Available at www.sciencedirect.com

INFORMATION PROCESSING IN AGRICULTURE 9 (2022) 378–396

journal homepage: www.elsevier.com/locate/inpa

FeedEfficiencyService: An architecture for the comparison of data from multiple studies related to dairy cattle feed efficiency indices

Heitor Magaldi Linhares [a], *Regina Braga* [a,*], *Wagner Antônio Arbex* [a,b], *Mariana Magalhães Campos* [b], *Fernanda Campos* [a], *José Maria N. David* [a], *Victor Stroele* [a]

[a] *Programa de Pós-Graduação em Ciência da Computação, Universidade Federal de Juiz de Fora, Campus Universitário, Rua José Lourenço Kelmer, s/n - São Pedro, Juiz de Fora, Minas Gerais 36036-900, Brasil*

[b] *Brazilian Agricultural Research Corporation - Embrapa, Av. Eugênio do Nascimento, 610 - Aeroporto, Juiz de Fora - MG, Minas Gerais 36038-330, Brasil*

A R T I C L E   I N F O

Article history:
Received 24 June 2020
Received in revised form
6 July 2021
Accepted 21 July 2021
Available online 27 July 2021

Keyword:
Dairy Cattle

A B S T R A C T

The increased demand for food worldwide, the reduced land availability for livestock production, the increasing cost of animal feed and the need for mitigating livestock-related greenhouse gas emissions have driven the search for animal feeding systems that proves more efficient. To tackle this problem, we propose the use of computational support to help researchers compare data on feed efficiency, therefore improving economic and environmental gains. As a solution, we present an integrative architecture capable of combining heterogeneous data from multiple experiments related to dairy cattle feed efficiency indices. The proposed architecture, called *FeedEfficiencyService*, classifies animals according to feed efficiency indices and allows visualizations through ontologies and inference engines. The results obtained from a case study with researchers from the Brazilian Agricultural Research Corporation – Dairy Cattle (EMBRAPA) demonstrate that this architecture is a supporting tool in their daily work routine. The researchers highlighted the importance of the proposed architecture as it allows analyzing animal data, comparing experiments, having reliable data analyses, and standardizing and organizing data from experiments. The novelty of our approach is the use of ontologies and inference engines to enable the discovery of new knowledge and new relationships between data from feed efficiency-related experiments. We store such data, relationships, and analyses of results in an integrated repository. This solution ensures unified access to the processing history and data from diverse experiments, including those conducted at external research centers.

1. Introduction

Regarding the world's population growth, Paddock [1] points out that the way dairy farming is carried out should be more efficient to ensure a sustainable production [2] and supply of

* Corresponding author at: Programa de Pós-Graduação em Ciência da Computação, Universidade Federal de Juiz de Fora, Campus Universitário, Rua José Lourenço Kelmer, s/n - São Pedro, Juiz de Fora, Minas Gerais 36036-900, Brasil.
E-mail address: regina.braga@ufjf.edu.br (R. Braga).
Peer review under responsibility of China Agricultural University.
https://doi.org/10.1016/j.inpa.2021.07.002
2214-3173

milk and dairy products. The demand for animal source foods (protein), the reduced land availability for livestock production, and the increasing cost of animal feed have led to a greater search for animal feeding systems that are more efficient. Ideally, such systems should use smaller areas and fewer natural resources to achieve production yields similar or superior to those of systems currently in use. However, on the one hand, consumers normally focus on food safety and on the nutritional quality of animal source foods. On the other hand, producers tend to prioritize their profitability by increasing efficiency in the production systems [3].

Concerning nutrient intake in beef cattle, Arthur et al. [4] report that there is a variation between animals with similar characteristics, due to underlying factors such as feed consumption, physiological mechanisms, animal live weight and weight gain. Montanholi et al. [5] stress that predicting the most efficient animals is possible by understanding the factors that regulate feed efficiency. However, selecting and understanding these factors is still a challenge.

Such a challenge has motivated the Feed Efficiency Group of the Brazilian Agricultural Research Corporation – Dairy Cattle (EMBRAPA) to conduct research on animal nutrition and to characterize efficient animals through blood, morphological, reproductive, and metabolic data, with the aim of simplifying and reducing costs for the classification and selection of these animals.

The EMBRAPA's Feed Efficiency Group carried out various experiments with animals. The experiments covered the multiple stages of an animal's life, from suckling to lactating. Through these experiments, they constructed hypotheses to relate feed efficiency with the following factors: (i) behavior in feed and water consumption; (ii) morphological measurements; (iii) blood markers; (iv) hormone markers; and (v) temperature. Through classification, the Group isolates efficient animals and analyzes their data in order to understand the factors that regulate feed efficiency. As no most suitable feed efficiency index is known, some experiments consider more than one index to classify the animals [6–9]. In this regard, it is important to compare these feed efficiency indices in the literature since there is no consensus about a most appropriate index for dairy cattle. Researchers lack instruments that allow them making such comparisons and analyzing and examining the distribution among the efficiency classes (labeled as efficient, intermediary, and inefficient) [6].

Among these indices are: Consumption and Residual Weight Gain (*GPR*[1]), Gross Feed Efficiency (*EA*), Feed Conversion Efficiency (*ECA*), Residual Feed Consumption (*CAR*), Residual Weight Gain (*GPR*), Relative Growth Rate (*TRC*), and Keiber Ratio (*RK*) [6,7,10–12]. In all of them, three variables are considered: Daily Weight Gain (*GPG*), i.e., the amount of weight that the animal gained (in kg/day) obtained from an average value; Dry matter intake (*IMS*), i.e., the amount of dry matter that the animal ingested (in kg/day) on average (the values determined in the field comprise the consumption of natural matter; subsequently, the daily feed used in the experiment is analyzed, thus calculating the proportion of dry matter present in the natural material used); Average Metabolic Weight (*MWh*), calculated as the *Average Live Weight* to the power of 0.75, as shown in Eq. (1). The *Average Live Weight* is defined as the average weight of the animal throughout the experiment.

$$MWh = (AverageLiveWeight)^{0.75} \quad (1)$$

The research on the best index for dairy cattle is conducted through comparative studies. Researchers at EMBRAPA and at other related research centers face problems in storing, classifying, and analyzing experiments. They store the experiments on heterogeneous and non-integrated repositories, and each researcher is responsible for producing and maintaining their own data. The availability of the data produced and the lack of standardization of the experiments is recurrent. Moreover, animal classification is laborious and time-consuming because it is done manually, with limited computational support. The data usually need to be transcribed in tables, tools are described, but mistakes are common.

In view of the above, adequate computational support is paramount for the classification of the experiments, as well as for the comparison and analysis of the animals' progress in the experiments. Furthermore, the data need to be stored, processed, cross-referenced, and accessed by researchers in an integrated manner so that the results can be compared using statistical support and analysis tools.

To tackle these problems, this paper details the specification of an architecture named *FeedEfficiencyService*. It is based on ontologies [16], inference engines, and data analysis, allowing for the discovery of new knowledge and new relationships between data from feed efficiency experiments, thus supporting researchers in comparing such data.

In this vein, an ontology[2] called Feed Efficiency Ontology (henceforth FEO) was specified to discover new information related to dairy cattle feed efficiency. Semantic Web Rule Language (SWRL) logical rules [13] were also specified. These rules are processed by inference engines (specifically the Pellet reasoner) to classify animal data and to, based on this classification, discover new information related to such data.

The proposed architecture aims to guarantee that experiments be adequately compared, once the data, relationships, and analysis results are stored in an integrated repository, ensuring access to the processing history and data from diverse experiments, including experiments conducted at other research centers. *FeedEfficiencyService* can also be used in other scientific experimentation contexts. Experiments that need analyses on data and cross-referencing of heteroge-

[1] Henceforth, the acronyms for the indices will appear in Portuguese, in the way that local researchers routinely use them, as these acronyms will be referred to as such in the proposed architecture.

[2] Bernes-Lee et al. [18] define ontology as a file or document that formally explains the relationships between terms and, through the use of inference rules, like SWRL [13] or Shapes Constraint Language [20] rules, provides additional power since terms not related can be associated. The standard language for defining ontologies is Ontology Web Language (OWL), proposed by the World Wide Web Consortium (W3C). This language is based on computational logic so that the knowledge expressed in OWL can be shared between systems [19].

neous information can also benefit from *FeedEfficiencyService* – the ontology must be related to the domain of the experiment and can be accessed by web portals, such as those presented in Drury et al. [14].

Moreover, researchers will be able to use scientific management systems, such as Taverna[3] or Kepler[4], to process their collaborative experiments in an integrated manner, even if they are in different research centers.

Specific contributions of the *FeedEfficiencyService* architecture include i) integration of heterogeneous data from multiple experiments; ii) definition of a generic architectural model for the discovery of new knowledge; iii) creation of visualizations to support feed efficiency analyses; iv) definition of an API (Application Programming Interface) for the services offered by the architecture, based on the principles of Software as a Service (SaaS); v) development of a Web application to allow researchers to access the services.

The Design Science Research (DSR) methodology [15] was used in our study. Through DSR, knowledge and understanding of a problem domain and its solution were achieved in constructing and using the designed artifact (*FeedEfficiencyService*). The artifact evaluation then provides feedback information and a better understanding of the problem to improve the quality of both the artifact and the design process.

The following research question was formulated from the scope definition: "*How does the FeedEfficiencyService architecture make it easy for researchers to analyze data from Feed Efficiency experiments?*".

This paper is organized into four sections, including this introduction. Based on the DSR methodology, section 2 details how the study was conducted, including a systematic literature review, as well as a description and evaluation of the *FeedEfficiencyService* architecture. Section 3 presents the results and discussion. Finally, section 4 presents final considerations.

2. Materials and methods

Following the DSR methodology, we identified the problem of relevance as: "Researchers need mechanisms that are more efficient in allowing for the storage, integration, and analysis of experiments related to dairy cattle nutrition". For this purpose, we investigated the domain area in order to conduct a Search Process. A systematic literature mapping was performed to identify the problem domain and the existing solutions in the field (Section 2.1). Because no proposals met the group's requirements for feed efficiency research, the *FeedEfficiencyService* (Section 2.2) was proposed. To collect evidence on the feasibility of our proposed solution, we conducted an evaluation by gathering data from scientific projects (Section 2.3) related to feed efficiency research. For that reason, the contributions of our study can be summarized as an architecture proposed to support and integrate feed efficiency data from experiments conducted at EMBRAPA. From ontologies and data analysis, new knowledge and new relationships can be discovered, thus strengthening data integration.

[3] https://taverna.incubator.apache.org/

[4] https://kepler-project.org/

2.1. Systematic literature mapping

As part of the DSR methodology, we conducted a Systematic Literature Review (SLR) considering ontologies applied to the agricultural context.

The SLR followed the protocol proposed by Kitchenhan and Charters [21] to describe the steps necessary for its conducting for later reproducibility of the results obtained.

The purpose of the study was defined as: *with ontologies as **the object of study**, the **intention/purpose** is to identify techniques, models, prototypes, architectures, frameworks, and tools that have the **effect** of supporting agricultural research through the use of ontologies, **from the point of view** of researchers in the agricultural context.*

We used two control articles, [22] and [23], to support the search string construction. The search string was as follows: *("rural industries" or "dairy farming" or "rural industry" or "livestock" or "cattle raising" or "dairy industry") and ("ontology") and ("model" or "architecture" or "framework" or "techniques" or "prototype")*, and the following databases were searched: IEEE, ACM, Engineering Village, SCOPUS, ScienceDirect.

A total of 82 articles were retrieved using the search string. All the databases returned results; however, Scopus was the database with the highest number of papers, 54 in total, followed by Engineering Village and ACM Digital Library with 9 papers, ScienceDirect with 7 papers, and IEEE Digital Library with 3 papers. These results are shown in Fig. 1.

None of the proposals in the retrieved papers uses ontology by combining inference engines with the integration and sharing of heterogeneous information or providing agricultural research support. However, some of these studies deserve attention, as discussed below.

