sharing in the precision dairy farming. This approach should permit us to translate, extend, enhance and dynamically maintain the domain knowledge already accrued and structured within the ISOagriNET Data Dictionaries, by using the representation techniques and tools of the Semantic Web.

Semantic Web technologies have reached considerable maturity, and have been already successfully demonstrated in different sectors, such as tourism, e-health, production and agriculture (Auer 2014). The World Wide Web Consortium (W3C) offers Semantic Web standards for representing data and knowledge in uniquely referenceable, linked data format with high expressivity, thus accelerating the development of expert services and opening the sector for new data and knowledge providers (Cardoso 2007).

The agriOpenLink project uses semantic and service technology to provide decision support functionality on the top of the ontology-based data integration and interpretation platform (Tomic et al, 2014). It uses existing dairy farming domain knowledge represented as semantic linked data, and extends and enhances it with additional relevant information. This facilitates integration of data from various sources, and encoding of the expert knowledge for decision support applications.

## **1.2. The Structure of the Paper**

Section 2 deals with challenges and approaches towards efficient and flexible data integration. It briefly presents the ISOagriNET standardization framework, introduces the Semantic Web representation and query standards, and provides references to some existing work that adopts ontology-based modelling in agriculture and in particular dairy farming. It also discusses the role of the Dairy Farming Ontology (DFO) within the agriOpenLink platform. Section 3 describes the process of DFO creation in agriOpenLink. Section 4 provides conclusions and future work.

## **2. Challenges and Approaches towards Flexible Data Exchange and Integration on the Farms**

### **2.1. The ISOagriNET Framework for Data Exchange in Livestock and Dairy Farming**

To resolve the need for standardized data exchange on the livestock and dairy farms, the working group ISO/TC 23/SC 19/WG 2 of the International Standardization Organization (ISO) created a collection of specifications also referred to as ISOagriNET ADIS/ADED standardization framework.

These standards specify electronic exchange based on the transmission of ASCII files. The ADIS/ADED interface applies to all systems at the farm including the systems for process control and the automatic or manual data entry (e.g., the milking and feeding systems), systems for farm management and data processing (e.g., programs for herd management and feed calculation) and systems of providers of services (e.g. control associations, breeding organizations and veterinary services).

ADIS – the Agricultural Data Interchange Syntax - describes how data is transmitted; ADED – the Agricultural Data Element Dictionary, describes what data is transferred. The ADIS syntax is defined in the international standard ISO 11787 (1995). The general structure of the ADED database is defined in ISO 11788-1 (1997). The Data Dictionaries for Beef, ISO 11788-2 (2000) and DIN 11800 (1999) have also been elaborated. The communication protocol is defined in ISO 17532 (2007).

In order to stimulate adoption of ADIS/ADED standards the organizations involved in ISOagriNET industrial association, and in particular the German milk control association (LKV-NRW) set up a specialized portal (LKV 2015), which offers information and API interfaces for developers who are implementing applications that integrate ADIS/ADED data exchanges. The portal provides access to a database with yearly published ADED Dictionaries, comprising the General International and German national dictionaries for cattle production, dairy production, pig production, and new extensions. The dictionaries can be queried (via GUI or a web service) and the ADED Data Dictionary files for any specific year can be obtained. Alternatively the description of any specific data element can also be obtained. Currently, the presentation of the data elements is also made available in HTML format, however the HTML documents are created automatically and have dynamic URLs.

The ADED data model uses following data modelling artifacts: (1) data objects (Entities) (2) data elements (Items), and (3) codes (Code Sets). Each of the defined Entities, Items and Code Sets is identified with a unique six-digit ADED number and is described in the corresponding section of the Data Dictionary.

ADED Items describe atomic data properties which values are exchanged (e.g., a unique ID number of the cow, or the milk protein content). An Item can be of a numeric, alphanumeric or of a specific value-restricted datatype described with a specific Code Set. A set of Items form an *Entity*, which is a data object for exchange. An example of an entity is a “Animal Master Data”. An Entity consists of mandatory Items and other optional Items. Each Item, Entity and Code Set is specified with specific features. For the Items these include e.g., name, field length, precision, type, etc. Each Entity is described with its key Items and its optional Items. Each Code Set is specified with a set of acceptable values. The Items, Entities and Code Sets can be either ISO-standardized, nationally defined, or manufacturer-specific, and respectively belong to a specific interval of ADED Numbers. The available ADED numbers are divided in intervals within which different Code Sets, Items and Entities pertaining to different Areas such as Beef production, Milk Quality, Pig production, etc., are defined, are dimensioned so as to guarantee that they are sufficiently large for further extensions. Manufacturers apply for a registered number, to replace manufacturer-specific number. New data elements are included in the national data dictionary, if they cannot be represented by other data elements according to their definition. For example an additional state variable can characterize computation or measurement methods. ADED dictionaries are updated and published on a yearly bases. The partners in ADIS/ADED exchange have to use the same dictionary.

