A new semantic resource responding to the principles of Open Science: The meat thesaurus as an IT tool for dialogue between sector actors

Moise Kombolo Ngah, Jérémy Yon, François Landrieu, Brigitte Richon, Sophie Aubin, Jean-François J.-F. Hocquette

A NEW SEMANTIC RESOURCE RESPONDING TO THE PRINCIPLES OF OPEN SCIENCE: THE MEAT THESAURUS AS AN IT TOOL FOR DIALOGUE BETWEEN SECTOR ACTORS.

Running title: THE MEAT THESAURUS: A NEW RESOURCE ACCESSIBLE TO ALL

Authors : Moise KOMBOLO[1], Jérémy YON[2], François LANDRIEU[3], Brigitte RICHON[3], Sophie AUBIN[2], Jean-François HOCQUETTE[1]

[1] INRAE, Université Clermont Auvergne, VetAgro Sup, UMRH, 63122, Saint-Genès-Champanelle, France

[2]INRAE, DipSO, 42 rue Georges Morel, 49070 Beaucouzé, France,

[3]Académie de la Viande, 207 rue de Bercy, 75012 Paris,

Abstract:

Nowadays, it is important to make the results of scientific research accessible in a simple and understandable way according to the Open Science policy. This movement uses tools to enhance findability and interoperability of data. This paper describes the transformation of the meat dictionary published by the French Meat Academy as a book into a machine actionable and freely accessible terminological resource based on the SKOS standard format. This thesaurus contains 1567 concepts describing the meat production chain. This work was carried out by experts in semantic web, meat biology and meat vocabulary. This thesaurus can be used to index articles, journals and datasets, thus facilitating consultation; it can also be used to facilitate interoperability of the indexed datasets and provide contextual definitions for building ontologies, i.e. formal descriptions of knowledge for reasoning on data. The thesaurus can be useful to enrich other vocabularies with new knowledge, such as French specificities in terms of meat cuts or definitions.

1. Introduction

Around the world, meat industries have become highly diversified with new products and technologies. Despite a decline in consumption in Europe, World meat production continues to grow globally, driven by increased population (Hocquette and Chatellier, 2011). These changes have brought a wide variety of products and manufacturing processes for which it has become necessary to have definitions that are as clear as possible (Seman et al., 2018) and regularly updated. Indeed, some terms used in the sector remain ambiguous and their usage may also vary between countries or between scientists, consumers, journalists or industrial actors (Seman et al., 2018). This lack of consensus on shared definitions can have an impact on research work, particularly for predictive biology approaches (Hocquette et al., 2020). The need for clear definition is also essential for commercial trade, and particularly for international trade where local source or purchase descriptions may differ widely. The word meat is itself the subject of controversy with the appearance of meat substitutes, notably "cultured meat", which some prefer to call "muscle fibres" or "cultured muscle" (Chriki and Hocquette, 2020).

Various resources have therefore been created with the aim of clarifying the vocabulary of the field. For instance, the UNECE (United Nations Economic Commission for Europe) provides standardised descriptions of cuts from different butchery animals (www.unece.org/trade/wp7/Meat-Standards) for international meat trade. The Encyclopedia of Meat Sciences (Dikeman and Devine, 2014), designed by an international team of experts, covers various important aspects such as animal husbandry, physiology, slaughter, meat preparation, packaging, welfare, food safety and many others. In France, Georges Chaudieu (Chaudieu, 1970) and Eric Glatre (Glatre, 2008) developed the first meat dictionaries and then the French Meat academy published the Meat dictionary in 2012. The latter contains clear and concise definitions of the names of cuts, professional butchery terms (including utensils), breeds of livestock, as well as technical or peripheral terms, applicable to livestock: cattle, veal, lamb, pork, poultry and game. It also lists the terms relating to the taste and texture of meat, as well as an inventory of some French and international dishes traditionally made with meat. Finally, it contains terms specific to the meat trade as well as French specificities of meat such as cuts and organisations of the sector. It contains 1357 entries, it is available as a book in French and English and also includes some terms in Spanish. Its most recent version was edited in 2019. Although they are well-known references, these dictionaries have been published in paper form and are therefore limited to human reading. The widespread use of digital approaches in research as well as in industry is creating new needs that require the use of semantic resources at the heart of information systems.

In the current context of Open Science development, with its challenges related to knowledge transfer and innovation, it is more necessary than ever to have terminological repositories that facilitate:

1) collaboration between disciplinary and linguistic communities,
2) the discovery of information and data from various sources,
3) the development of methods and tools for decision support based on the integration of heterogeneous data and domain knowledge. To achieve objectives 2) and 3), these repositories must be available in formats that can be used by a software.