Parrott et al. [24] developed a multi-agent collaborative architecture to support decisions in the dairy industry. In the approach, the authors used ontology to map the meanings of different domains to establish communication among agents, i.e., they use ontologies as a communication language. However, they did not use ontology to process new knowledge based on inference engines. Thus, this multi-agent collaborative architecture does not produce new knowledge through the ontology and does not propose an integration of experiments, unlike the *FeedEfficiencyService* architecture, which uses direct relationships to integrate the

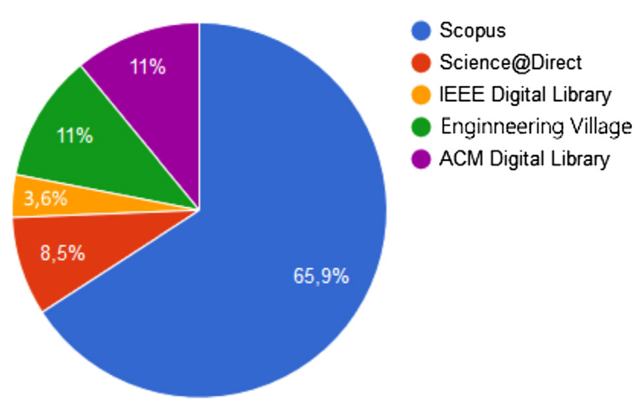

Fig. 1 – Percentage of articles retrieved from each database.

data and inference algorithms (Pellet) to discover new knowledge.

An ontology for the nutritional management of herds was proposed by Sivamani et al. [25]. The authors considered aspects related to the variation in nutritional requirements. The life stages of the animal and the cycle of milk production were considered for creating the ontology. The authors discovered that confining heifers, calves, and mature cattle together is common on rural properties. Through logical rules and the animals' unique identification, the ontology can classify them according to the stage of life and select the ideal diet. The authors defined an ontology with a focus on animal feeding, considering the nutritional variations required in different stages of their lives, unlike the ontology proposed through the *FeedEfficiencyService* architecture. This architecture considers aspects of feed efficiency and four main classification indices, in addition to assisting in the integration of heterogeneous data from different experiments.

Janssen et al. [26] described the System for Environmental and Agricultural Modeling (SEAMLESS) architecture, which integrates databases from different domains, such as climatic conditions, soil, and cropping patterns. The authors developed a collaborative ontology to facilitate the study's interdisciplinarity nature, focusing on mapping those databases. The *FeedEfficiencyService* architecture seeks to discover new knowledge and share it with EMBRAPA's research centers and external partners.

Hulsegge et al. [27] discussed the development of the Animal Trait Ontology (ATO), an ontology for the livestock sector. They identified that there are still few ontologies for livestock production, quality and health aspects. Accordingly, the authors developed two ontologies for livestock production: Reproductive Trait and Phenotype Ontology (REPO) and Host-Pathogen Interactions Ontology (HPIO). The REPO, HPIO, and the FEO (proposed in the present paper) have different applications in the same context. The REPO focuses on female fertility in dairy cattle; while the HPIO focuses on the interactions between pigs and salmonella; and the FEO focuses on feed efficiency indices.

Jonqueta [17] presents a platform called AgroPortal that receives and hosts ontologies, aligns them, and enables their reuse in agriculture. The initiative can store and share the FEO and reuse ontologies related to other agricultural contexts through the *FeedEfficiencyService* architecture. Therefore, *FeedEfficiencyService* can be integrated with the AgroPortal platform. In Drury et al. [14], the authors provide a survey that discusses the use of semantic web technologies to address agricultural problems. The survey presents several initiatives that use ontologies to solve agricultural problems. However, none presents an ontology related to feed efficiency capable of processing inferences and deriving new information, coupled with a service that analyses the data considering ontological rules. On the other hand, the FEO can be stored into the portals described in the survey, such as the AGROVOC or Agroportal [17], and *FeedEfficiencyService* can also use ontologies provided by these portals. Therefore, *FeedEfficiencySercive* can be used in different domains.

The results of the systematic review show that none of the papers met all the feed efficiency requirements. Therefore, this pointed to the need for an architecture able to provide experiment data integration and analysis through the support of a visualization mechanism, i.e., **the creation of an innovative purposeful artifact according to the DSR methodology**.

2.2. *Feed efficiency service*

The researchers at EMBRAPA working in the context of feed efficiency need computational tools to assist them with data storage, and classification and analysis of animals in their experiments. These researchers need to analyze the data produced throughout the experiments and compare them between animals or other experiments daily. Tools such as spreadsheets, usually used by them, do not provide an interface that facilitates such analyses.

The *FeedEfficiencyService* architecture was proposed to deal with these issues, supporting researchers in scientific experimentation related to feed efficiency, based on data integration from multiple experiments and integrated analysis tools. Besides, *FeedEfficiencyService* has a secondary objective, which is to guarantee the reproducibility of the experiments, enabling access to and reuse of the experiments and the results thereof.

The *FeedEfficiencyService* architecture was specified following the layered architectural model [28] to provide better modularization and decoupling between the services, facilitating integration with other applications. Fig. 2 presents a high-level vision of the architecture, composed of the following layers: Service Layer, Front End Layer, Data Layer, and Ontology Layer.

For the use of data from multiple experiments with different data models, an integrator data model was specified based on information from the heterogeneous databases and on the researchers' information. Fig. 3 displays an overview of the model. The model presents metadata that directly or indirectly affects the feed efficiency experiments. For example, the metadata related to dairy food directly impacts the various related experiments, considering that we measured it in kilograms of natural matter. The variable dry matter intake (*IMS*) is obtained through its processing, and it calculates feed efficiency indices.

Another example is the temperature and coverage metadata, used to correlate these metadata with the efficient animals to identify some characteristics investigated. Thus, the model has information related to animals, experiments,

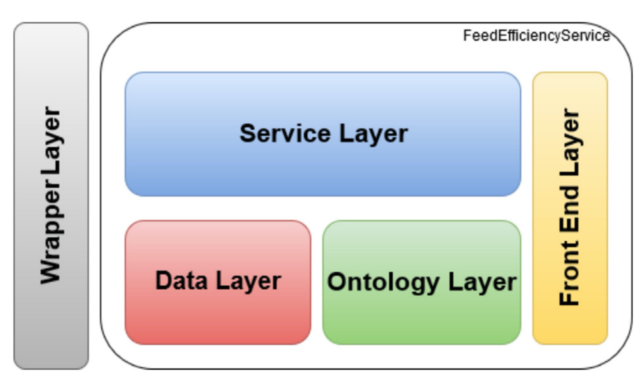

Fig. 2 – *FeedEfficiencyService* high level vision of the

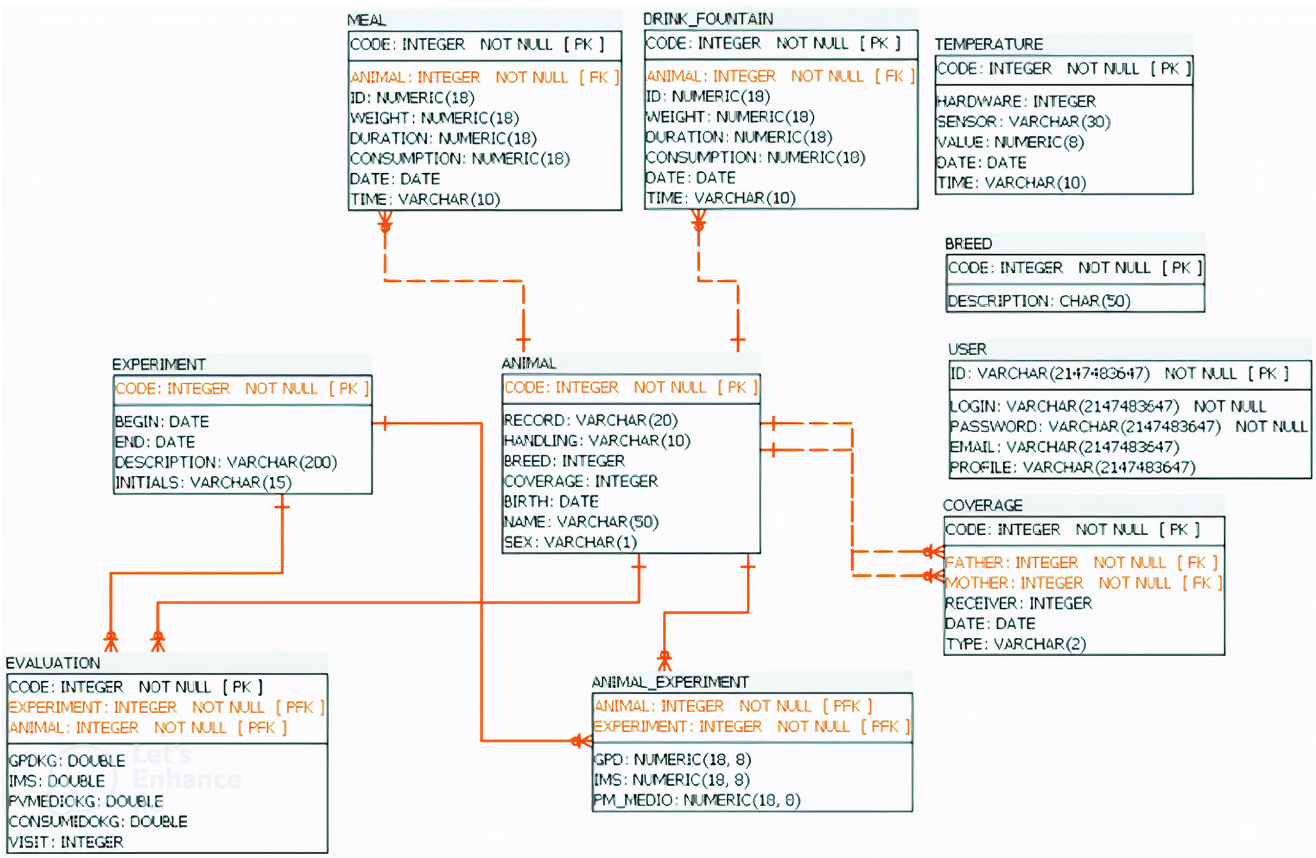

Fig. 3 – Data model of the *FeedEfficiencyService* (Data Layer) architecture.

water consumption, feed consumption, coverage, breed, and climatic information, to name a few. We proposed an ontology, as previously mentioned, named FEO to discover new connections between metadata and to analyze these metadata considering feed efficiency indices. The FEO provides specific classes to analyze feed efficiency indices and to discover new connections between animal data. We use these specific classes to process feed efficiency indices and provide new connections between data, for example, by identifying similarities between animals across the experiments.

2.2.1. *The feed efficiency ontology (FEO)*

The Ontology Layer aims to support the integration and analysis of the data from experiments. This layer encompasses the FEO[5] (see Fig. 4). The FEO allows semantic integration between related experiments.

It supports researchers in animal classification and data interoperability to perform cross-analysis and discover new connections between experiments. Besides, due to the need to classify the animals according to the *CAR*, *GPR*, *CGPR*, and *ECA* indices, specific classes and rules were created for efficiency classification, considering three possible levels: efficient, intermediary and inefficient.

The language used to implement the ontology was Ontology Web Language (OWL) 2.0, recommended by World Wide Web Consortium (W3C). The FEO structure is composed of three main classes: Cattle, Classification, and Evaluation. The Classification class has three subclasses in the feed efficiency index: efficient, intermediary and inefficient. There are other four subclasses for feed efficiency indices: *CAR*, *GPR*, *CGPR*, and *ECA*. We specified the ontology to make it easily extensible, i.e., if new feed efficiency indices are needed, it is easy to create new subclasses in the hierarchy and classification and evaluation types. Considering the ontology implementation in OWL, we used object properties to implement relationships between classes.

To discover new associations between experiments and animals and enable the processing of classifications related to feed efficiency, we created SWRL rules in the FEO. These rules classify animals as efficient, intermediary or inefficient for each of the feed efficiency indices and enable discovering new associations between animals and experiments.

