Toward a Dairy Ontology to Support Precision Farming

Victor Fuentes[1,2], Tomas Martin[1], Petko Valtchev[1], Abdoulaye Baniré Diallo[1,2], René Lacroix[3] and Maxime Leduc[2]

[1]CRIA, Département d'informatique, UQÀM, Montréal (QC), Canada
[2]LACIM, Département d'informatique, UQÀM, Montréal (QC), Canada
[3]LactaNet, Sainte-Anne-de-Bellevue, Canada

Abstract

Precision farming is about improving farming processes through in-depth analysis of the generated data. Dairy farming, in particular, is being intensively computerized and hence a fertile soil for such applications. In our own project, we investigate the benefit of data analytics in optimizing dairy production. To that end, the Valacta centre of expertise shares a dataset recording the performances of dairy cows and farms in Eastern Canada. Here, we tackle the design of a domain ontology (DONT) on top of it. The *dairy cattle performance ontology* (DCPO) reconciles the complex structure to the heterogeneous nature of dairy data within a unified framework that ensures extensibility to external data. It also provides a common vocabulary for both stakeholders and automated knowledge management tools, and, in the longer term, should support explainability for predictive neural models. We present here the bottom-up process of DCPO design and summarize its current content. We also illustrate its present and future usages.

Keywords

Precision agriculture, Dairy farming, Domain ontologies, Knowledge discovery from data, Graph mining

1. Introduction

Agriculture 4.0[1] refers to future trends helping the sector face the main challenges pertaining to the demands of the future: demographics, scarcity of natural resources, climate change, etc. It puts a special emphasis on precision agriculture, the internet of things (IoT) and the use of big data to drive higher business efficiencies. Precision farming, in particular, is a very active area for both research and technology transfer [1, 2, 3, 4]. It is about improving the overall farming process through in-depth analysis of its various aspects as reflected in their data imprint, i.e. historical data generated by farming devices, produce/crop processing entities, regulatory bodies, etc. This requires all the stakeholders (e.g. producers, managers, analysts, consultants, etc.) to work together to leverage available data as a competitive advantage.

A typical approach is the design of machine learning or data mining-based analytical tools

International Conference on Biomedical Ontologies 2021, September 16–18, 2021, Bozen-Bolzano, Italy
fuentes.victor_eduardo@courrier.uqam.ca (V. Fuentes); martin.tomas@courrier.uqam.ca (T. Martin); valtchev.petko@uqam.ca (P. Valtchev); diallo.abdoulaye@uqam.ca (A. B. Diallo); rlacroix@lactanet.ca (R. Lacroix); maximeleduc@gmail.com (M. Leduc)
0000-0001-7037-7565 (V. Fuentes)

[1]https://www.worldgovernmentsummit.org/api/publications/document?id=95df8ac4-e97c-6578-b2f8-ff0000a7ddb6

to, *inter alia*, predict outcomes in daily-life situations the stakeholders face or to detect major trends and/or exceptional events in the data. As living beings are involved, data are typically heterogeneous and complexly structured: They may cover such aspects as the well-being and health issues for farming animals, nutrition, yield, genetics, etc. Inner structure, e.g. time series, and inter-record relations, e.g. animal pedigree, would also appear in the data.

Constituting such complexly-structured datasets requires a significant data-modelling effort. Moreover, as ever more aspects of the farming process get computerized, extensibility to further datasets is often a prime concern. This motivates a full-scale domain modelling in the form of a dedicated domain ontology (DONT). DONTs have a wide range of benefits beyond mere rich/extensible data schema. For instance, they provide a standardized vocabulary to support stakeholder collaboration while representing a centralized repository for domain expertise, thus enabling the design of decision-support systems for various domain tasks [5].

We tackle the design of a DONT for dairy cattle performance evaluation within a larger project on optimizing the production process in Canadian dairy farms. In the context of ever increasing competition and the anticipated reform/abolishment of the current quota system, it is crucial to provide the necessary decision support systems to dairy farmers to help them adapt to the new realities. To that end, a large corpus of data about milk production and milk control (components such as fat, protein, urea, lactose, etc.), that was gathered for the last two decades is to be leveraged. More specifically, predictive models should help anticipate various metrics of the production process on both single cow and whole herd levels while further analytical tools are intended to pinpoint typical farming practices as well as outlying animals/farms.