The research communities in agriculture, food and environment are particularly active in the creation and provision of these semantic resources as evidenced by the study conducted by the Agrisemantics working group in the framework of the "Research Data Alliance" (Aubin et al., 2017). These scientific communities are now also part of the FAIR approach; a set of principles broken down into actions that make data easier to find, more accessible, more interoperable and therefore more easily re-used (Wilkinson et al., 2016). One of the FAIR principles puts vocabularies as a cornerstone centre of the approach: "*I2: Data and metadata use vocabularies that respect the FAIR principles*", I2 being the second principle supporting data interoperability, that is, the ability for a software to exchange and process data with heterogeneous representations as they originate from several information systems. A number of vocabularies for agriculture and food are thus accessible from public portals, including Agroportal (Jonquet et al., 2018) where the Meat thesaurus is exposed. This type of portal facilitates the discovery of existing ontologies and thesauri, the reuse of which helps to harmonise the definitions given to the objects handled by each community.

Among resources publicly available for the agri-food domain, we found the physical and chemical characteristics of beef and pork muscles (https://bovine.unl.edu/ and https://porcine.unl.edu/) and generalist thesauri such as the FAO's Agrovoc (http://www.fao.org/agrovoc/fr) or the National Agricultural Library Thesaurus (https://agclass.nal.usda.gov/). Another interesting resource is the Langual thesaurus (https://www.langual.org), which provides a standardised language to describe food, in particular to classify food products for information retrieval. The FoodOn ontology (Dooley et al., 2018), derived from Langual provides a more formal representation of Langual, which allows reasoning to be constructed. During our research, we were unable to identify a semantic resource that was both precise enough and covering completely our specific needs and that was easily reusable for the meat sector. On the field of meat, let us mention the work of (Pizzuti et al., 2017a, 2014) which focused on the construction of ontologies within the framework of the FTTO (Food Track and Trace Ontology) project to represent traceability in the meat industry without, however, making the result accessible and reusable.

In line with the “Open Science” approach, our aim is to convert the Dictionary of the French Meat Academy, which provides us with terms and precise definitions specific to the meat sector in French and English, into a thesaurus to make these terms and concepts freely available to the professional and scientific meat community. A thesaurus is an “organised authority list of descriptors and non-descriptors that obey their own terminological rules and are linked together by semantic relationships (hierarchical, associative, or equivalence). This list is used to translate concepts expressed in natural language into an unambiguous artificial language” (AFNOR 1981). The thesaurus is a resource accessible to human readers, with a logical structure and textual definitions. By choosing to represent it using the standard SKOS (Simple Knowledge Organization System) schema, and to associate rich metadata and identifiers with it, we also make it usable by various computer applications carrying out the indexing of documents, datasets, content annotation, or the display of definitions in context.

In this article, we present our approach to the construction and online publication of the Meat Thesaurus through collaborative work between specialists in knowledge and data engineering on one hand and meat science and technology on the other hand.

2. Material and methods

The transformation of the dictionary from its documentary form to a structured thesaurus was carried out in three phases:

1) analysis of the data leading to the choice of the standard for representing the information contained in the thesaurus, that is, the SKOS model,

2) transformation of the data, carried out semi-automatically, and

3) intellectual and manual work on organising the concepts within the hierarchies

2.1. Choice of the SKOS model

The creation of the Meat Thesaurus is based on the terms of the Meat Dictionary by the French Meat Academy. The source data file provided was made of two Word documents, including one for French, with 1357 entries and inserts. The inserts were not kept in the resulting thesaurus. The content for English was integrated in a second phase of the project, after the transformation and organisation of

the data for French, as the English version was not available at the beginning. Similar though simpler procedures were used for English. In this paper, we will focus on the processing for data in French.

The structure for describing an entry in the source file for French is pretty regular (Figure 1), which allows some automation of the process, although this is made difficult by the presence of a large number of optional elements. On the first line, the term is followed by its grammatical characteristics (its nature, gender and number) and a definition as in the example shown in Figure 2. Some entries have one or more synonyms, notes, or a reference to one or more other entries indicated by the mention "**Voir**" (See). Except from the grammatical information, term properties are prefixes with the "I" (pipe) character. Synonymy is indicated heterogeneously either using the fields "Synonym" or "See" or as part of the definition or the "To got further" information.

term (required) nature.gender.number (optional) **|** definition (optional)
| Pour aller plus loin : (optional)
| Synonyme : (optional)
| Notes : (optional)
| *Voir* *: name (optional, several possible values separated by commas or full stops)*

Figure 1: Structure of a dictionary entry (Pour aller plus loin = To go further =; Synonyme= Synonym; Notes = Notes; Voir = See)

emboucheur n.m. **|** Éleveur ou marchand qui pratique l'engraissement du bétail à l'herbe sur des pâtures.
| Synonyme : herbager.
| Voir : Embouche.