The rules were constructed based on the information provided by the EMBRAPA's researchers. The classification of animals into efficient, intermediary and inefficient draws on the standard deviation of the indices obtained. Thus, for the *CAR*, *GPR* and *CGPR* indices, the range of animals with indices between a standard deviation below and above zero is considered intermediary (see Fig. 5). In the *ECA* index, we consider the average of the indices rather than the zero marks. Consid-

[5] https://github.com/heitormagaldi/FeedEfficiencyServiceBase/tree/master/data

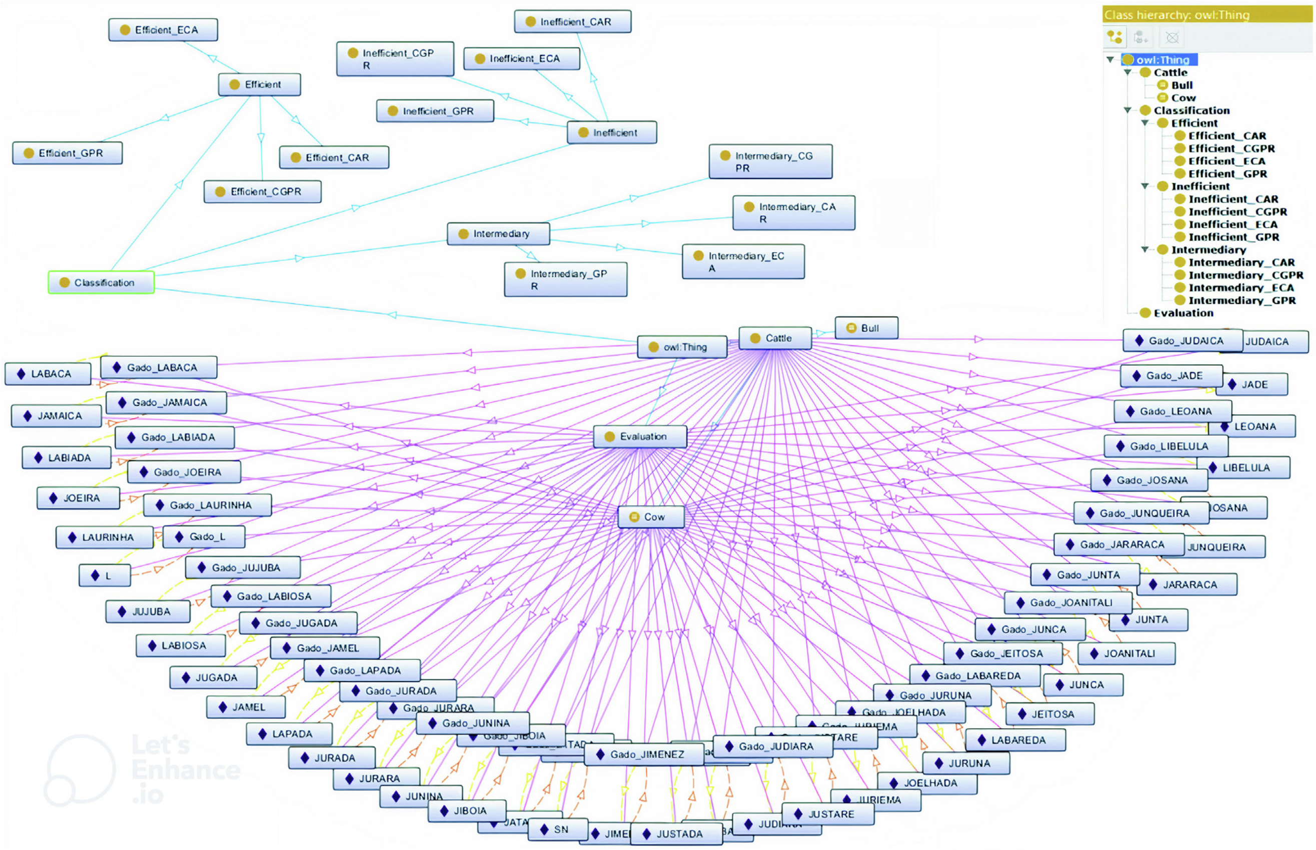

Fig. 4 – Feed Efficiency Ontology (FEO) Defined (or Declared) Model.

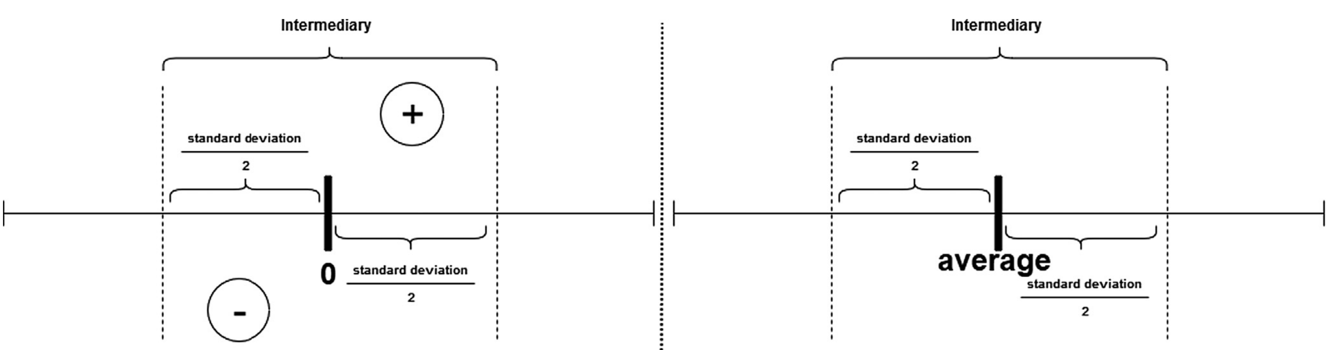

Fig. 5 – Range for the classification of animals as intermediary as per the CAR, GPR, ECA and CGPR indices.

ering that the efficient, intermediary and inefficient classes are disjointed in the FEO, the classification in the efficient and inefficient bands considers only the indices not classified as intermediary.

In this sense, the *ECA*, *GPR* and *CGPR* indices consider the animals in the highest value range as the most efficient. For the CAR index, the opposite is considered, i.e., the most efficient animals are in the lowest value range.

The classification ranges are related to the standard deviation, which is related to the set of animals. This approach brings dynamism to this calculation. It is also important to highlight that the addition or removal of an animal in this set can adjust the standard deviation values and averages and affect all the bands. In this case, static labels can incur an error or unnecessary reprocessing. Thus, the adoption of SWRL rules provides flexibility to the architecture, leaving the inference engine (Pellet reasoner[6]) in charge of the respective classifications. Table 1 presents some of the SWRL rules.

The construction of the logical rules for the classification of the animals' instances uses data from one experiment, with animals previously classified by the animal nutrition/feed efficiency team of EMBRAPA. In order to evaluate the accuracy of the classifications, we used data from other five experiments.

6 https://github.com/stardog-union/pellet

Table 1 – SWRL rules created for classifications in the *Feed Efficiency Ontology*.

Name	Type	Classification	SWRL Rule
S1	CAR	Efficient	EMBRAPA:Cattle(?cattle) ^ EMBRAPA:isEvaluationOf(?cattle, ?evaluation) ^ EMBRAPA:Experiment_CAR(?evaluation, ?EvaluationCAR) ^ swrlb:lessThan(?EvaluationCAR, (-1)*X) ->EMBRAPA:Efficient_CAR(?cattle)
S2	CAR	Intermediary	EMBRAPA:Cattle(?c) ^ EMBRAPA:isEvaluationOf(?c, ?y) ^ EMBRAPA:Experiment_CAR(?y, ?EvaluationCAR) ^ swrlb:lessThanOrEqual(?EvaluationCAR, X) ^ swrlb:greaterThanOrEqual(?EvaluationCAR, (-1)*X) ->EMBRAPA:Intermediary_CAR(?c)
S3	CAR	Inefficient	EMBRAPA:Cattle(?c) ^ EMBRAPA:isEvaluationOf(?c, ?y) ^ EMBRAPA:Experiment_CAR(?y, ?EvaluationCAR) ^ swrlb:greaterThan(?EvaluationCAR, X) ->EMBRAPA:Inefficient_CAR(?c)
S4	GPR	Efficient	EMBRAPA:Cattle(?c) ^ EMBRAPA:isEvaluationOf(?c, ?y) ^ EMBRAPA:Experiment_GPR(?y, ?EvaluationGPR) ^ swrlb:greaterThan(?EvaluationGPR, X) ->EMBRAPA:Efficient_GPR(?c)
S5	GPR	Intermediary	EMBRAPA:Cattle(?c) ^ EMBRAPA:isEvaluationOf(?c, ?y) ^ EMBRAPA:Experiment_GPR(?y, ?EvaluationGPR) ^ swrlb:lessThanOrEqual(?EvaluationGPR, X) ^ swrlb:greaterThanOrEqual(?EvaluationGPR, (-1)*X) ->EMBRAPA:Intermediary_GPR(?c)
S6	GPR	Inefficient	EMBRAPA:Cattle(?cattle) ^ EMBRAPA:isEvaluationOf(?cattle, ?evaluation) ^ EMBRAPA:Experiment_GPR(?evaluation, ?EvaluationGPR) ^ swrlb:lessThan(?EvaluationGPR, (-1)*X) ->EMBRAPA:Inefficient_GPR(?cattle)
S7	ECA	Efficient	EMBRAPA:Cattle(?c) ^ EMBRAPA:isEvaluationOf(?c, ?y) ^ EMBRAPA:Experiment_ECA(?y, ?EvaluationECA) ^ swrlb:greaterThan(?EvaluationECA, X) ->EMBRAPA:Efficient_ECA(?c)
S8	ECA	Intermediary	EMBRAPA:Cattle(?c) ^ EMBRAPA:isEvaluationOf(?c, ?y) ^ EMBRAPA:Experiment_ECA(?y, ?EvaluationECA) ^ swrlb:lessThanOrEqual(?EvaluationECA, X) ^ swrlb:greaterThanOrEqual(?EvaluationECA, Z) ->EMBRAPA:Intermediary_ECA(?c)
S9	ECA	Inefficient	EMBRAPA:Cattle(?cattle) ^ EMBRAPA:isEvaluationOf(?cattle, ?evaluation) ^ EMBRAPA:Experiment_ECA(?evaluation, ?EvaluationECA) ^ swrlb:lessThan(?EvaluationECA, X) ->EMBRAPA:Inefficient_ECA(?cattle)
S10	CGPR	Efficient	EMBRAPA:Cattle(?c) ^ EMBRAPA:isEvaluationOf(?c, ?y) ^ EMBRAPA:Experiment_CGPR(?y, ?EvaluationCGPR) ^ swrlb:greaterThan(?EvaluationCGPR, X) ->EMBRAPA:Efficient_CGPR(?c)
S11	CGPR	Intermediary	EMBRAPA:Cattle(?c) ^ EMBRAPA:isEvaluationOf(?c, ?y) ^ EMBRAPA:Experiment_CGPR(?y, ?EvaluationCGPR) ^ swrlb:lessThanOrEqual(?EvaluationCGPR, X) ^ swrlb:greaterThanOrEqual(?EvaluationCGPR,(-1)*X) ->EMBRAPA:Intermediary_CGPR(?c)
S12	CGPR	Inefficient	EMBRAPA:Cattle(?cattle) ^ EMBRAPA:isEvaluationOf(?cattle, ?evaluation) ^ EMBRAPA:Experiment_CGPR(?evaluation, ?EvaluationCGPR) ^ swrlb:lessThan(?EvaluationCGPR, (-1)*X) ->EMBRAPA:Inefficient_CGPR(?cattle)

SWRL rules (Table 1) enable associations. For example, rule S1 classifies efficient animals in the CAR index. To do so, it uses previously known information (explicit knowledge), e.g. being an instance of *Cattle* and having an instance of the *Evaluation* class associated with it, besides having a data property *Experiment_CAR,* and that being less than X[7]. Thus, an animal instance with such combinations is classified in the *Efficient_CAR* class (implicit knowledge) as efficient CAR.

As a result, through its declared model (explicit knowledge) with the addition of specific SWRL rules and inference engine, the FEO infers the classification of the animals' instances under the feed efficiency indices (implicit knowledge). The animal classification is considered new knowledge, produced from the processing of SWRL rules and inference engines over the ontological instances.