While dairy farming has been targeted by at least two prior ontology-designing exercises [6, 7], the resulting DONTs collide with our goals. First, their broader coverage of the production process mismatches our focus on milk control. Second, our starting point is an existing dataset with a partially available schema that provides both guidelines and limits to domain modelling. Third, we need to support a variety of analytical tools by means of a rich vocabulary to enable the expression of regularities/anomalies in the data at various abstraction levels.

We present here the design of our *dairy cattle performance ontology* (DCPO), its current state and intended usage. We also discuss the original aspects of both the process and its end-product, DCPO, and illustrate the regularities it allows us to mine. The remainder of the paper is as follows: Section 2 presents our motivations while section 3 lists relevant prior work. Next, section 4 details our iterative modeling process and our tool set. Finally, section 5 concludes.

2. Motivation

Valacta[2] is the Dairy Production Centre of Expertise covering the province of Quebec and the Atlantic regions of Canada. Its core business is to improve the profitability and the sustainability of dairy farms by helping the producers with various aspects of technico-economic performance of their herds and its management. *Valacta*'s employees provide services to 4,500 dairy farms.

The accumulated data about dairy production and milk control represents more than two decades (first records date back to 1998). It describes 6,670 herds and 1.5M cows, over periods of varying duration, yet the most rigorously recorded data covers the decade ending in 2017. Indeed,

[2]http://www.valacta.com/

just as farms can enter and, more rarely, exit the Valacta-controled set of farms, individual cows can move between herds and farms till they definitely exit the controlled livestock. Key concepts reflected in the data include milk control samplings and the associated laboratory-based analyses that estimate the principal milk components: fat, protein, milk urea nitrogen, somatic cells, lactose, etc. *Milk controls* are performed roughly on monthly basis during the milk-producing part of a cow's life cycle, the *lactation*. The latter is a periodic process starting with a *calving* and typically ending with the cow getting dry (no more milk). Lactations can also be curtailed, i.e. end abnormally, due to low productivity, health issues, production quota concerns, etc. A lactation splits into *early*, *mid* and *late* stages, each ca. 100-day long. The latter covers a major milestone, the *305th day*, marked by the cumulative values of milk components. Just like lactation, the herd membership for a cow can be ended for a variety of reasons, inclusive death or exit from the controlled livestock, which are divided into *voluntary* and *involuntary*.

Overall, the records provided by Valacta amount to 3+ billion data end points. This huge dataset hides potentially *meaningful concepts*, e.g. unproductive cows admitting improvement vs those to quickly sell, and *behavioral patterns* for cows or farmers, that need to be uncovered. In order to allow richly-structured heterogeneous datasets to be: (1) properly built and (2) analyzed to yield meaningful and intelligible patterns, we decided to design a DONT.

A number of our dairy analytical tools are symbolic-level, inclusive a suite of graph mining methods whose cornerstone is a DONT-powered generalized pattern miner. This novel pattern flavor, introduced in [8], is illustrated in Figure 5. Additionally, a set of predictive models exploiting deep neural net architectures have been designed targeting a variety of yield metrics such as milk production and overall cost [9]. The way these can benefit from the ontology and the graph mining tools' output is currently under investigation.

3. Related Work

A variety of ontological sources have been developed that pertain to dairy production and livestock. For instance, the *Animal Trait Ontology for Livestock* (ATOL)[3] models phenotypical animal traits of livestock. These are represented from an environment-aware and animal breeding-driven point of view. The stated goal behind ATOL is to support database design, fine-grained domain modelling and semantic analyses. A *Common Dairy Ontology* (CDO) [6] has been designed towards assisting on farming decision making and semantic alignment [4]. Additionally, it was provided with suitable similarity metrics and other measures. Yet CDO is primarily focused on sensor data and lacks a transverse view of the domain (e.g. nutrition, health, environment, etc.). *AgroRDF* [10] is a data exchange standard designed for agro-industrial purposes and built with semantic technologies. However, it lacks a unifying broader framework able to precisely describe the dairy domain. The *agriOpenLink* [3] system provides open interfaces and linked services to enable the development of new processes with a plug-and-play architecture. The *Dairy Farming Ontology* (DFO) is among the many created within the agriOpenLink project. Albeit strongly appealing for our own goals, it is not publicly available. The FAO (Food and Agriculture Organization of the United Nations) project develops agricultural

[3]https://www.ebi.ac.uk/ols/ontologies/atol
[4]http://www.smartdairyfarming.nl

standards such as *AgroVOC* vocabulary [11]. While it covers a wide range of subjects (e.g. food, nutrition, agriculture, fishing, etc.) it lacks middle-level concepts involved in dairy production, hence it is too generic for our needs.