To illustrate the Entity/ Item / Code Set Structure we can examine the Entity “Animal Master Data” (unique ADED ID 880005). The key Item is the “Cattle ID” (900080). It is described as “official (in law) Identification Number on national or regional level”. It is of the numeric data type (N) with 15 characters length. There is 32 possible Items that can be used to describe this Entity. For example “Coat color” (810026) which is described with the CodeSet (810026) with the individual values “other” (0), black (1), brown (2), etc. The Item “Race” (800030) uses the CodeSet with the ADED ID 2. The race is coded in ADR-2 where “Holstein” is identified with the value “01”, “Red Holestein” “02”, “Simmental” “11”, etc.

While ADED provides the semantics for the exchanged data, the ADIS describes syntax of the file which encodes data. An ADIS file consists of ASCII characters and is structured in lines. Each line in

the ADIS file belongs to a specific type which is encoded in the first character of the line. The most important line types are the Definition (D), and Value (V) types, where one D line describes the format of all the V lines that directly follow it. The D line starts with the Entity number followed by a number of triples <Item number, field length, resolution>, one for each Item describing the Entity. The V lines contain the concrete values formatted as specified in the D line. Each ADIS file begins with the header lines (DH, VH) that describe the dictionaries used, the file itself, parties in transmission, checksum etc. The header is important for the decryption of the data of the ADIS file. Here both the sender and the recipient of the ADIS file must use the same version of the data dictionary, which is not transmitted with the data file. The header is followed by normal lines, most importantly (1 x DN, n x VN) blocks.

For the transmission, the Data Dictionary is also described in the ADIS syntax. The Dictionary ADIS file contains the detailed description of the features of all the Items, Entities and Code Sets contained within the dictionary. On the other hand, a Data File in ADIS format includes a number of (1xDN, nxVN) blocks, a block for each Entity including arbitrary number of instances with respective Item values. With its compact format the content of the ADIS/ADED file is optimized for the text-based transmission and exchange between the systems.

For illustration purposes a DN line for the Entity 800050 (already mentioned Animal Master Data) and one corresponding VN line from an ADIS data file are included in Figure 1.

```

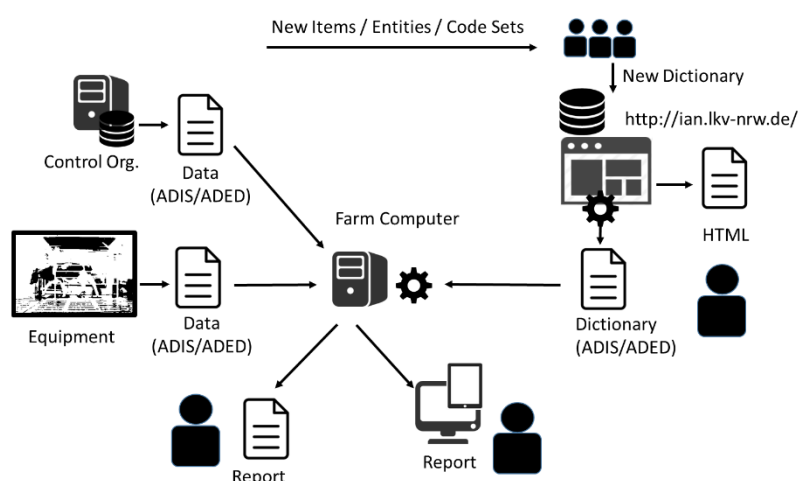
DN880005009900080150009000701500080003002000900053080008000351500080003215000800094150008000950
2000800042010008001110100080011201000800114010008001130200080002601000800096030008000970200083
00011500083011515000900045240
VN880005 | .40000033205422 | ?????????????? | 11 | 20120614 | .40000202042816 | ?????????????? | 276000940
100513 | ?? | ? | 2 | 1 | ? | ?? | ? | ??? | ? | ?????????????? | ?????????????? | HERTA

```

**Figure 1.** An example of a DN and its corresponding VN line from an ADIS data file

The DN line lists all Items (marked bold) which values are included in the VN line. The length and precision of the data representation is also included. Based on that we can interpret the VN line (we inserted vertical bars “|” between the values of different items for easier reading). The VN line describes a cow (800111=sex, value=0, female) of the race Simmental (800030=race, value=11, Simmental), with date of birth June 14, 2012 (900053=date of birth, value=20120614). Where the values are not known the VN line contains “?” characters.