2.2.2. *Implementation*

For applications that need to consume data and provide information to the architecture, such as the works presented in Drury et al. [14], a RESTful web service in JAVA was implemented. It provides services for the storage, management and query of data, and interoperability with other applications and services.

For the development of the *FeedEfficiencyService* architecture and the services layer, the Software as a Service (SaaS) paradigm [28] was adopted. The architecture enables the researcher to use data and services related to dairy cattle feed efficiency and share information with other researchers while also providing remote access from multiple devices.

In this way, the researcher does not need to worry about infrastructure, implementation of routines, external tools, and storage, among other technical details. Besides, the composition of services can be modified and be used by other external or internal researchers. We implemented 56 services that communicate through JavaScript Object Notation (JSON) files. Table 2 presents the services related to feed efficiency analyses. These services do not require prior knowledge of statistical tools nor the use of third-party tools.

Services related to the ontology were also made available and used to discover new connections between the data and to process the animals' classification in the feed efficiency indices. For this purpose, the Service Layer has a service that supports the processing of OWL files. This service considers the information present in the integrator model, available in the Data Layer, as well as the data from the analysis services and the parameters passed in the experiment's selected service request. The animals' classifications as efficient, intermediary and inefficient in the four feed efficiency indices are found through inference algorithm processing. Then, through SPARQL queries (see Fig. 6), this service returns the animals associated with each of the classes and, consequently, the classification into the feed efficiency indices (see Fig. 7 – the green highlights indicate the classification type (in Portuguese) and the animals returned).

[7] X is the value obtained by computing the standard deviation of the indices obtained, varying according to the classification of efficient, intermediary or inefficient (as presented in Fig. 5 for animals classified as intermediary).

Table 2 – Feed Efficiency data analysis services.

Operation	Method	URI	Output	Functionality
GET	Classification	http://www.servidor.com.br/EMBRAPA.site2/services/regressao/CAR/{codigoexperimento}	JSON	Regression of the CAR Index
GET	Classification	http://www.servidor.com.br/EMBRAPA.site2/services/regressao/ECA/{codigoexperimento}	JSON	ECA index formula
GET	Classification	http://www.servidor.com.br/EMBRAPA.site2/services/regressao/GPR/{codigoexperimento}	JSON	Regression of the GPR index
GET	Classification	http://www.servidor.com.br/EMBRAPA.site2/services/regressao/CGPR/{codigoexperimento}	JSON	Regression of the CAR and GPR index + calculation of the CGPR index

```
PREFIX owl:<http://www.w3.org/2002/07/owl#>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX embrapa: <http://www.semanticweb.org/heitor/ontologies/2016/6/Embrapa#>
SELECT distinct ?Nome ?Numero ?Registro ?Nascimento ?Codigo
WHERE {
?Cattle a embrapa:Efficient_CAR.
?Cattle embrapa:Cattle_Name ?Nome.
?Cattle embrapa:Cattle_Number ?Numero.
?Cattle embrapa:Cattle_Birth ?Nascimento.
?Cattle embrapa:Cattle_RGD ?Registro.
?Cattle embrapa:Cattle_Code ?Codigo.
}
```

Fig. 6 – Ontology's SPARQL query of animals associated with the Efficient_CAR class.

{"CAR":{"ineficientes":[{"cobertura":0,"codigo":22,"nascimento":"0012-02-20","raca":0,"manejo":"4581","nome":"JATADA JUNEAU DO CECP","sexo":"","registro":"0669-AL"},{"cobertura":0,"codigo":20,"nascimento":"0012-02-20","raca":0,"manejo":"4575","nome":"JOEIRA JUNEAU DO CECP","sexo":"","registro":"0668-AL"},{"cobertura":0,"codigo":28,"nascimento":"0012-02-20","raca":0,"manejo":"4587","nome":"JUNQUEIRA JUNEAU DO CECP","sexo":"","registro":"0646-AL"},{"cobertura":0,"codigo":35,"nascimento":"0001-03-20","raca":0,"manejo":"5408","nome":"LAPADA IPINION DO CECP","sexo":"","registro":"0644-AL"},{"cobertura":0,"codigo":15,"nascimento":"0011-02-20","raca":0,"manejo":"4554","nome":"JURADA JUNEAU DO CECP","sexo":"","registro":"0658-AL"},{"cobertura":0,"codigo":30,"nascimento":"0001-03-20","raca":0,"manejo":"5401","nome":"LABACA OPINION DO CECP","sexo":"","registro":""},
{"cobertura":0,"codigo":16,"nascimento":"0011-02-20","raca":0,"manejo":"4557","nome":"JUNCA JUNEAU DO CECP","sexo":"","registro":"0674-AL"},{"cobertura":0,"codigo":34,"nascimento":"0001-03-20","raca":0,"manejo":"5407","nome":"LABIOSA OPINION DO CECP","sexo":"","registro":"0649-AL"},{"cobertura":0,"codigo":18,"nascimento":"0012-02-20","raca":0,"manejo":"4572","nome":"JOELHADA JUNEAU DO CECP","sexo":"","registro":"0671-AL"},{"cobertura":0,"codigo":29,"nascimento":"0012-02-20","raca":0,"manejo":"4588","nome":"JIMENEZ OPINION DO CECP","sexo":"","registro":"0667-AL"},{"cobertura":0,"codigo":32,"nascimento":"0001-03-20","raca":0,"manejo":"5403","nome":"L BIA OPINION DO CECP","sexo":"","registro":"0650-AL"}],"intermediarios":
[{"cobertura":0,"codigo":17,"nascimento":"0011-02-20","raca":0,"manejo":"4560","nome":"JUNINA JUNEAU DO CECP","sexo":"","registro":"0657-AL"},{"cobertura":0,"codigo":13,"nascimento":"0011-02-20","raca":0,"manejo":"4550","nome":"JUDAICA JUNEAU DO CEC","sexo":"","registro":"0660-AL"},{"cobertura":0,"codigo":25,"nascimento":"0012-02-20","raca":0,"manejo":"4584","nome":"JURIEMA OPINION DO CECP","sexo":"","registro":"0666-AL"},{"cobertura":0,"codigo":9,"nascimento":"0011-02-20","raca":0,"manejo":"4545","nome":"JURUNA JUNEAU DO CECP","sexo":"","registro":"0661-AL"},{"cobertura":0,"codigo":19,"nascimento":"0012-02-20","raca":0,"manejo":"4573","nome":"JIBOIA TALENTO DO CECP","sexo":"","registro":"0654-AL"},{"cobertura":0,"codigo":36,"nascimento":"0001-03-20","raca":0,"manejo":"5409","nome":"LAURINHA OPINION DO CECP","sexo":"","registro":"0648-AL"},{"cobertura":0,"codigo":12,"nascimento":"0011-02-20","raca":0,"manejo":"4549","nome":"JOSANA JUNEAU DO CECP","sexo":"","registro":"0659-AL"},{"cobertura":0,"codigo":14,"nascimento":"0011-02-20","raca":0,"manejo":"4551","nome":"JUDIARA JUNEAU DO CECP","sexo":"","registro":"0655-AL"},{"cobertura":0,"codigo":24,"nascimento":"0012-02-20","raca":0,"manejo":"4583","nome":"JUSTARE OPINION DO CECP","sexo":"","registro":"0662-AL"},{"cobertura":0,"codigo":4,"nascimento":"0010-02-20","raca":0,"manejo":"4528","nome":"JUBA TALENTO DO CECP","sexo":"","registro":"1432-AF"},{"cobertura":0,"codigo":33,"nascimento":"0001-03-20","raca":0,"manejo":"5404","nome":"LABIADA OPINION DO CECP","sexo":"","registro":"0647-

Fig. 7 – Returned JSON with the animal classifications in the experiment.

We developed a web application (see Fig. 8) to provide researchers with an interface for the direct use of the *FeedEfficiencyService* services.

As previously stated, researchers working with feed efficiency need to evaluate the animals' performance throughout several experiments, besides evaluating animals' general performance in each experiment. The selection of several experiments may result in data overload, which requires visualization techniques. The visualization of the classification allows researchers to monitor the animals' performance in the experiment and monitor how they evolve throughout the experiments. One of the research challenges is to understand how this process occurs. Thus, by visualizing the classification, researchers can understand issues such as "*Has the animals' performance on the initial assessment been maintained? Would efficient animals be classified as efficient in other experiments?*". These analyses of classifications and their evolution in the experiments are possible through classification visualization, which presents the data from several experiments.

To exemplify the performance and evolution analyses, Fig. 9 presents information on the "4557_JUGADA" animal, and experiments 1, 3, and 6, which refer to the animals at 30 days, 56 days, and 80 days of life, respectively.

According to the color palette adopted, it can be seen in Fig. 9 that in experiment 1 (30 days feeding), the architecture classifies the animal as efficient in the four indices, and this performance was maintained in experiment 3 (56 days) for the *ECA* index but its classification dropped to intermediary. In experiment 6 (80 days of lactation), the animal showed an even more significant drop, becoming inefficient in all indices. The advantages of dynamic visualizations can be observed in terms of development, usability and performance. It is possible to use one single visualization model,

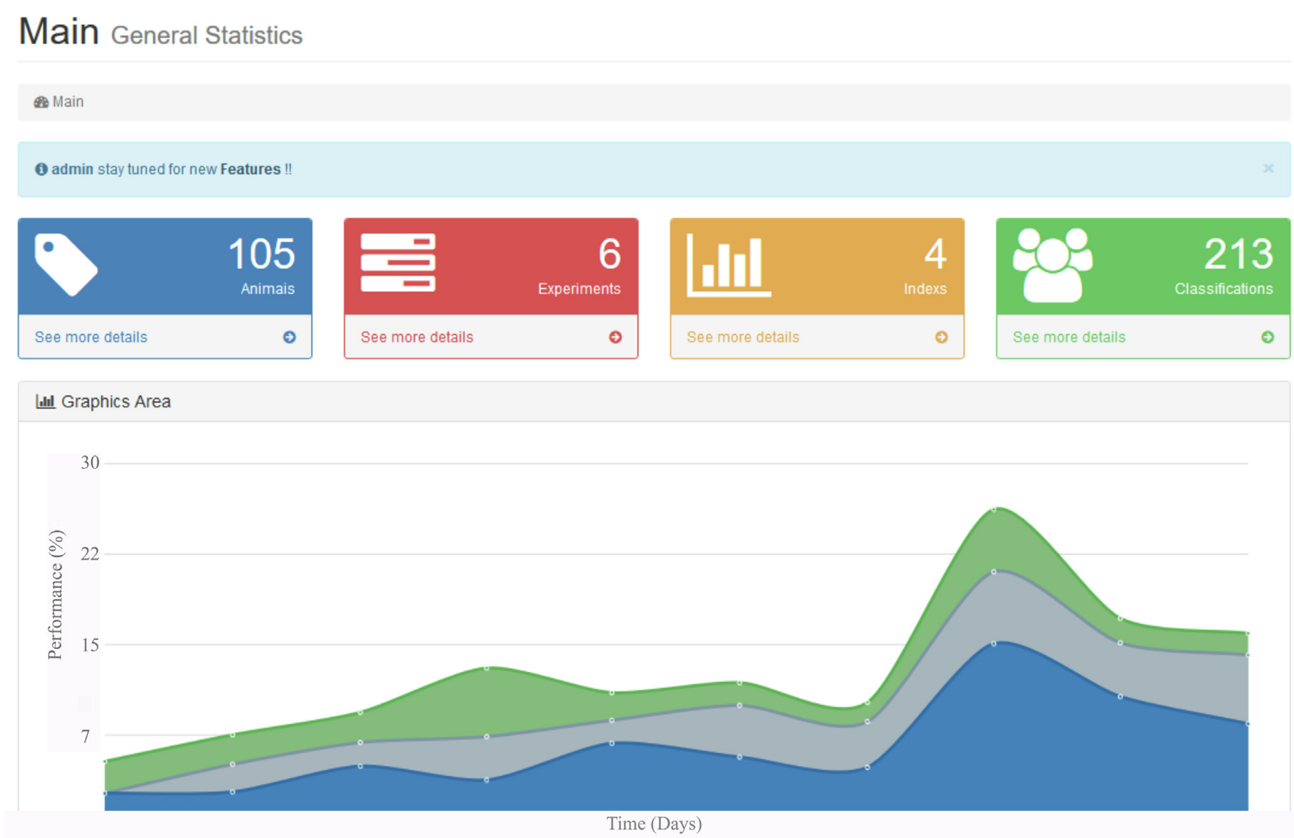

Fig. 8 – *FeedEfficiencyService* Web Application Interface for experiments data access.

i.e., use a single button to trigger the visualization, reducing the number of steps to build the desired visualization while also reducing processing and network resources.