Figure 2: Example of a dictionary entry from the file source in French : emboucheur (grass finisher)

Considering the English version of the dictionary, the useful information to retrieve are the English term, its correspondancy in French, its definition and reference to other entries (« See » information) as shown in Figure 3.

Grass finisher/ Emboucheur
n.
Livestock producer or trader who fattens livestock on grass in pastures.
See: Grass Finishing.

Figure 3: Example of a dictionary entry from the file source in English : grass finisher. The equivalent term in French is indicated in red and preceded by the / character

Considering the source data structure and the community practices, we chose the SKOS (Simple Knowledge Organisation System) model, a W3C standard (https://www.w3.org/) used to represent simple knowledge organisation systems, typically thesauri. We will see that choosing a concept-centered representation of the data implies some difficulties when working with a term-centered

source (dictionary). The SKOS model is based on RDF (Resource Description Framework), a framework for describing structured data as a graph made of assertions of the form [Subject - Predicate - Object] as shown in Figure 4.

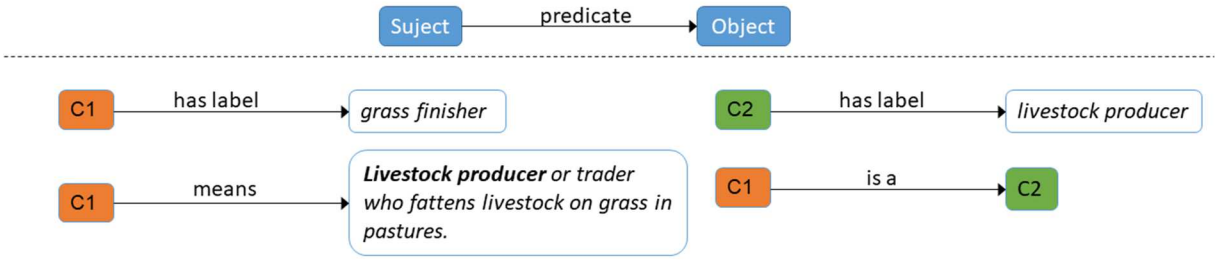

Figure 4: Examples of information representation in the form [Subject - Predicate - Object] or RDF triplets

A unique identifier called a URI is used to reference each concept. The URI is a global identifier, that is, unique throughout the web space. When published, requesting the URI will allow to display or retrieve information on a concept. This identifier is composed of a unique prefix for the thesaurus (here *http://opendata.inrae.fr/ThViande/*) and a unique suffix for each resource described (for example *C1* for concept 1).

A partial representation of the SKOS model is shown in Figure 5. The central object of the model is the concept (*skos:Concept*) which can be seen as a unit of thought corresponding to a type of real-world object (e.g. the notion of “herd”) or to an abstract entity (e.g. the notion of “meat quality”). It is designated by a preferred term (*skos:prefLabel*) in one or more languages and possibly alternative terms (*skos:altLabel*). The *skos:definition* property is used to add a textual definition to the concept. Unspecified relationships between two concepts can be encoded in the *skos:related* relation, which is equivalent to “see also” like between the concepts for “grass finisher” and “grass finishing” in our example. The whole thesaurus is a *skos:ConceptScheme*, where concepts are organised in a hierarchy using the relationships *skos:narrower* and *skos:broader*. For example, [livestock producer] *skos:narrower* [grass finisher] means that “grass finisher” is a specific concept of “livestock producer”. This hierarchical structuring did not exist in the dictionary and was built with the experts as discussed in section 3.