To illustrate the problems with data exchange based on ADIS/ADED concepts, we show in Figure 2 a hypothetical exchange scenario where the ADIS/ADED interfaces to the milking robot and to the milk control organization are implemented at the farm computer.



**Figure 2.** Data Exchange and Interpretation based on ADIS/ADED

In this scenario, the central farm computer interprets the ADIS Data files coming from two sources using the ADED Dictionary file obtained via the Data Dictionary portal. The central application is interpreting data and generating reports. The Data Dictionary provides necessary descriptions of the



Code Sets, Items and Entities and can be used to create a local database in the farm computer with the ADED relational schema. In this scenario, a data integration application running on the farm computer might still encounter problems in respect with data interpretation and integration. Firstly, the ADIS/ADED standards do not provide query logic to filter out entities based on the values of different Items, and therefore data processing or interpretation applications need to save the data in the local database and then query the local database (e.g., with SQL queries). The implementation is therefore vulnerable to changes in the Data Dictionary, as these must be reflected in the database schema, i.e., they lead to the change of the database schema. Secondly, the ADED Dictionary represents an information island as it does not explicitly and uniquely reference descriptions of the used concepts. For example, as we have previously shown, the race is encoded with two characters, and the reference to the unique race description is included only implicitly in the name of the race included in the Data Dictionary. An internal application cannot thus e.g., reach the Web Page describing the Simmental Race, based only and directly on the information in the database. These two problems can be efficiently addressed by using the standards of the Semantic Web.

## 2.2. Semantic Web Technology for Flexible Data Exchange and Integration

The Semantic Web extends the “Web of documents” with the meaning (semantics) through standards that promote common data formats and exchange protocols. The Semantic Web expresses resources and relationships between them in a formalized and structured way, in the so called linked data set, i.e., a knowledge graph. The atomic unit of this graph is a triple which describes the relationship between two resources in the form (subject, predicate, object). Semantic Web uses 1) the Hypertext Transfer Protocol (HTTP) and unified resource identifiers (URIs) for resource identification and access, 2) the Extensible Markup Language (XML) and XML-Schema as a syntax and a content structure of the documents, 3) the Resource Description Framework (RDF) and RDF-Schema for expressing the data objects, their relationships, properties and classes, 4) the Web Ontology Language (OWL) for formal hierarchical knowledge representation based on description logic, and 5) the SPARQL query language and protocol for structured queries on linked data sets over HTTP (W3C SW 2015), (W3C SPARQL 2015).

The Semantic Web standards and tools currently support the whole lifecycle of creating and maintaining large linked data sets that combine both schema (ontologies, vocabularies, ...) and data. The available Semantic Web stack include tools for 1) extraction of structured data from unstructured sources, such as text, 2) linked data authoring or creation via “triplification”, enrichment, interlinking and fusing and 3) data maintenance in a semantic repository – the triplestore (Auer 2014).

A linked data set can be open or freely accessible over the Web, which is referred to as Open Linked Data, or closed, thus accessible only for registered users, e.g., enterprise Linked Data (Mezaour et al. 2014). The enterprise systems that do not open their data can still benefit from the linked data concepts by re-using established semantic descriptions and uniquely defined resources, making it easier to further interlink external and internal information based on the shared semantics. Semantic Web technology thus lowers the entrance barrier for data publishers and users in both open and closed scenarios. So far, the benefits of the Semantic Web technology have already been demonstrated in numerous domains. The result of the Linked Data Community is Open Linked Data Cloud<sup>6</sup> which today include enormous number of data sets (ontologies, vocabularies, and real data) with geographic information, publications, user generated content, government, life sciences, and cross-domain.

The semantic technology has been tested in applications requiring knowledge management in numerous domains such as open government solutions, but also for corporate information systems in eHealth, tourism, media publishing, smart energy (Auer, 2014). Within the agricultural domain the Semantic Web approach already inspired a number of solutions in the research spectrum, e.g., (Eckartz et al. 2013; Chaplinskyy et al. 2013; Gao 2005; Athanasiadis et al. 2009; Grimnes et al. 2012) and is also embraced by the Food and Agriculture Organization of the United Nations (FAO; <http://aims.fao.org>) in their global initiatives for agricultural information management systems (AIMS), AGROVOC vocabulary, and agricultural ontology service (Lauser et al. 2006).