On the other hand, considering the need to find the most appropriate feed efficiency index for dairy cattle, a comparative analysis of the indices is also necessary. Therefore, *FeedEfficiencyService* has a clustering visualization[8]. Similar to the previous visualization, this one also has interaction resources. Fig. 10 shows four frames obtained through interactions in the same visualization. Quadrant 1 of Fig. 10 presents the selected experiments; Quadrant 2 displays the four indices obtained for the experiment "aleitamento_80"; Quadrant 3 shows the quantity obtained for each classification, and finally Quadrant 4 presents the efficient animals in the CAR index in the "aleitamento-80" experiment.

2.3. *Evaluation*

The DSR methodology emphasizes the importance of a proper evaluation. Hevner et al. [15] state that the selection of evaluation methods must be matched with the designed artifact and the selected evaluation metrics. Such evaluation methods comprise observational research designs, divided into two types: case study and field study. Case study is the most appropriate type for our research, considering that we address a specific artifact (the *FeedEfficiencyService* architecture) in a business domain (scientific experiments on feed efficiency).

Case study is also the best instrument in this particular context because the present evaluation aims to verify if the proposed approach offers an adequate mechanism for the analysis and understanding of experiment data.

2.3.1. *Study definition*

The scope of this evaluation drew on the Goal, Question, Metrics (GQM) method, that is, to "**Analyze** the use of the *FeedEfficiencyService* architecture from the researchers' **point of view, in the context** of the Feed Efficiency Research Group of EMBRAPA".

Having defined the objective of the study, we formulated the research question (RQ), as already presented in the introduction section:

RQ. "*How does the FeedEfficiencyService architecture make it easy for researchers to analyze data from Feed Efficiency experiments?*".

Based on the above RQ, four secondary research questions (SRQ) were specified:

SRQ1. *How can the use of ontologies support feed efficiency research?*

[8] In this visualization, different levels of abstraction for each experiment are presented in the 4 quadrants, by the selection in each quadrant, i.e., selecting a given experiment (quadrant 1), selecting a specific index (quadrant 2), the values obtained for each index (quadrant 3), and quadrant 4 presents experiment-specific data for a selected index. Thus, there are different levels of abstraction, grouped in a single visualization.

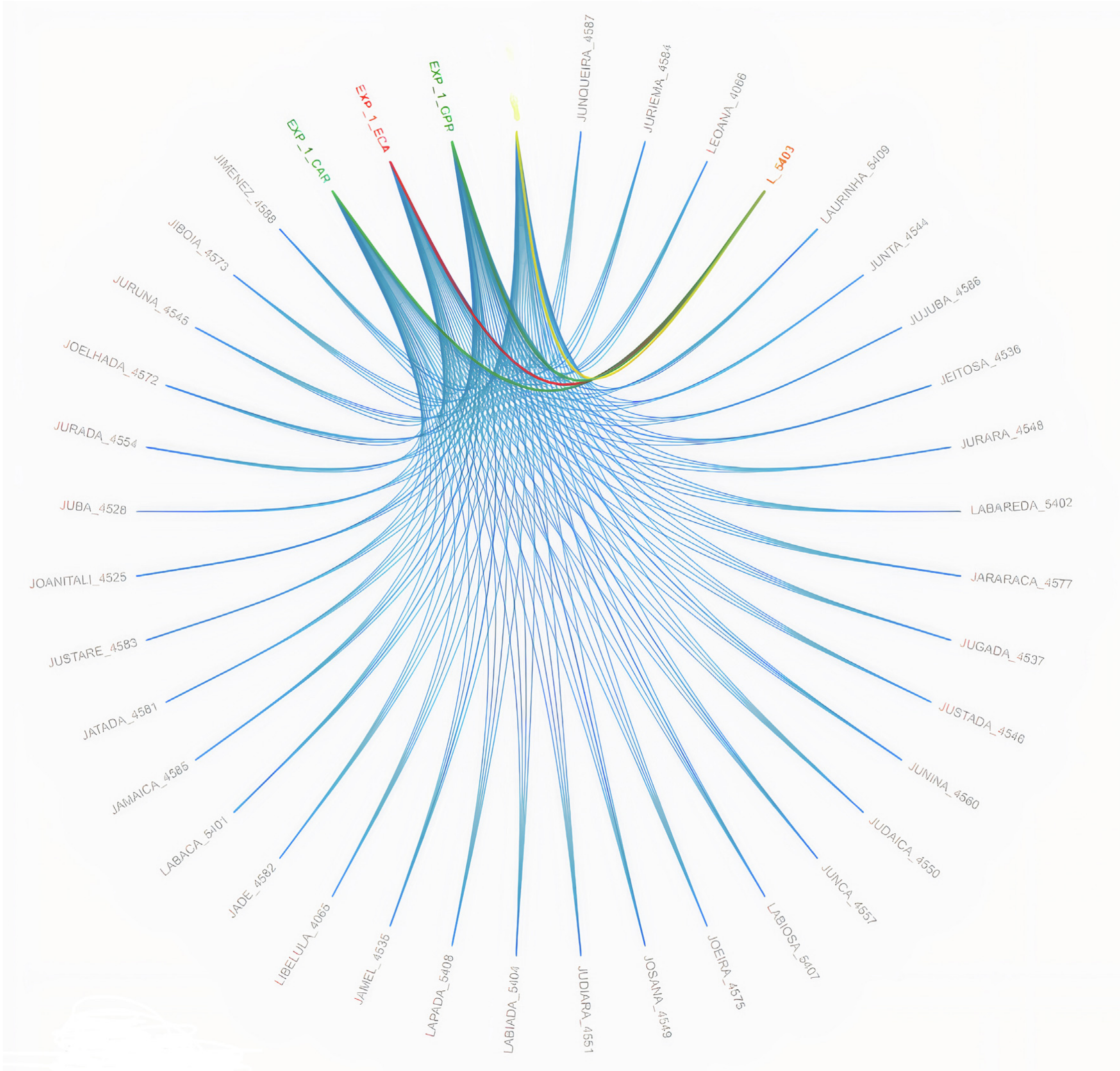

Fig. 9 – A case study of the *FeedEfficiencyService* architecture – Classification visualization of the animal known as "4557_JUGADA".

SRQ2. *How do classification visualizations contribute to researchers' analyses?*
SRQ3. *How do clustering visualizations contribute to researchers' analyses?*
SRQ4. *How was feed efficiency/animal nutrition-related knowledge relevant in the evaluation process of the architecture?*

2.3.2. *Planning*
2.3.2.1. *Context definition.* We conducted the case study using data from 6 experiments performed by researchers from EMBRAPA between 2014 and 2017. These data were made available from a server with the following configuration: Intel Core i7-5500 2.40 GHz processor, 16 GB DDR3 RAM, and Windows 10 64bit operating system. Subjects used a web interface accessible through personal computers to carry out the experiment. The duration of each experiment was approximately 30 min. The collection instruments are available in a public folder[9], and the *FeedEfficiencyService* architecture is available on GitHub[10].

The case study began with the selection of subjects. We created two groups: participants related to the context of animal nutrition/feed efficiency (Group A) and the subjects not related to this context (Group B). The second group was created so that we can observe whether researchers from other contexts, using the *FeedEfficiencyService* architecture, can analyze the experiments, which is a way of encouraging interac-

[9] https://www.dropbox.com/s/kasyqklt7hxgg5k/QUESTIONNAIRE%20AND%20CHARACTERIZATION%20OF%20SUBJECTS.docx?dl=0

[10] The architecture design was divided into two parts: the first part contains the implementation base classes, the access to the data layer and its configurations; the second part contains the services and the web application.https://github.com/heitormagaldi/FeedEfficiencyServiceBasehttps://github.com/heitormagaldi/FeedEfficiencyService

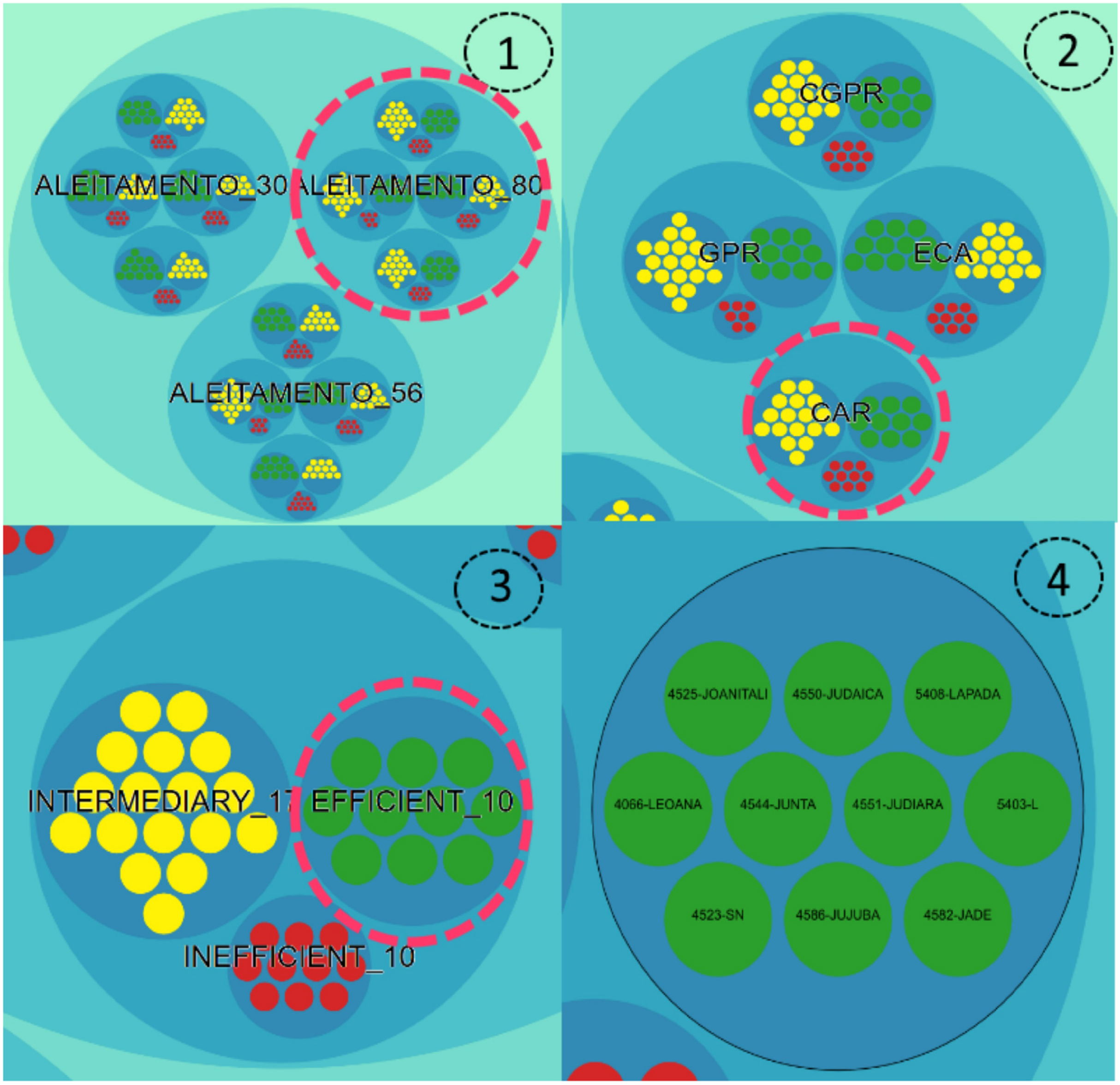

Fig. 10 – A case study of the *FeedEfficiencyService* architecture – interactive clustering visualization.

tion between different EMBRAPA research centers and external researchers. For the selection of the subjects in each group, a characterization form[11] was used.