In the recent past, DONTs have been used to support a semantically rich data mining process. Indeed, they expose domain knowledge to machine processing while providing a rich vocabulary that is easily intelligible for domain experts [12]. Pattern mining [13, 14] aims at discovering recurrent data fragments in a dataset that might represent potentially useful trends and regularities (combinations of descriptors). Depending on data record topology and how much thereof is preserved in the patterns, various flavors of patterns have been studied, from itemsets (sets of products) to sequences to graphs. Independently, *generalized* patterns [15] have been introduced to deal with cases where abstracting from concrete data items (e.g. *Corona virus* instead of *SARS-CoV-2*) can bring insights absent in the ground level of data records. Generalized patterns are defined on top of an item taxonomy.

Graphs are among the most challenging pattern formats and adding a DONT on top of their vertex and edge labels further compounds the issue. Partial solutions to the graph mining with a DONT problem were investigated in [16, 17, 18]. Both [17] and [18] under-exploit the ontological structure by focusing only on parts of it (object properties and classes, respectively). In comparison, our DONT is intended to support abstraction on edges as well, e.g. use *parent* property in patterns to match the *dam* property (female parent of a bovine) in data. In [16], abstraction from both vertices and edges was formalized, yet for graphs built around a vertex sequence which largely eases the mining task. In contrast, we deal with unrestricted graphs.

Finally, the problem of feeding the knowledge from a DONT into a neural learning process was approached in [19] with class-embedding-based techniques. Prior studies have investigated mimicking the ontological structure by the neural network architecture [20, 21]. Unlike these, we rely on discovered graphs patterns for data augmentation [8].

4. Building the Dairy Cattle Performance Ontology

We were provided with several non ontological resources such as datasets of various provenance and coverage pertaining to dairy production, together with their data dictionaries. These covered milk production, quality control, genetics, etc., with records for 1.5M cows, 6.67K herds and 10+ years of milk tests. Additionally, we followed the International Committee for Animal Recording (ICAR)[5] guidelines that establish definitions, guidelines, rules and standards for (1) identifying animals, tracing their parentage, recording their performance and evaluating their genetics, and (2) identifying characteristics of production systems and their bearing on animal health, care, productivity, food safety and the environment. Moreover, the calculation and publication of all dairy cattle genetic evaluations in Canada is the responsibility of the Canadian Dairy Network (CDN)[6], whereby the data and dictionaries are publicly available. Starting from all these resources, we apply an iterative modeling process inspired by the Ontology Summit 2013 Communiqué's life cycle[7]. Below, we describe its main steps and their outcomes.

[5] https://www.icar.org/
[6] https://www.cdn.ca/
[7] http://ontolog.cim3.net/OntologySummit/2013/communique.html

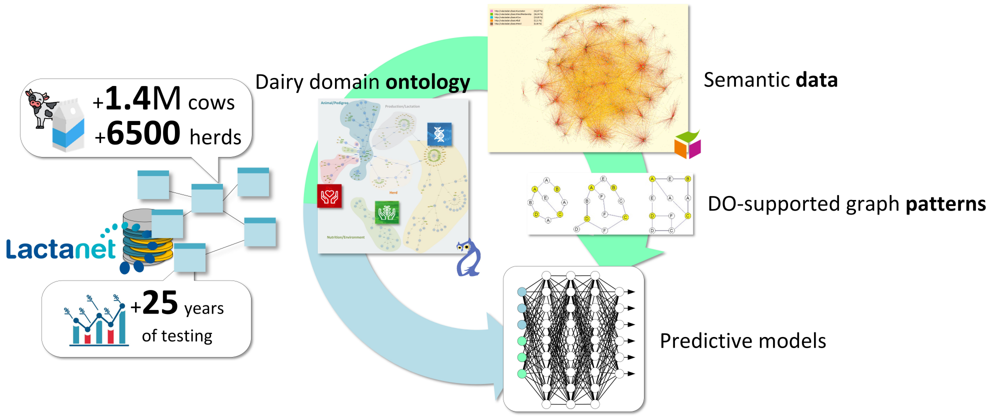

Figure 1: Global project architecture

4.1. Requirements

Our initial focus was on purpose and scope of the ontology. Precisely speaking, DCPO is intended to support three use cases: 1) power unsupervised data mining tools in order to extract interesting, actionable and previously unknown insights that could further support predictive machine learning; 2) enable user-formulated cross-domain queries over genetics, nutrition, health, etc. datasets; 3) facilitate dataset federation when exploiting external data sources.