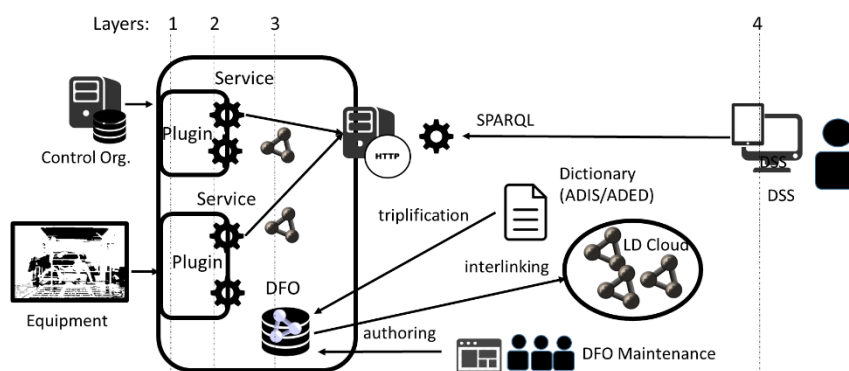
---

<sup>6</sup> <http://lod-cloud.net/>

Two problems in the data integration that have been identified as related to the ADIS/ADED based exchange and have been described previously in Section 2.1, can be efficiently solved by the Semantic Web approach. The first problem of the fixed data schema, and querying of the local database can be solved by using the RDF/XML linked data format and by expressing the schema in the ontology (OWL) format. The flexibility in schema description is the most significant feature that distinguishes semantic technology from the traditional relational database schema modelling. While the relational approaches, e.g., as one for dairy farming presented in (Schulze et al. 2007) can exploit schema-centric optimizations, they are difficult to extend and interconnect. In the RDF knowledge graph, on the other hand, both schema and data are represented as nodes and links in the graph, and accordingly the schema can be flexibly extended by linking new classes and establishing new relationships between them. In addition, the SPARQL based query can be customized to filter specific data and this information is available over the HTTP interface. The second problem of the information island is solved by representations in which each resource is assigned a unique identifier based on URI. In the same time the RDF schema and OWL enable linking of the similar or identical objects together. SPARQL can query both internal and external interlinked data. This way any concept used in describing the data can be uniquely identified and disambiguated. Accordingly the SPARQL query can easily find out external Web Page describing the Simmental Race within the query on the local knowledge graph.

### 2.3. The Role of the Ontology in the agriOpenLink Platform for Dairy Farming

The agriOpenLink project uses Semantic Web technology to provide a data integration and decision support platform for adaptive process control in agriculture. The layered architecture of the platform corresponds to the 4-Level functional model that we introduced in Section 1.1, and is illustrated in Figure 3.



**Figure 3.** agriOpenLink 4-Layer Architecture

The Dairy Farming Ontology (DFO) is a unifying domain knowledge, including the artefacts modeled within the ISOagriNET framework, represented in the form of an extensible knowledge graph that has been created by using RDF and OWL modeling standards. DFO is therefore a central component of the Data Integration Layer (Layer 3). As illustrated in Figure 3, it is used both by the Measurement and Interpretation Layer (Layer 1 and Layer 2) and the Decision Support Layer (Layer 4). The Layer 1 and 2 are implemented by the means of the so called Plugins. A Plugin is a software components that wraps the data-exchange interfaces between the agriOpenLink platform and a farm device (e.g., a milking robot in Figure 3) or an external data source (e.g., a milk quality organization in Figure 3). The equipment publishes specific data on demand (on the query). The data is expressed in a semantically unified format (RDF/OWL) and are semantically aligned with the DFO. More specifically, each Plugin implements specific Semantic Web Services - HTTP REST services - that return data in the RDF Format. These Plugin Services translate the equipment data from the equipment-internal proprietary data format into data instances of the corresponding DFO Classes encoded in the RDF/OWL format. The Plugin Services are therefore the corner stone of the Interpretation Layer (Layer 2). At the Decision Support Layer (Layer 4) the knowledge about the relevant correlations between the heterogeneous data

properties (e.g., the knowledge that a cow in heat has higher activity, lower feed intake) is encoded into complex SPARQL queries. These queries define search and filtering patterns over linked data objects where the linking structure is provided by the standardized classes and properties of DFO. DFO specifies also the so-called defined classes –specified with restrictions on various properties - for the purpose of automatic classification of data by means of automated reasoning. Hence, each complex SPARQL query is resolved within the platform by successively collecting data from identified Plugin Service, reasoning on the collected data, filtering results by simple SPARQL queries, invoking further Plugin Services, and so on until the final result is obtained. The final result satisfy the initial complex SPARQL query constraints. The example of the query is a search of a lame cows among all the cows in the second lactation with a specific milk yield. The complex query resolves the definitions of the “lame cow”, and the “2<sup>nd</sup> lactation” cow and filters the instances of these cows based on the milk yield data.