Once the groups were defined, we formulated questions in order to address the research questions previously presented. We specified two sets of questions[12]: the first, addressing aspects prior to the architecture, and the second, considering aspects after the architecture had been adopted.

The questionnaires were the same for both groups. However, because we had one group of subjects with no relation to the animal nutrition/feed efficiency context, a brief explanation about the domain was necessary, presenting the researchers' daily routine and explaining how data analyses used to be carried out before adopting the *FeedEfficiencyService* architecture. Then, *FeedEfficiencyService* was presented to the subjects, who were given access to the available services. To describe the functionalities of the architecture, we used two scenarios.

The first scenario focused on access to experiments and animal data analyses, communication between the services, and the animals' classification under the efficiency indices (see Fig. 9 and see Fig. 10). The second scenario presented the classification visualization feature as a support for analyzing the animals' evolution throughout the experiments. This scenario used data from the six experiments. These were conducted at different stages of the animals' lives (see Fig. 11).

2.3.2.2. *Selection of subjects.* The study had the voluntary participation of 35 subjects, of which 16 were related to the

[11] https://www.dropbox.com/s/kasyqklt7hxgg5k/QUESTIONNAIRE%20AND%20CHARACTERIZATION%20OF%20SUBJECTS.docx?dl=0

[12] https://www.dropbox.com/s/kasyqklt7hxgg5k/QUESTIONNAIRE%20AND%20CHARACTERIZATION%20OF%20SUBJECTS.docx?dl=0

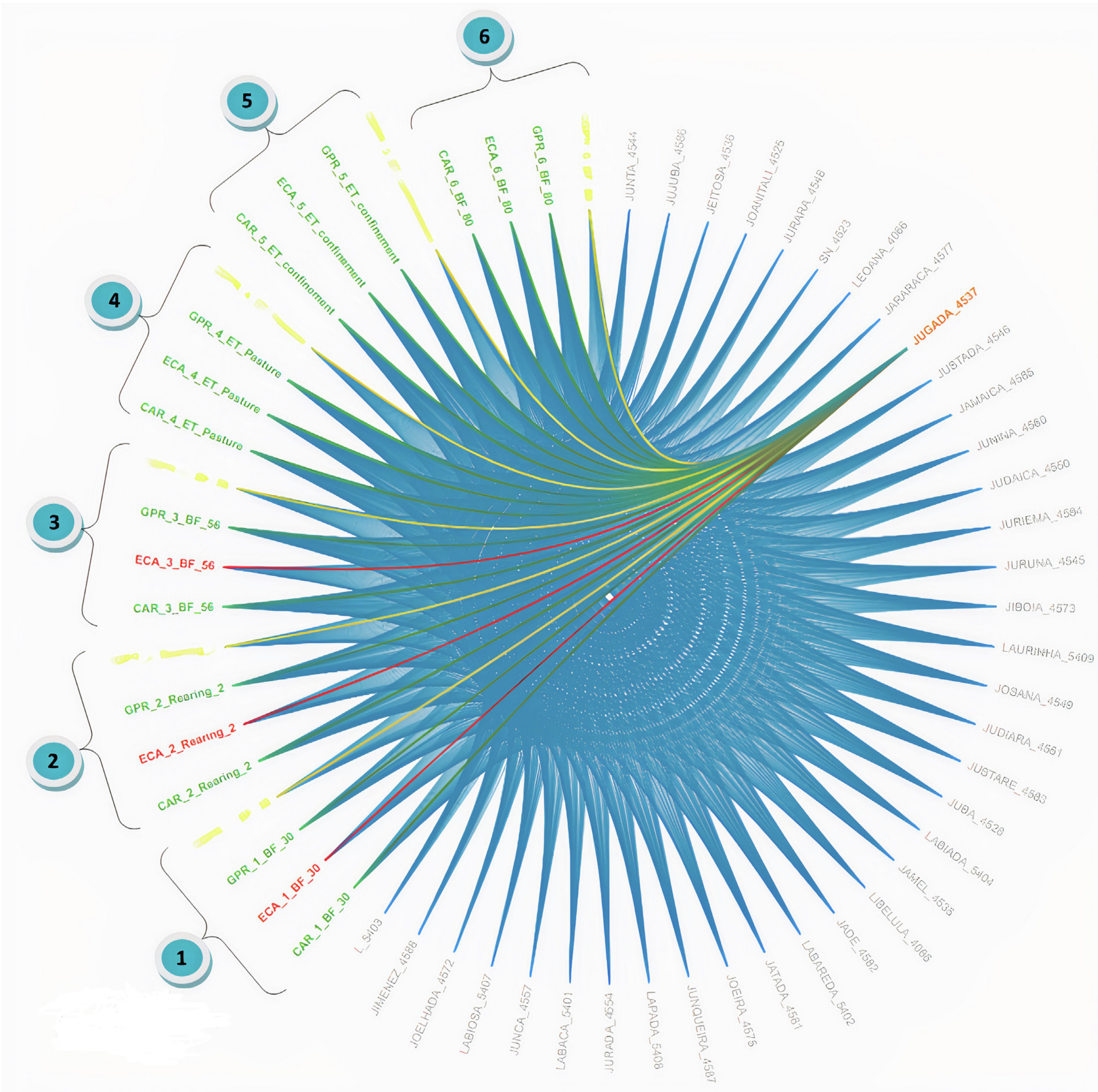

Fig. 11 – A case study of the *FeedEfficiencyService* architecture, classification visualization to support analyses of animals' evolution in experiments (numbered in the visualization).

context of animal nutrition/feed efficiency and 19 were not related. Among those 16, there were subjects holding degrees in Zootechnics, Veterinary Science, and Computing. Among the nonrelated ones, 4 were researchers from EMBRAPA, but from other contexts. Although these researchers do not work directly with feed efficiency research, they have degrees in Veterinary Science and Zootechnics, and carry out statistics and research on genetic engineering of milk, precision farming, and genetic enhancement. The other 15 nonrelated subjects have degrees in Computer Science and work with scientific experiments.

The subjects nonrelated to feed efficiency are of great importance for this study, as they will allow verifying whether researchers from other contexts, using the *FeedEfficiencyService* architecture, can analyze the experiments, hence encouraging the reuse of data among the different EMBRAPA research centers. We elaborated a group of questions to understand this group's evaluation from outside the feed efficiency context.

2.3.2.3. Data collection sources. A case study can rely on different data sources, and these data can be obtained through direct, indirect, and independent methods. For this case study, we chose to collect data through the direct method, considering direct interaction with the subjects and adopting a questionnaire. As for the questionnaire, we adopted the semi-structured category as it encompasses open-ended and closed-ended questions.

Open-ended questions aimed to detect relevant aspects not considered in the closed-ended questions. The closed-ended questions were answered through a scale of values, ranging from 1 to 5. Value 1 refers to the answers that dis-

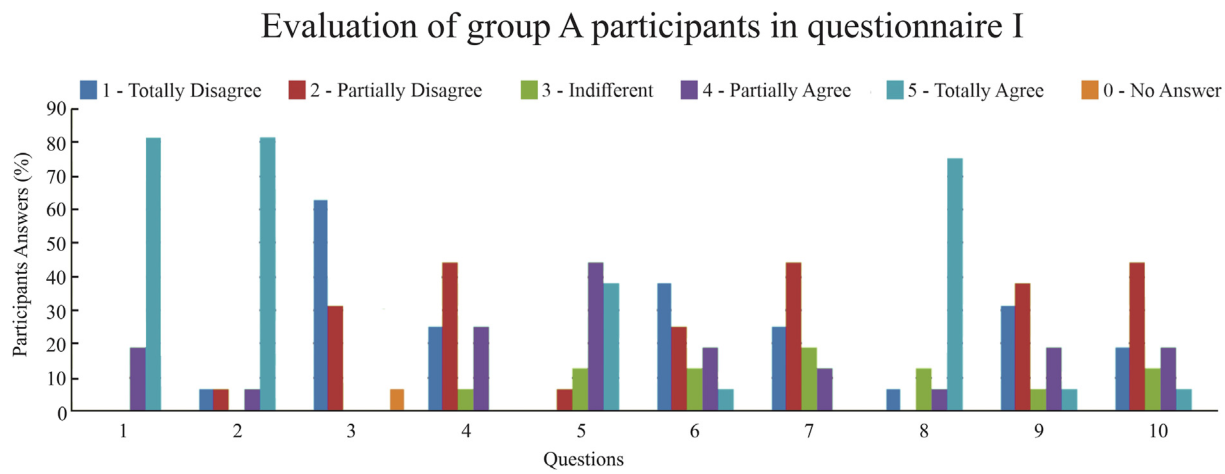

Fig. 12 – Answers from Group A to Questionnaire I.

agree with the statement and value 5 to answers that totally agree with the statement.

2.3.3. *Execution*

We conducted the case study in three stages according to the availability of subjects. The first stage took place at the headquarters of EMBRAPA, Brazil, and was attended by 7 subjects, being 6 researchers from the institution and 1 from an external research center. The second stage took place at the Federal University of Juiz de Fora, Brazil, and with 14 external researchers. Finally, the third stage took place in the experimental area of EMBRAPA, Brazil, with 14 subjects.

The *FeedEfficiencyService* architecture was presented at each stage, allowing access to data from the feed efficiency experiments carried out at EMBRAPA. Furthermore, the group not related to the context of feed efficiency received information on the concept of efficient animal, the classifications adopted by EMBRAPA, and the rationale of the study.

During the case study, each subject randomly selected the visualization context (clustering or classification). No script was used, nor any interference was made in each subject's choice.

It took 30 min on average for participants to complete each questionnaire, except for the first stage of the study, which took them about 150 min. Of the 6 researchers from EMBRAPA who participated in the case study at the first stage, 4 were not related to the feed efficiency context, and they considered the possible impact of the *FeedEfficiencyService* architecture on other research contexts at EMBRAPA. This directly influenced the duration of the case study.

All the data under analysis were obtained through the questionnaires and from direct observation of the subjects. We grouped these data and organized them in a tabular form[13].

3. Results and discussion

Considering the *FeedEfficiencyService* architecture and the previous evaluation, we analyzed the results of the questionnaires through a qualitative method. Questions regarding the method for evaluating animals and experiments before the *FeedEfficiencyService* architecture had been adopted by EMBRAPA were analyzed based on the Group-A questionnaires. This group had subjects with experience in a previous scenario without *FeedEfficiencyService*. We considered the data from Groups A and B to evaluate the questions related to the contents and explanations of the classifications, visualizations and analyses.

Fig. 12 presents the data obtained after completion of Questionnaire I[14], related to data processing before the use of the *FeedEfficiencyService* architecture. Questionnaire II[15] encompasses data related to the use of the *FeedEfficiencyService* architecture, shown in Fig. 13. For the sake of clarity, 0 in the graphs means no answer.

The objective of the analysis with Group B was to evaluate whether researchers from other contexts, using the *FeedEfficiencyService* architecture, could analyze the experiments, which is a way of encouraging the reuse of experimental data between different research groups. Fig. 14 shows the data obtained through Questionnaire II for Group B. Only Group A answered Questionnaire I, and both groups answered Questionnaire II. The analyses of the data obtained from these groups are presented below.

The answers related to questions 1, 2 and 5 revealed problems such as the need for knowledge about statistical tools, the absence of data storage and patterns, and the difficulty of associating the same animals in different experiments. Questions 3, 7, 8 and 9 addressed aspects of the analysis of the experiments. In these questions, the subjects agreed on the difficulty of classifying the animals, monitoring the animals' performance, obtaining the animals' indices and comparing the experiments, an issue that stems from the heterogeneity of the databases and the lack of a standard procedure to conduct experiments. Regarding the data and the form of storage thereof, the answers to questions 4 and 6 evinced the subjects' dissatisfaction with the previous form of storage of and access to experimental data.