Figure 1 depicts the global architecture of the information system the ontology is intended to support. The DCPO, in the center, plays the role of: (1) a federated schema for external data sources on the left, (2) a semantic interface to query the linked data produced from the integration of external and internal sources, and (3) a knowledge base for graph mining algorithms on the right side. The produced generalized graph patterns are injected into further machine learning tools as additional ontology-based features to make their results more intelligible.

A number of high-level requirements are drawn from the above scenarios. First, the ontology should be capable of integrating a series of datasets representing the dairy data given as starting point knowledge, but flexible and standardized enough as to represent an interface to federate any new dairy dataset. Second, it should provide enough low level detail as to satisfy data requirements of data mining algorithms, and third, albeit modular, provide the minimal necessary inter-domain connections to perform complex queries. As the results of this ontology supported system are intended to be used by anyone in the dairy community, it is of utmost importance to select a vocabulary all stakeholders feel comfortable with. To that end, a key further requirement is compliance to international standards in the field, in particular to the ICAR guidelines vocabulary and CDN's genetic data format.

4.2. Scope

Precision dairy farming is about optimizing the dairy production process. To that end, it uses a variety of technologies in measuring dairy cattle indicators of significant impact on their performance. Such indicators reflect complementary aspects of the dairy process that have to be encoded in our DCPO. More precisely, in the long term, DCPO will embody knowledge across six distinct perspectives over dairy production: breeding (pedigree), genetics, production and milk quality control, environment, health and nutrition. Currently, it encompasses only the first four due to the limited availability of datasets and data dictionaries to be used as departing point. Note that additional information about entities such as farms (e.g. location or cleanliness) or farmers (preferences, history, etc.), even though appealing, are out of scope here since not yet properly formalized and hence not recorded.

4.3. Ontological Analysis and Design

To understand the dairy farming field, we started by frequent interactions with our domain experts, inclusive some already experienced data scientists. The ontological analysis for DCPO has been further guided and simplified by the available structured description of the dairy data recording procedures within the ICAR documentation. Even if it does no amount to a specification of the dairy process, it is informative enough as to provide a skeleton for our own work at the most detailed data level. In general, each ICAR guideline provides definitions for the most relevant terms describing the data to record, a minimal set of attributes to be recorded for each particular trait and an optional set of attributes with extra information for improved recording. Additionally, the rules and recommendations on how the data should be captured are a source of terms and knowledge for identifying the relevant entities in the field. ICAR and CDN standardized terms are important resources, as defining a core set of entities in an intuitive, shared vocabulary is fundamental to achieve a good communication in the interdisciplinary team. This core will be gradually enriched with lower level concepts and properties and aligned to an upper ontology to provide a foundational theory.

To take advantage of the available resources mentioned before, a bottom-up approach is performed. To identify the key entities of the ontology, we first extract the names of our datasets and their columns from available data dictionaries and match them to the terms defined in the ICAR documents, so that we could link them to the standardized dairy domain terminology, Figure 2 illustrates this process. On the left, the candidate term *Lactation* is retrieved from the dataset name and it's matched to the terms defined in the standards document, where an occurrence is found. In the matched definition, the related candidate terms/phrases *Calving* and *DryPeriod* are retrieved. Additional examination of the document identifies another candidate term *ProductionPeriod* and its relationship to the other terms are inferred (e.g. *hasLactation*). To keep track of the process, one use of annotations was to mark classes and properties with the name of the original field in the data dictionary (e.g. $ANIMAL\_ID$, $HERD\_ID$). This is useful for both documentation and conversion between semantic and relational data formats.

In defining hierarchies (i.e. classes and properties) we usually provide an abstract level to factor out the common characteristics of the elements in a particular module, and one or more specialization levels below which inherit and refine these characteristics. This facilitates the management of the overall ontology architecture and inter-module connections, its extension, readability, and better grouping of similar entities. In some cases the generalization process leads to the finding of *ontology patterns* (OP) that can then be reused across the ontology to provide modularity and even be used in other projects as readily available design solutions. One such OP is the *ascertainment pattern* shown in figure 3 (left). In detail, a target, *Thing*, undergoes