The major benefit of the DFO-based data exchange and integration approach is in the extensibility of the DFO as a knowledge graph. In Figure 3 we depicted sources of the data that can continuously enhance the DFO. First of all, the existing domain knowledge (e.g. ADIS/ADED Data Dictionary) can be made a part of DFO by means of triplification (as we describe later in Section 3) Secondly, external knowledge graphs (e.g., already mentioned AGOVOC dictionary of FAO) can be inter-linked with DFO. Thirdly, the necessary extensions to the data model which potentially comes with new equipment that produces new data can be authored by the interested vendors, who can insert new descriptions in DFO, and use them to implement their corresponding Plugins. The life-cycle management of DFO starts with the initial engineering and making the decisions on which concepts shall be included within the ontology. The next step is establishing the tools for the DFO life-cycle management and identifying the clear roles of all possible DFO editors in this process. Section 3 describes the experiences with the DFO engineering that we obtained within the agriOpenLink project.

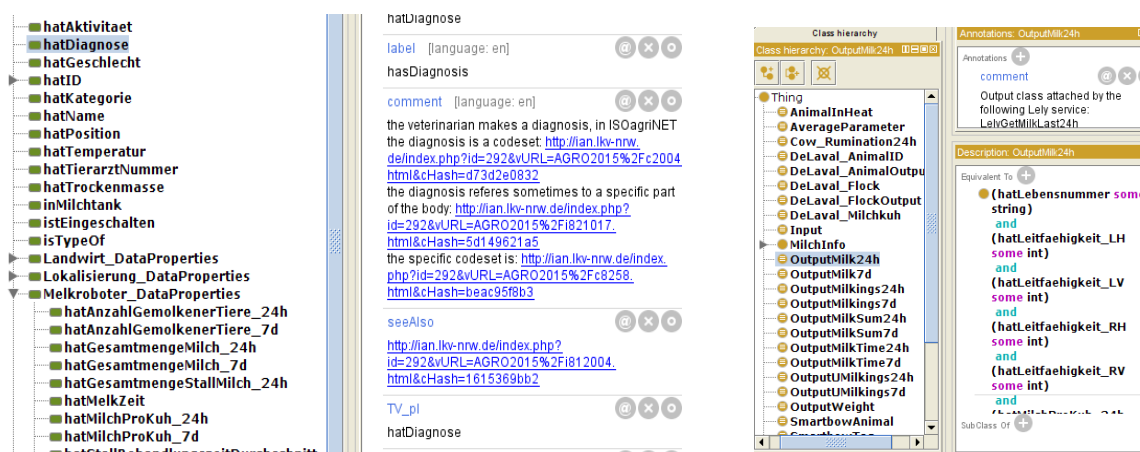
### **3. Experience with the DFO Engineering**

#### **3.1. 1<sup>st</sup> Phase - Domain Analysis and Operation Support**

The creation of the Dairy Farming Ontology in the agriOpenLink project was performed in two phased. The 1<sup>st</sup> phase focused on the benefits of using the ontology editing tools, in particular the OWL ontology editor Protégé (Protégé 2015) in facilitating the analysis of the dairy farming domain models, and in encoding of the domain knowledge. The goal of this phase was to include all relevant concepts into DFO so as to be able to collect initial repository of the animal master data for the farm. Accordingly this initial DFO provides the basis for collecting data from the herd management applications, or animal registers. We further concentrated on the scenario where this initial knowledge graph is constantly extended with the data obtained from different farm equipment and external information sources via Plugin services. These data describing the milk quality, feed quality and quantity, activity, fertility, and health information. We started with a small set of OWL classes including Animal, Farm, Farmer, Equipment, Organization and with the object properties that reflected specific relationships such as parent-child relationship between Animals, similar to the approaches (Gao 2005; Seamless 2015).

In the process of DFO engineering for the domains of milk quality, feeding, breeding, fertility, both the ADED Dictionaries and the schemas of the milking robots, concentrate feeder, and heat and activity monitoring equipment have been analyzed. Consequently for the included properties the annotations about which equipment or external data sources created data – an information missing in the ADED model- was added in form of specific annotations. In addition, the classes and data properties that were included in DFO based on their existence in the ADED dictionary were annotated referencing the definitions in the ADED data portal. Figure 4 (a) show a screen shot of DFO in Protégé Editor.





**Figure 4.** 1st phase DFO a) including comments to reference ADED, b) focus on defined classes for Plugin Services (Protégé Editor)

The property “hasDiagnose” is annotated with the link to a specific HTML page defining this property in already mentioned LVW Portal. However, the problem with this approach to referencing arises due to the fact that these HTML pages do not have unique URI but are periodically re-generated and new URIs are created in this process. Accordingly it became clear that ADED Dictionaries need to be available in the linked data format in order to reference ADED concepts in a formal way. Consequently, we defined ADED “triplication” as a goal for the 2<sup>nd</sup> phase of the DFO Engineering.