[13] https://www.dropbox.com/s/wilbaxwkz3f5975/analise%20dos%20dados.xlsx?dl=0

[14] https://www.dropbox.com/s/wilbaxwkz3f5975/analise%20dos%20dados.xlsx?dl=0

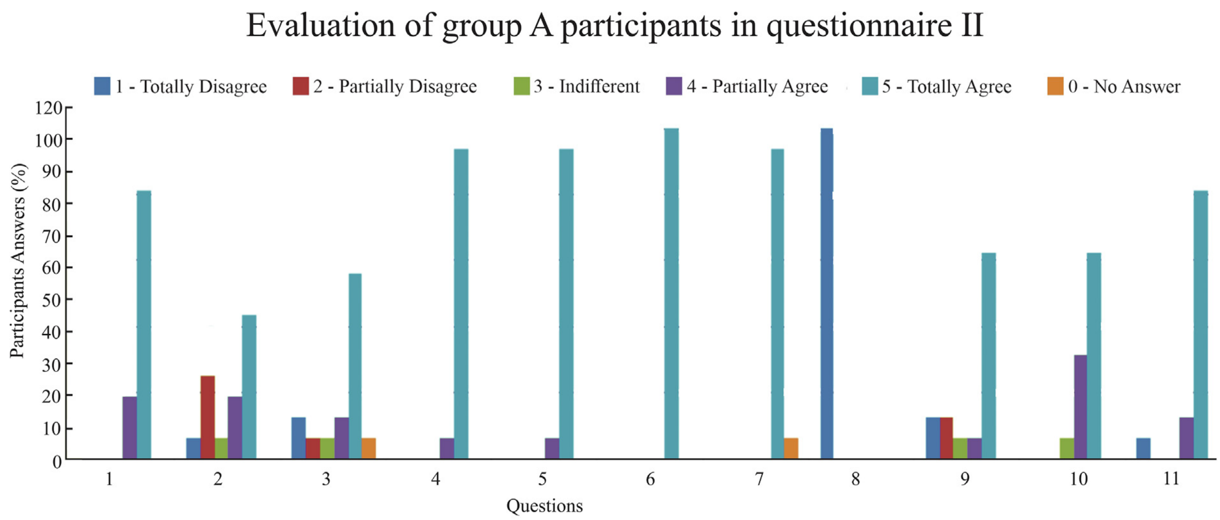

Fig. 13 – Answers from Group A to Questionnaire II.

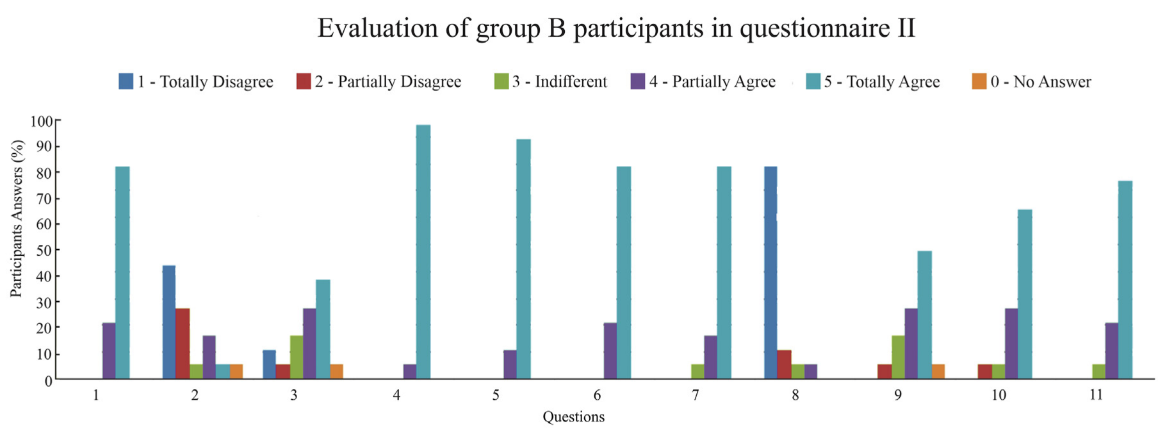

Fig. 14 – Answers from Group B to Questionnaire II.

The open-ended questions allowed observing that most subjects pointed out the possibility of errors in data transcription or storage in heterogeneous databases, which is a recurrent concern for them as researchers. Another difficulty raised regards the analysis of the experiments and the possibility of comparison between them. This difficulty was justified by the absence of patterns in data storage and availability.

The subjects emphasized that the architecture in question will be of great importance for the evaluations of animals, comparisons between experiments, security of data analyses, and the standardization and organization of experimental data. Some subjects highlighted the academic and industrial importance of the architecture for rural producers. However, a point to consider is that the question was answered by feed efficiency researchers, not by rural producers.

Data triangulation[15] analysis was used to analyze the answers of Groups A and B and to increase the accuracy and strengthen the validity of this study. Question 1 (from Questionnaire II) was not considered for Group B. On the other hand, 100 % of the subjects in Group A corroborated this question and agreed that the architecture brought agility to the analysis of experiments.

[15] Runeson [30] point out that triangulation is important to increase accuracy and strengthen the validity of empirical research and is of great importance in qualitative data analyses. The authors define triangulation as an analysis from multiple perspectives, providing a broader view of the object of study. Data triangulation uses more than one data source or collection at different times. Observation triangulation uses more than one observer in the study. The triangulation methodology combines different types of data collection methods (qualitative and quantitative); and the triangulation theory uses alternative theories or points of view.

Table 3 – Results of comparison between Groups A and B.

Question	Test	*p-value*	Results
1	Anova	0.870	Acceptance of H0
2	Anova	0.001	Acceptance of H1
3	Anova	0.764	Acceptance of H0
4	Anova	0.904	Acceptance of H0
5	Anova	0.664	Acceptance of H0
6	Anova /Kruskul-Wallis	0.053 / 0.055	Acceptance of H0*
7	Anova	0.878	Acceptance of H0
8	Anova / Kruskul-Wallis	0.086 / 0.055	Acceptance of H0*
9	Anova	0.899	Acceptance of H0
10	Anova	0.730	Acceptance of H0
11	Kruskul-Wallis	0.741	Acceptance of H0

Question 2 (from Questionnaire II) discussed the aspects before and after the use of the *FeedEfficiencyService* architecture. In this study, most of the subjects in Group A (73.68 %) disregarded the architecture's statistical services, and either agreed with the statement or were indifferent to it. In contrast, most subjects in Group B noticed the architecture's statistical services and disagreed with the statement. Therefore, this suggests that future work may lead to a better understanding of this difficulty, as it was not clear whether the subjects in Group A made a correct interpretation of the question.

Question 3 evaluated the standardization of experiments (Questionnaire II). We perceived that the architecture provided a better procedure for storing animal and experimental data because both groups agreed with the statement.

Questions 4, 5, 6, 7, 8 and 9 evaluated the clustering and classification visualizations. The analyses of questions 4, 5, 6 and 7 showed no divergent opinions between the groups, hence an agreement rate of 100 % among the subjects. Therefore, the architecture facilitates the observation and comparison of animal data, and the color highlights aid in the analysis. Although Group B was not knowledgeable about feed efficiency, it is worth mentioning that they managed to evaluate the characteristics stated in the above questions.

Regarding Question 8, Group A was unanimous in rejecting the statement that the architecture had not improved animal analysis throughout their experiments. In Group B, only 5.27 % of the subjects were indifferent to that, only 5.27 % agreed partially with the statement, and most of them rejected the statement. We can argue that Group B contained subjects from outside the context of feed efficiency, with limited knowledge of any previous efficiency methodology or analysis processes, and the lack of such knowledge may have influenced their evaluation of the system.

In Question 9, 68.75 % of the subjects in Group A and 73.68 % in Group B agreed that clustering visualization may help researchers to identify imbalance in index distribution. Clustering visualization can be useful in choosing the efficiency index that is most appropriate for the experiment. However, the indices used are intended for beef cattle, and an index that is most suitable for the context of the dairy cattle is still unknown.

Through the FEO, the SWRL rules, and inference engines, it was possible to obtain the animals' classification under four feed efficiency indices (*CAR, ECA, GPR, CGPR*) and classify them under the labels of efficient, intermediary and inefficient. This statement was assessed in Questions 10 and 11. In Question 10, 93.75 % of the subjects in Group A and 89.46 % in Group B agreed that these labels improved the analyses. Moreover, in Question 11, 93.75 % of subjects in Group A and 94.73 % in Group B agreed that the ontology can encourage interaction with other researchers. Consequently, this ontology can aid in the reuse of experimental data by researchers from other contexts.

Therefore, considering the previous analyses, we can answer the research questions, and an understanding of how the architecture can impact EMBRAPA researchers' daily routine was successfully achieved.

> RQ. "*How does the FeedEfficiencyService architecture make it easy for researchers to analyze data from Feed Efficiency experiments?*".

The use of the *FeedEfficiencyService* architecture can simplify analyses, reducing complexity and bringing agility to data evaluations, providing organization and standardization in the conducting of experiments and reliability in the analyses, according to the answers to the questionnaires. Besides, classification and clustering visualizations facilitate data analyses of animals and experiments. They support the comparison of animals' evolution and experiments in the same visualization, through interactions and the color palette, visually providing the researcher with the experiments' general classification.

> SRQ1. *How can the use of ontologies support feed efficiency research?*

Through the FEO, the SWRL rules, and the inference engines, the architecture classifies and labels the animal data, providing the researcher with an easy way to identify efficient animals, share and reuse data. Besides, there is evidence of improvement in the interaction between researchers from different contexts. However, another evaluation instrument will be necessary to substantiate this statement, as the present study did not specifically address this aspect.

> SRQ2. *How do classification visualizations contribute to researchers' analysis?*

Classification visualization allows the researcher to analyze more experiments simultaneously in the same data chart. As a result, it enables evaluating the animals' performance throughout the experiments in a single visualization and provides the researcher with information necessary for its analysis, according to the answers to the questionnaires. Another point to consider, also present in the visualizations, is the palette of colors associated with the animals' classifications and indices. This visualization facilitates analyses and supports researchers, helping them find answers to questions such as: "*Has the animals' performance on the initial assessment been maintained? Would efficient animals be classified as efficient in other experiments, or as intermediary or inefficient?*".

SRQ3. *How do clustering visualizations contribute to researchers' analyses?*

Clustering visualization provides the researcher with the indices' distribution analysis in one or several experiments. The color palette for the classification into efficient, intermediary and inefficient is also present. This visualization supports the researcher in identifying the most appropriate efficiency index in the dairy context, as confirmed by the questionnaires.

After completing the analyses of Groups A and B, it is relevant to compare them with Questionnaire II to find evidence of improvement with the use of the proposed architecture, and, consequently, to investigate whether the participants' knowledge related to the feed efficiency context (Group A) interfered with the results found.

Norman [29] stresses that statistical methods are usually criticized in studies and these critiques are based on sample size, normality of distribution, and Likert scale data. However, other studies show that parametric statistics are robust and that such statistical methods can be used without the concern of producing incorrect results. Another point to be considered when choosing a method is the homoscedasticity of the data, which refers to the concentration of data around the regression line of the model. In this way, the choice should include the normality and homoscedasticity of the data.

To evaluate the questions answered by Groups A and B, we used two tests: Anova for homocystic samples with normal distribution, and Kruskul-Wallis for samples that did not present these characteristics. Thus, two hypotheses were defined for this evaluation:

H0 (null hypothesis): the samples are the same.
H1 (alternative hypothesis): the samples are NOT equal.

Because Questions 6 and 8 of Group A did not meet the requirements of the previous evaluations, we adopted two tests in this new evaluation. The results (Table 3) demonstrate that Question 2: "*Knowledge about some statistical tool remains indispensable for obtaining the CAR, GPR and CGPR indices*" was the only question that accepted the H1 (alternative hypothesis), confirming the difference between the groups.

Additionally, based on these results, we observed that in Questions 6 and 8, p-value was close to the significance level (0.05) but confirmed the H0 (null hypothesis), rejecting the difference between the groups. In the other cases, with a more considerable margin, the H0 (null hypothesis) was also confirmed. The full results of Table 3 are available at a public folder[16].

A significance level of 0.05 (or 5 %) showed that the groups could be considered statistically equal. Thus, knowledge about feed efficiency context did not interfere in the evaluation of the architecture.