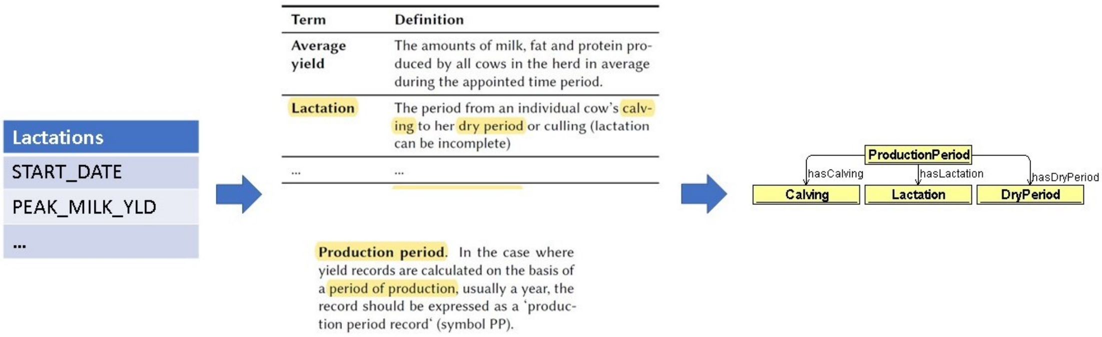

Figure 2: Bottom-up analysis: From table/column names, via ICAR term definitions, to DCPO entities.

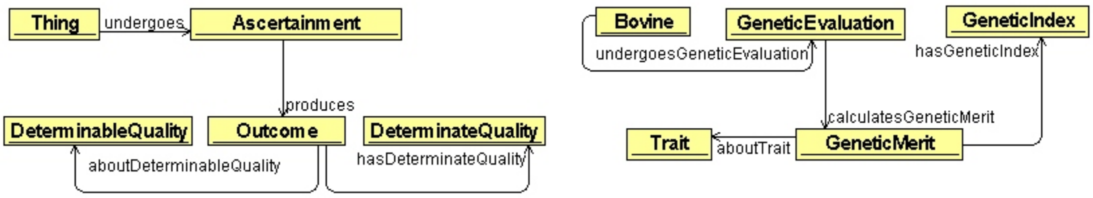

Figure 3: The OP *Ascertainment* pattern (left) and its instantiation around *GeneticEvaluation* (right).

an assessment procedure of some kind or *Ascertainment*, that *produces* an *Outcome about* some *DeterminableQuality* or other characteristic of the target and is quantified by some measure or *DeterminateQuality*. This OP abstracts different ways of acquiring certain knowledge concerning the target entity. Under the umbrella of this OP, one finds such dairy farming activities as milk composition tests, genetic evaluations and cow conformation scoring, to name a few. For instance, genetic evaluation is depicted on the right of Figure 3.

In searching for abstractions and OPs, we combine the bottom-up strategy of generalizing from concrete entities with the top-down strategy of making them specializations of a foundational ontology, the *Basic Formal Ontology* (BFO) from the OBO Foundry in our case. This greatly simplifies the integration of the two ontologies as the specialization approach gradually refactors the DCPO using BFO as a design guide, trying to align our entities to entities in the upper ontology. This has the effect of forcing our design to comply to the upper ontology, and thus absorb its principles. As an example, a genetic trait is any measurable characteristic of a cow that is heritable with some probability. Using the bottom-up strategy, we found a hierarchy of trait classes associated with concrete measures. From the top-down perspective we understand that traits are *BFO:Quality* specializations. So, we created the classes *DeterminableQuality* and *DeterminateQuality* for general use in our patterns, which are both specializations of *BFO:Quality* and generalizations of our concrete classes for traits and measures correspondingly (see Figure 3). Finally, as integration with BFO is an ongoing task, we will not extend on the subject.

This strategy enabled the rapid design of a coarse first model made of candidate classes and properties. Diagrams using a UML-based OWL representation were used to capture the current state of elicited knowledge and design decisions. We choose OWL since it is a standard Semantic Web technology built on top of RDF, a data format designed for interoperability, and provides valuable inference capabilities. This choice brought three main benefits: First, through the UML graphical visualization, it facilitated the communication among domain experts and ontologists. Second, the graphical editor OWLGrEd[8] enabled a smooth production of the OWL formalization. Third, the ontology consistency could be checked with a reasoner, thus greatly reducing the production effort for the formal ontology artifact.

Domain experts challenged this first model/design. From that point onward, the analysis/design process followed an iterative feedback loop (refining, updating and adding new entities) with domain experts regularly challenging the latest changes.

4.4. Ontology Description

From the start, this ontology has been designed with modularity as a key component. To this end, we have divided the ontology according to the different aspects of the dairy process covered at this stage: core, production and quality control, testing, breeding and genetic evaluation. In the following paragraphs we describe the ontology at the highest level of abstraction, as depicted in Figure 4. Notice the use of italics to highlight ontology entities where classes begin with an uppercase letter and properties with a lowercase one.