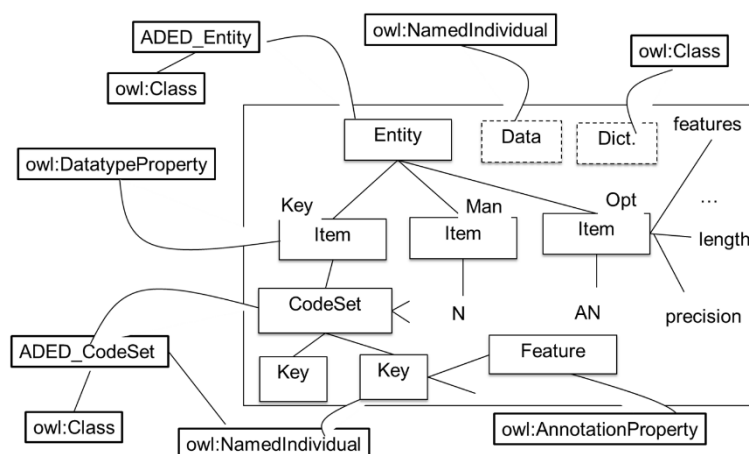
Finalizing the 1<sup>st</sup> phase we revisited the requirements of our initial scenario of a growing knowledge graph. This scenario still presents a large number of challenging issues related to an unclear business model of the data storage on the farm. Consequently we re-focused the DFO engineering on the scenario in which SPARQL queries coming from Decision Support Systems (DSS) trigger search, access, interlinking and filtering of semantic data available via Plugin Services. The resulting DFO is a minimal collection of properties that belong to the animal master data set that can be extracted from equipment and external sources.

DFO is also extended with a number of defined classes that are used by Plugin Services. More precisely, each Plugin Service consumes a property of a specific defined Input Class and creates an Instance of a specific defined Output Class. While the input class is for example the Animal Class identified with the “unique animal number”, the Output Class describes properties that can be extracted from the equipment by means of the Plugin Services. Figure 4 (b) shows a screen shot of DFO in the Protégé Editor with a set of defined Output Classes of the milking robot Plugin. The selected Output Class of one Plugin Service is defined as including the properties “unique animal number” and four “milk conductivity” properties for each of the cow’s tits. In this phase DFO is hence extended with the properties and classes that support the platform operation, particularly in our demonstrator that included Plugins of two milking robots, of a concentrate feeder and of an animal monitoring system.

### 3.2. 2<sup>nd</sup> Phase - Triplication of the ADED Data Dictionary

Triplication stands for translating the data from some external format into the RFD data graph. Atoms of this graph are triples comprising the subject, the object and the relation. A triple can describe schema, e.g. linking two classes with their relationship, or can include data, where subject is a data instance, the object is a specific value, and they are linked by a specific data property of the instance that holds this value. We developed a triplifier tool that translates an ADED Dictionary into Linked Data using RDF/OWL standards. We followed the best practices for creating the linked data (Mezaour 2014), and automatically created HTML documents to support publishing of the newly defined URI. Our triplifier tool also transforms ADIS Data Files into linked data.

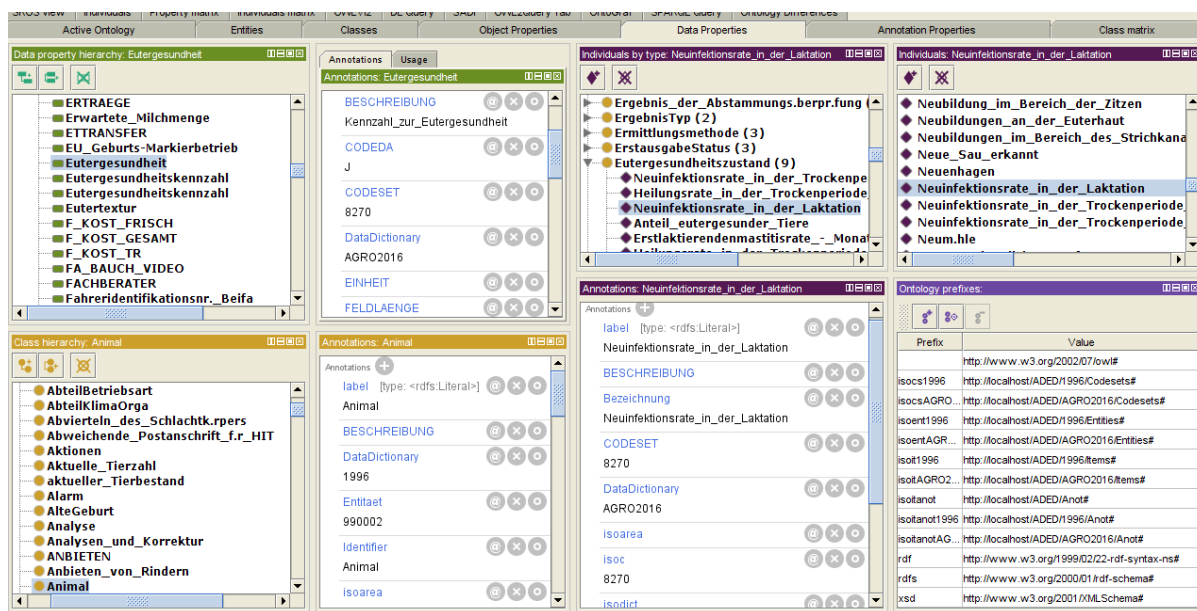
The triplication of the ADED schema is illustrated in Figure 5.