Through the statistical comparison between Groups A and B, it was possible to answer the last research question and understand how previous knowledge impacted the evaluation process of the *FeedEfficiencyService* architecture.

SRQ4. *How was feed efficiency/animal nutrition-related knowledge relevant in the evaluation process of the architecture?*

The participants' knowledge about animal feed efficiency was not relevant in the evaluation process of the architecture because there was no difference in the groups' responses (*p-value* > 0.05). However, to a certain extent, this knowledge was essential for the evaluation regarding Questions 6 and 8, which addressed the contributions of the *FeedEfficiencyService* architecture to researchers concerning the efficiency of dairy cattle feed at EMBRAPA and, indirectly, made a comparison with the methodology adopted prior to the use of *FeedEfficiencyService*. Thus, Group B evaluations for these questions were considered statistically equal to those for Group A, but with percentages close to acceptance of H1 (alternative hypothesis).

3.1. *Threats to validity*

In terms of **construction validity**, we can mention that the architecture was developed to meet the researchers' needs as regards dairy cattle feed efficiency, and we did not consider aspects of other domains. Therefore, new studies in new contexts are necessary to measure the architecture's impact on data interaction and reuse. Despite using triangulation to analyze the answers, the use of only one instrument of data collection can be considered a threat to validity. Probably, employing multiple instruments, both qualitative and quantitative, could enrich the evaluations.

In terms of **internal validity**, we can state that the instruments used in this case study were chosen using the resources available in the *FeedEfficiencyService* architecture. In this vein, in other contexts, it is necessary to evaluate new instruments. All the subjects took a single assessment of the architecture. Therefore, there is no increase in response capacity, demotivation of the evaluation process, or the possibility of a same participant being present in both groups. The groups were selected based on the aforementioned characterization form. However, it is worth mentioning that the groups did not have an equal number of participants, and therefore the selection of subjects could be identified as a threat to validity. There was no interaction between evaluator and participants in any of the three evaluation stages of the case

[16] https://www.dropbox.com/s/wo96hy2cv72mw6t/STATISTICAL%20EVALUATION.docx?dl=0

study. However, the participants requested clarification on some aspects, and the evaluator did not consider it as a threat. We need to emphasize that we applied the same questionnaires to both groups.

The number of participants in the case study is a threat, as a larger number of participants could influence the results. This study relied on the participation of researchers with a computer science background, composing Group B (participants not related to the context of feed efficiency). Although that group does not include all participants, this may have influenced the results.

The analyses and conclusions expressed in the present study can only be applied to its particular context (external validity). However, it is possible to transfer the conclusions and analyses of this study to case studies with similar scenarios. As a result, it is necessary to conduct additional evaluations from other points of view. Aspects not addressed in this study may provide additional evidence.

4. Conclusion

This paper presented the *FeedEfficiencyService* architecture designed to support researchers in analyzing, classifying and integrating experiments into feed efficiency research at EMBRAPA.

We used the Design Science Research methodology to conduct the research. We carried out a systematic literature review and identified deficiencies in the existing approaches. As a result, open research challenges were raised. None of the reviewed proposals provides an approach that captures, integrates, analyzes and visualizes data from multiple experiments to facilitate understanding (**a problem space and a mechanism posed or enacted to find an effective solution**). Therefore, we proposed the creation of an architecture (**creation of an artifact**), named *FeedEfficiencyService*, to assist scientists in data integration and analysis to understand their data and the relationships between experiments (**problem domain**). To evaluate the feasibility of our approach, we conducted a study with data from scientific experiments performed at EMBRAPA (**an evaluation of the artifact is crucial**). The results point to the feasibility of the proposal.

These results verify that "the integration of feed efficiency experiments through an architecture that uses ontologies and visualization techniques can support the analysis and evaluation of animals in experiments and promote the reuse of data by researchers from other contexts", underpinning the proposed research question.

As specific contributions of the *FeedEfficiencyService* architecture, we can highlight:

- Creation of a Feed Efficiency Ontology named FEO;
- Specification of SWRL rules and inference engines for classifying animal data;
- Use of an integration model for the storage of experiments and elimination of data heterogeneity;
- Definition of a generic architectural model that can also be used by other research groups out of the feed efficiency context;
- Creation of visualizations to help in the analyses of feed efficiency data;
- Definition of an API for services offered by the proposed architecture, based on SAAS principles; and
- Development of a web application to simplify access for researchers.

Researchers evaluated the *FeedEfficiencyService* architecture and emphasized its importance in the analyses of animal data, comparisons of experiments, reliability of data analyses, and standardization and organization of experimental data. However, we built the *FeedEfficiencyService* architecture to meet relevant needs in the realm of feed efficiency research, and its functionalities are restricted to the particularities of the research group in question. Thus, its use cannot be generalized, although it could be easily used in other contexts.

Regarding the volume of data used in this study, we can affirm that the ontology met the expectations. However, we know that the ontology has restrictions for processing large volumes of data, which can cause slowness in the processing of visualizations and returning of services.

The creation of new ontologies from the services added to this architecture could result in new knowledge. Likewise, the addition of computational intelligence techniques could also be relevant to this process of knowledge discovery.

Finally, the impacts of this architecture on the production and dissemination of information, the contributions thereof to decision-making, and the interactions that it allows between researchers from other contexts should be further studied. In this vein, it is necessary to conduct new experiments to evaluate the impact of the proposed architecture in these new contexts.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

We would like to thank the researchers who participated in the evaluation, Embrapa and Mr. Rian das Dores Alves for image manipulation. This work was partially funded by UFJF/Brazil, CNPq/Brazil (grant 311595/2019-7), and FAPEMIG/Brazil (grant APQ-02685-17), (grant APQ-02194-18).

REFERENCES

[1] Paddock ZD. Energy Expenditure in Growing Heifers with Divergent Residual Feed Intake Phenotypes. Effects and Interaction of Metaphylactic Treatment and Temperament on Receiving Steers. Doctor Thesis. Texas: A & M University; 2011.
[2] Herrero M, Gerber P, Vellinga T, Garnett T, Leip A, Opio C, et al. Livestock and greenhouse gas emissions: The importance of getting the numbers right. Anim Feed Sci Technol 2011;166 (167):779–82. https://doi.org/10.1016/j.anifeedsci.2011.04.083.

[3] Boichard D, Brochard M. New phenotypes for new breeding goals in dairy cattle. The Animal Consortium 2012;6 (4):544–50. https://doi.org/10.1017/S1751731112000018.
[4] Arthur JPF, Herd RM. Residual feed intake in beef cattle. Revista Brasileira de Zootecnia 2008;37(SPE):269–79 (in Portuguese).
[5] Montanholi YR, Odongo NE, Swanson KC, Schenkel FS, Mcbride BW, Miller SP. Application of infrared thermography as an indicator of heat and methane production and its use in the study of skin temperature in response to physiological events in dairy cattle (Bos taurus). J Therm Biol 2008;33(8):468–75. https://doi.org/10.1016/j.jtherbio.2008.09.001.
[6] Crowley JJ, Mcgee M, Kenny DA, Crews Jr DH, Evans RD, Berry DP. Phenotypic and genetic parameters for different measures of feed efficiency in different breeds of Irish performance-tested beef bulls. J Anim Sci 2010;88(3):885–94. https://doi.org/10.2527/jas.2009-1852.
[7] Berry DP, Crowley JJ. Cell biology symposium: genetics of feed efficiency in dairy and beef cattle. J Anim Sci 2013;91 (4):1594–613. https://doi.org/10.2527/jas.2012-5862.
[8] Pryce JE, Gonzalez-Recio O, Nieuwolf G, Wales WJ, Coffey MP, Hayes BJ, et al. Hot Topic: Definition and implementation of a breeding value for feed efficiency in dairy cows. J Dairy Sci 2015;98(10):7340–50. https://doi.org/10.3168/jds.2015-9621.
[9] Connor E. Invited review: Improving feed efficiency in dairy production: Challenges and possibilities. Animal 2015;9:395–408. https://doi.org/10.1017/S1751731114002997.
[10] Gomes RC, Santana M, Ferraz J, Leme P, Silva S. Food intake and feed efficiency of beef cattle and sheep. Funpec-Editora, Ribeirão Preto. 2012;77 (in Portuguese).
[11] Grion AL, Mercadante M, Cyrilo N, Bonilha S, Magnani E, Branco R. Selection for feed efficiency traits and correlated genetic responses in feed intake and weight gain of Nellore cattle. J Anim Sci 2014;92(3):955–65.
[12] Bonilha S, Cyrillo N, Santos G, Branco R, Ribeiro E, Mercadante M. Feed efficiency, blood parameters, and ingestive behavior of young Nellore males and females. Trop Anim Health Prod 2015;47(7):1381–9. https://doi.org/10.1007/s11250-015-0875-7.
[13] Horrocks I, Patel-Schneider PF, Boley H, Tabet S, Grosof B, Dean MSWRL. A semantic web rule language combining OWL and RuleML. W3C Member submission. 2004;21(79).
[14] Drury B, Fernandes R, Moura M, Lopes A. A survey of semantic web technology for agriculture. Information Processing in Agriculture 2019;6(4):487–501. https://doi.org/10.1016/j.inpa.2019.02.001.
[15] Hevner AR, March ST, Jinsoo P, Ram S. Design science in information systems research. MIS Quarterly 2004;28 (1):75–105. https://doi.org/10.2307/25148625.
[16] Guarino N. Formal ontology and information systems. In: In: Proceedings of Formal Ontology in Information Systems (FOIS). p. 81–97.
[17] Jonqueta C, Touleta A, Arnaudc E, Aubind S, Dzale E, Emoneta V, et al. AgroPortal: A vocabulary and ontology repository for agronomy. Computers and Eletronics in Agriculture 2018;144:126–43. https://doi.org/10.1016/j.compag.2017.10.012.
[18] Berners-Lee T, Hendler J, Lassila O. Scientific American: Feature Article. The. Semantic Web. 2001.
[19] Mcguinness D. OWL 2.0 overview. W3C recommendation. link: https://www.w3.org/OWL/. 2012.
[20] W3C, SHACL (Shapes Constraint Language). link: https://www.w3.org/TR/shacl/. 2017
[21] Kitchenham B, Charters S. Guidelines for performing systematic literature reviews in software engineering (version 2.3). Technical Report. Keele University and University of Durham; 2007.
[22] Tomic D, Drenjanac D, Wöber W, Hörmann S, Auer W. Enabling Semantic Web for Precision Agriculture: a Showcase of the Project agriOpenLink. CEUR Workshop Proceedings. 2015:26–9.
[23] Miah SJ, Gammack J, Kerr D. Ontology development for context-sensitive decision support. In: Proc. of Third International Conference on Semantics, Knowledge and Grid, IEEE. 2007. p.475-478.
[24] Parrot L, Lacroix R, Wade KM. Design considerations for the implementation of multi-agent systems in the dairy industry. Comput Electron Agric 2003;38(2):79–98. https://doi.org/10.1016/S0168-1699(02)00139-4.
[25] Sivamani S, Hong-Geun K, Changsun S, Jangwoo P, Youngyun C. A Balance Diet Management System for Livestock using an OWL-based Ontology Model. Indian Journal of Science and Technology, 9(20). In: Proceedings of the 2010 ACM-IEEE International Symposium on Empirical Software Engineering and Measurement. 2010. 33.
[26] Janssen S, Andersen E, Athanasiadis IN, Van Ittersum MK. A database for integrated assessment of European agricultural systems. Environ Sci Policy 2009;12(5):573–87. https://doi.org/10.1016/j.envsci.2009.01.007.
[27] Hulsegge B, Smits MA, Pas MFW, Woleders H. Contributions to an animal trait ontology. J Anim Sci 2012;90(6):2061–6. https://doi.org/10.2527/jas.2011-4251.
[28] Bushmann F. Patterns of Software Architecture. A system of patterns, Wiley; 1, 1996.
[29] Norman G. Likert scales, levels of measurement and the "laws" of statistics. Adv Health Sci Educ 2010;15(5):625–32. https://doi.org/10.1007/s10459-010-9222-y.
[30] Runeson P. Case study research in software engineering: Guidelines and examples. John Wiley & Sons; 2012.