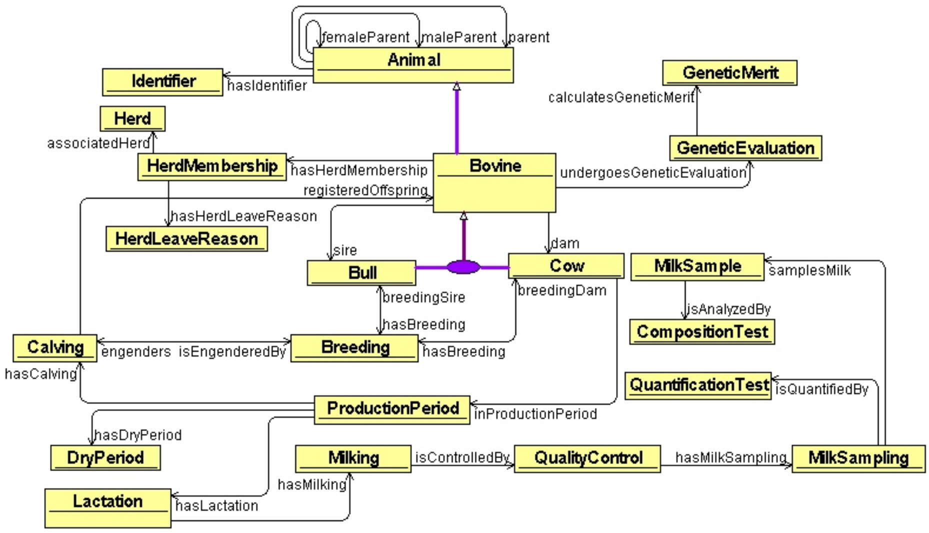

Figure 4: Dairy ontology and its modular design.

At the core of the ontology, the central abstract entity *Bovine*, factors out common characteristics of main actors: *Cow* and *Bull*, regardless of their particular role in the process or their life stage, allowing these concrete specializations to refine a common base by inheritance.

[8] http://owlgred.lumii.lv/

Bovine is derived from *Animal*, used to enable extensions of the DCPO to other dairy species. A *parent* property and its specializations *femaleParent* and *maleParent* are defined on *Animal* to allow the construction of a parentage graph tracking the pedigree of each animal, with further specializations *dam* and *sire* for the cows and bulls involved in breedings, respectively.

The productive life of a cow is represented by its associated individuals of the class *ProductionPeriod* which has three main stages: *Calving*, representing the birth of a new calf; *Lactation*, the milk production periods the cow has went through and *DryPeriod*, the time the cow is not producing milk. During *Lactation*, the *Milking* of cows undergoes *QualityControl* whose instances represent the different milk quality checkpoints performed during lactation. Quality control performs *MilkSampling* to produce a *MilkSample* that *isAnalyzedBy* a *CompositionTest*. During *MilkSampling* a *QuantificationTest* is performed to measure the milk yield . The class *LactationEndReason* represents the several causes a lactation terminates, the regular cause being the cow goes *Dry* or lactation may be interrupted because the cow *Died*. A *Breeding* between a *breedingDam* and a *breedingSire engenders* a new *Calving* producing a new *registeredOffspring Bovine*. Each *Bovine*, undergoes a *GeneticEvaluation* that calculates the *GeneticMerit* of the animal on several traits, to assess its value (a full description is available though CDN).

Finally, a *Herd* entity is associated with the concept of *HerdMembership*, representing the fact that a cow belongs to a herd. The class *HerdLeaveReason* associated to *HerdMembership*, represents the cause(s) for a cow to leave the herd. Two categories of reasons exist: *Voluntary* and *Involuntary*. Current statistics of our ontology are as follows:

Table 1: Dairy ontology main metrics

Construct	Number	Construct	Number
Classes	150	Subclass axioms	136
Object properties	67	Sub-object property axioms	48
Data properties	125	Sub-data property axioms	35

4.5. Ontology Usage and Evaluation

4.5.1. Query-based Evaluation

As a preliminary evaluation of the DCPO, we adopted a query-based approach. The motivation behind was two-fold: (1) assess the practical usability of the populated ontology and (2) ensure the correctness of applied data transformations. Led by domain experts, we implemented SPARQL queries that reflect the typical questions experts might ask, e.g. to estimate the impact of cow management w.r.t. to genetic potential.

```
PREFIX valacta: <http://valactadairy/basic#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>

select ?breed COUNT(?lact) as ?nbr_lact
AVG(xsd:double(?d305m)) as ?day_305_milk
AVG(xsd:double(?d305f)) as ?day_305_fat
```

```
AVG(xsd:double(?d305p)) as ?day_305_protein where {
?cow valacta:breed ?breed .
?cow valacta:hasLactation ?lact .
optional { ?lact valacta:day305Milk ?d305m . }
optional { ?lact valacta:day305Fat ?d305f . }
optional { ?lact valacta:day305Protein ?d305p . }}
GROUP BY ?breed ORDER BY ASC(?breed) LIMIT 100
```