**Figure 5.** Mapping of ADED Entities, Items and CodeSets into DFO in OWL

First the OWL Classes `ADED_CodeSet`, and `ADED_Entity` are defined. The features of `CodeSet`, `Entity` and `Item` (e.g., `NAME`, `IDENTIFIER`, `SYNONYME`, `BESCHREIBUNG`, `KOMMENTAR`,...) are mapped into OWL `AnnotationProperties`. Additional `AnnotationProperties` are defined to capture ADED Number for easy search, ADED Dictionary name, and area (Cattle, Milk, General,...). Each ADED Code Set is translated into a subclass of the `ADED_CodeSet`. For each specified value of the `CodeSet` a unique OWL `NamedIndividual` of this Class is created. The Class and the `NamedIndividuals` are annotated based on respective `CodeSet` features. Each ADED `Item` is translated into a unique OWL `DatatypeProperty`, and is annotated based on the values of the defining features using respective OWL `AnnotationProperties`. Each ADED `Entity` is translated into a unique subclass of the `ADED_Entity` Class. It is annotated based on the values of the defining features using respective OWL `AnnotationProperties`. The relationship among an `Entity` and its defining `Items` is captured in a respective `AnnotationProperty`. Each `Code Set`, `Item` or `Entity` is annotated with its ADED Number, and ADED Dictionary Name. A `Named Individual` is created to represent an ADED Dictionary, for easy search. Finally, to translate the ADIS Data File into linked data, the values of all the `Items` captured in VN lines for particular `Entities` are processed in the following way: An `Entity` VN line translates into an OWL `NamedIndividual` with all the datatype properties (for corresponding `Items`) initialized with the respective values. To test the procedure we triplified all dictionaries available via <http://ian.lkv-nrw.de> as well as exemplary ADIS Data File.

Figure 6 shows a set of owl data properties, classes, and code set values in the triplified ADED Dictionary for 2016 in Protégé. The selected OWL Class `Animal` (Figure 6, b) is annotated with the ADED Number of its originating ADED Entity (990002), with an ADED identifier “`Animal`”, which is also used as an OWL Class Name. A selected OWL data property “`Udder Health`” (Figure 6, a) is annotated with the ADED features including the `CodeSet` (8270) which defines all possible values of this property. Figure 6 (c) shows how the `CodeSet` 8270 is mapped into the OWL Class labeled with “`Udder Health Status`” and its keys into OWL `Named Individuals`. For example, a selected OWL `Named Individual` that belongs to the `CodeSet` “`Udder Health Status`” is labeled with “`The rate of new infections in the lactation period`”. This OWL `Named Individual`, as well as all other resources contained in DFO, have unique URL which can be used as global references. By means of further annotations any of these resources can be interlinked with an additional information that can be useful in searching in the knowledge graph.



**Figure 6.** ADED Dictionary (2016) in Protégé Ontology Editor: (a) OWL data property “Udder health”, (b) OWL Class Animal, (c) OWL Named Individuals for the values of the CodeSet “Udder Health status)

#### 4. Conclusions and Outlook

In this paper we presented a motivation for, and the two phases of the ontology engineering process in which the Dairy Farming Ontology (DFO) has been created in the project agriOpenLink. The agriOpenLink platform uses DFO as a unifying description of the heterogeneous dairy farm data for efficient data interpretation and integration. DFO is a knowledge graph which can be further enhanced with links to an external body of agricultural knowledge which already exists in the linked data format, making it assessable for the decision support applications. To use the existing livestock and dairy farming knowledge we proposed the mapping of the ADIS/ADED model into an ontology and we implemented a translation tool – a triplifier. We are currently preparing a SPARQL-endpoint for experimental accessing and manipulation of the DFO. Accounting for the fact that the DFO is not a static structure, agriOpenLink platform also integrates ontology editing and maintenance tools. The next steps will focus on data enrichment and interlinking and in exploiting their benefits in the design of the decision support tools.