For example, the above query computes the average values on Day 305 estimates for milk, protein and fat by cow breed. Simply put, the goal is to compute averages for cows, herds and regions for both production metrics (i.e. milk, fat and protein) and estimated genetic potential (i.e. estimated breeding values). By substracting – relative to average – values for production and genetics rough estimates of the quality of management practices for cows and herds are computed. While this query is rather straightforward, more complex ones have been developed.

4.5.2. Generalized Graph Pattern Mining with a DONT

Structural regularities, or patterns, in the data can provide useful insights as to the general trends it reflects: They may lead an expert to discover unknown phenomena or, more realistically, to confirm an already formulated hypothesis. Therefore, such regularities, are worth mining and presenting to experts for an in-depth examination.

The immediate benefit of using a DONT as vocabulary for pattern graphs is to enable the shared structure in data graphs to be explicitly described at the conceptual level, even though it may manifest in diverging ways at the data level. In other words, isomorphic graphs on the data level with diverging vertex and edge labels, which are thus, seemingly, unrelated, can become identical once their respective labels are generalized to the respective classes and generic properties from the DONT.

Here, our DCPO and its instances act as a dual graph model where the former is used as a blueprint while the latter acts as the actual data to explore and analyze. Another way to picture it is to consider it as meta-data to formulate relevant hypothesis whereby graph data is used to (in-)validate such hypothesis.

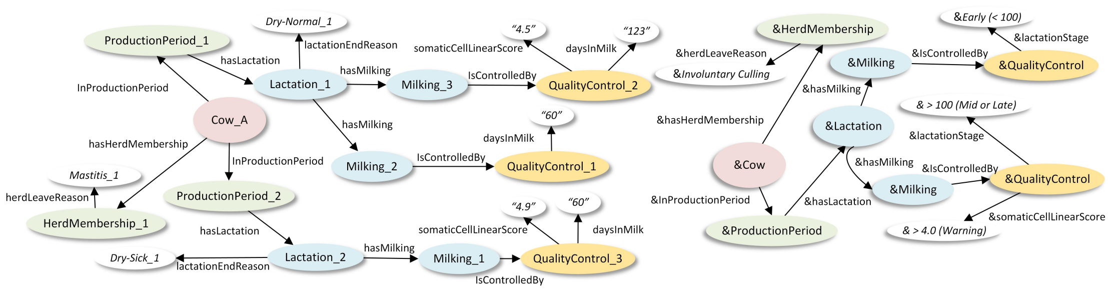

Figure 5: Example pattern (right) from a data graph (left), both supported by the dairy ontology.

As an illustrative example, Figure 5 represents a data graph and a matching pattern that refers to DCPO. The pattern –found in an *ad-hoc* manner– was deemed potentially useful by our experts. It reflects the fact that a number of cows culled for reasons that were not under farmer's control (*Involuntary Culling* class) had, prior to that event, at least one lactation with two quality controls, one of which indicates worrisome values of somatic cells. Such a co-occurrence is perfectly plausible as increased somatic cell counts are major signals for *mastitis* (inflammation of the udder tissue). Consequently, larger patterns contextualizing recurrent health issues could very well reveal the actual trigger for the involuntary culling. Therefore such patterns deserve to be investigated so that the underlying phenomena could be better understood and, if necessary, more closely monitored.

It is noteworthy that in order to support a finer-grained pattern language for the mining tools, we started enhancing the hierarchical structure in DCPO. To that end, we developed a dedicated version in which many data properties were transformed into object ones so that a hierarchy of OWL classes could express the generality between various groups of values in a property range. This proved particularly suitable for the variety of codes expressing the reasons for particular outcomes at the end of notable periods in the dairy cow life-cycle (lactation, herd membership, etc.). Such values are readily grouped into categories, e.g. the aforementioned *Involuntary* class is one such artefact. Further examples of such transformations include the somatic cell count (in its linear score version) for which threshold values exist: its usage illustrated by the pattern in Figure 5. Overall, one could envision extending this type of transformation to all data properties in DCPO. While doing it manually is hardly conceivable, automated methods based on clustering- or formal concept analysis [22] are conceivable.