#### Acknowledgment

The work presented here is partially funded by the Austrian Research Agency (FFG) within the project agriOpenLink (research fund 898398).

#### References

- Athanasiadis IN, Rizzoli AE, Janssen S, Andersen E, Villa F. 2009. Ontology for Seamless Integration of Agricultural Data and Models. MTSR 2009; Milan, Italy, October 1-2, 2009, 282-293. doi: [10.1007/978-3-642-04590-5\\_27](https://doi.org/10.1007/978-3-642-04590-5_27)
- Auer S., 2014. Introduction to LOD2, in Linked Open Data – Creating Knowledge Out of Interlinked Data, 1-17, Springer, 2014. doi: [10.1007/978-3-319-09846-3\\_1](https://doi.org/10.1007/978-3-319-09846-3_1)
- Bewley, J.M. 2013. Exciting Dairy Breakthroughs: Science Fiction or Precision Dairy Farming? Proceedings of the Precision Dairy Conference, Rochester, MI, June 26-27, 2013.
- Cardoso, J. 2007, The Semantic Web Vision: Where Are We?, IEEE Intelligent Systems, Sept/Oct 2007, 84-88. doi: [10.1109/mis.2007.4338499](https://doi.org/10.1109/mis.2007.4338499)

- DIN 11800, 1999. Agricultural engineering - Electronic data interchange in agriculture - data interchange in the field of cattle production, available at <http://www.beuth.de/>
- Eckartz, S, Verhoosel, J, Folmer E and E Somers 2013. Semantics for Smart Dairy Farming: a milk production registration standard v1.0, Report, available at <http://www.smartdairyfarming.nl>, last accessed May 2015.
- Elite, 2014 'Drei Programme für die Organisation Ihrer Herde', in German, Elite Magazin 6/14, 2014.
- Gao Yongchun, 2005. The application of Web Ontology Language for information sharing in the dairy industry, PhD Thesis, McGill Univ. Montreal. 2005.
- Grimnes MK, Abufouda, M and A Schröder. 2012, Semantic Integration through Linked Data in the iGreen project. GIL Jahrestagung, volume 194 of LNI, page 107-110.GI,(2012)
- ISO 11787, 1995. Machinery for agriculture and forestry - Data interchange between management computer and process computers - Data interchange syntax, available through <http://www.iso.org>. doi: [10.3403/00655990](https://doi.org/10.3403/00655990)
- ISO 11788, 1997/2000. Electronic data interchange between information systems in agriculture - Agricultural data element dictionary - Part 1: General description Part 2: Dairy farming, Part 3 Pig farming
- LKV, 2015 'ISOagriNet, <http://ian.lkv-nrw.de/index.php>, <http://ian.lkv-nrw.de>, <http://webapp.lkv-nrw.de/AdedDataDictionary>, last accessed Mai 2015
- Jepsen, T. C., 2009. Just What Is an Ontology, Anyway? IEEE IT Pro, Sept/Oct 2009.
- Lauser, B., Sini, M., Liang, A., Keizer, J. and S. Katz, Stephen, 2006. From AGROVOC to the Agricultural Ontology Service / Concept Server, available at <http://aims.fao.org>
- Mezaour, A-D., Van Nuffelen, B., and C. Blaschke. 2014. Building Enterprise Ready Applications Using Linked Open Data, in Linked Open Data – Creating Knowledge Out of Interlinked Data, Springer, 2014.
- Protégé, 2015, <http://protege.stanford.edu/>, last accessed Mai 2015
- Rutten , C. J., Velthuis , A. G. J., Steeneveld, W. and H. Hogeveen. 2013. Invited review: Sensors to support health management on dairy farms, J. Dairy Sci. 96 :1928–1952. doi: [10.3168/jds.2012-6107](https://doi.org/10.3168/jds.2012-6107)
- Schulze, C., J. Spilke, and W. Lehner. 2007. Data modeling for PDF within the competitive field of operational and analytical tasks. Comput. Electron. Agric. 59(1-2):39-55. doi: [10.1016/j.compag.2007.05.001](https://doi.org/10.1016/j.compag.2007.05.001)
- Seamless , 2015. <http://ontologies.seamless-ip.org/livestock.owl>, last accessed Mai 2015
- Tomic SDK, Hoermann S., Handler F., Wöber W., Otte, M., and W. Auer, 2014. agriOpenLink: Semantic Services for Adaptive Processes in Livestock Farming, AgEng 2014.
- W3C Semantic Web, <http://www.w3.org/standards/semanticweb/>, last accessed: May, 2015.
- W3C SPARQL 1.1, <http://www.w3.org/TR/sparql11-overview/> last accessed May 2015.