5. Conclusion

We reflect here on our efforts on the design and implementation of DCPO, unifying several key aspects of dairy production. A major challenge we faced was the trade-off between plausible domain modeling and support for expressive knowledge discovery tools. At the current stage, it proved possible to reach both goals within a unique ontlogy.

Next, we shall look at how to exploit *ontology design patterns* [23] and conform to a foundational theory. In particular, given the biological nature of the data, we envision an integration to the OBO ontologies[9], with alignments to the relevant ontologies of the library. Eventually, the ontology will be publicly released to the community. In longer run, we shall look at enhancing the data-centered ontology with knowledge discovered from the data by mining tools.

References

[1] A. Goldstein, et al., A framework for evaluating agricultural ontologies, Sustainability 13 (2021) 6387.

[2] B. Drury, et al., A survey of semantic web technology for agriculture, Information Processing in Agriculture (2019). doi:`https://doi.org/10.1016/j.inpa.2019.02.001`.

[9] http://www.obofoundry.org/

[3] S. D. K. Tomic, et al., agriopenlink: Towards adaptive agricultural processes enabled by open interfaces, linked data and services, in: MTSR, 2013.
[4] C. Jonquet, et al., Agroportal: A vocabulary and ontology repository for agronomy, Computers and Electronics in Agriculture 144 (2018) 126–143.
[5] C.-J. Su, S.-F. Huang, Ontology-supported knowledge management with case-based reasoning for intelligent health projection, in: BIBE, 2018, pp. 1–4.
[6] J. Verhoosel, J. Spek, Semantics for big data applications in the smart dairy farming domain, in: Precision Dairy Farming Conference, Leeuwarden (NL), 2016.
[7] D. Tomic, et al., Experiences with creating a precision dairy farming ontology (dfo) and a knowledge graph for the data integration platform in agriopenlink, Agrárinformatika/Journal of Agricultural Informatics 6 (2015) 115–126.
[8] T. Martin, et al., Bridging the gap between an ontology and deep neural models by pattern mining, in: CSSA@CIKM, CEUR Workshop Proceedings, volume 2708, 2020.
[9] C. Frasco, et al., Towards an effective decision-making system based on cow profitability using deep learning, in: ICAART 2020, volume 2, 2020, pp. 949–958.
[10] D. Martini, et al., agrordf as a semantic overlay to agroxml: a general model for enhancing interoperability in agrifood data standards, in: CIGR Conference on Sustainable Agriculture through ICT Innovation, 2013.
[11] C. Caracciolo, et al., The agrovoc linked dataset, Semantic Web 4 (2013) 341–348.
[12] F. Kramer, T. Beißbarth, Working with ontologies, in: Bioinformatics, 2017, pp. 123–135.
[13] C. C. Aggarwal, H. Wang (Eds.), Managing and Mining Graph Data, volume 40 of *Advances in Database Systems*, Springer, 2010.
[14] C. C. Aggarwal, J. Han, Frequent Pattern Mining, 2014 ed., Springer, 2014.
[15] R. Srikant, R. Agrawal, Mining generalized association rules, Future Generation Computer Systems 13 (1997) 161–180. doi:http://dx.doi.org/10.1016/S0167-739X(97)00019-8.
[16] M. Adda, et al., Toward Recommendation Based on Ontology-Powered Web-Usage Mining, Internet Computing, IEEE 11 (2007) 45–52.
[17] T. Jiang, et al., Mining generalized associations of semantic relations from textual web content, IEEE TKDE 19 (2007) 164–179.
[18] A. Cakmak, G. Ozsoyoglu, Taxonomy-superimposed graph mining, in: 11th EDBT, 2008, pp. 217–228.
[19] U. Kursuncu, et al., Knowledge infused learning (k-il): Towards deep incorporation of knowledge in deep learning, in: AAAI-MAKE Symposium, 2020.
[20] H. Wang, et al., Ontology-based deep restricted boltzmann machine, in: DExA, 2016, pp. 431–445.
[21] N. Phan, et al., Ontology-based deep learning for human behavior prediction with explanations in health social networks, Information sciences 384 (2017) 298–313.
[22] M. Rouane-Hacene, et al., Relational concept analysis: mining concept lattices from multi-relational data, Annals of Mathematics and Artificial Intelligence (2013) 1–28.
[23] A. Gangemi, V. Presutti, Ontology design patterns, in: Handbook on ontologies, Springer, 2009, pp. 221–243.