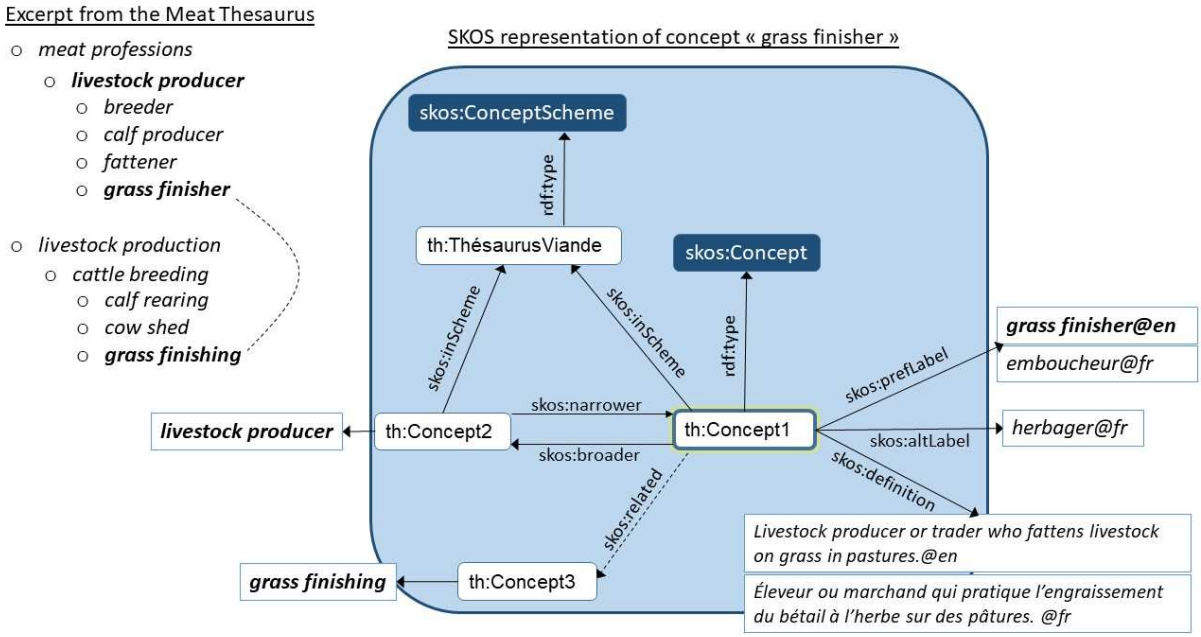

Figure 5: Thesaurus data model based on SKOS

2.2. Conversion of the dictionary from its textual form into a structured and formalised resource

The work for the conversion of the dictionary into the thesaurus includes the recovery and formalisation of data (blue elements in Figure 6) which precedes the structuring phase (grey).

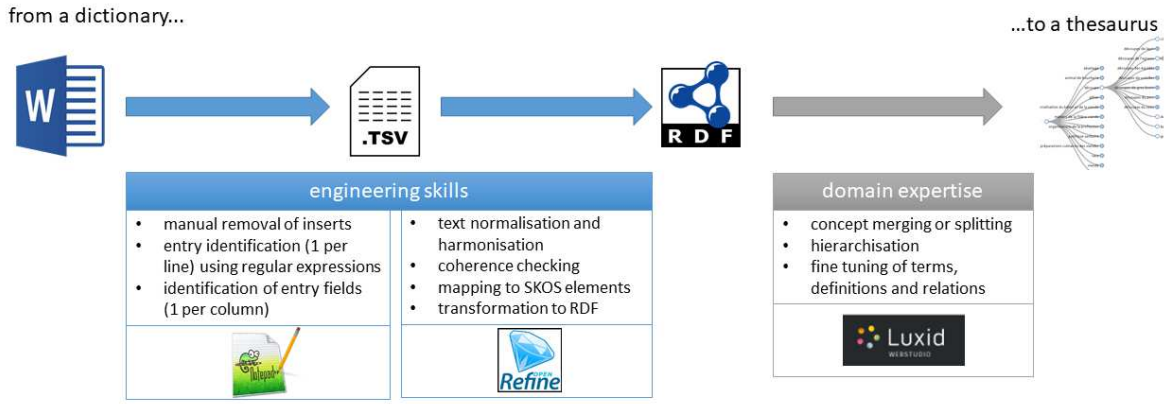

Figure 6: Thesaurus construction process

Much of the retrieval and transformation work was done semi-automatically and required the following:

- An expertise in using regular expressions that we applied in a simple text editor (Notepad++) in order to isolate the dictionary entries from each other (1 per line) and to organise their respective information in a tabular format;

- OpenRefine, a free software for data cleaning and transformation. It allowed us to identify and massively correct some encoding errors, and to refine the information segmentation of each entry. Finally, the RDF extension of OpenRefine was used to transform the data from a tabulated format to the standardised SKOS representation in RDF/XML. Openrefine allows the definition of a mapping procedure from table cells to RDF predicates.

The data was finally uploaded into *Webstudio*, a vocabulary editor providing an ergonomic interface for experts to curate and organise the concepts into a hierarchy.

2.3. Concept curation and organisation

This part of the work was entirely manual and consisted in moving from a term-centered (dictionary) to a concept-centred (thesaurus) representation, and in proposing a thematic organisation. This task required in-depth knowledge of the field, which was provided by four experts, two from the French

Meat Academy and two from INRAE (French National Research Institute for Agriculture, Food and Environment). The experts from the French Meat Academy provided knowledge on cuts, cooking, trades and specificities of the sector as well as on the different breeds of meat animals. The INRAE experts provided knowledge on muscle biology. About 32 meetings of 2 hours each were held for this work. In some cases, the experts relied on other knowledge sources such as the AHOL ontology (Salaun et al., 2020) or the Animal Science Dictionary by CIRAD (Meyer et al., 2022).

When transforming the Meat dictionary into the Meat thesaurus, a major challenge was to either properly group terms into classes of synonyms, i.e. concepts, or on the opposite, split dictionary entries into unequivocal notions. Grouping terms into concepts was facilitated by either the "Voir" (See) information on entries with no definition or explicit mentions of synonymy in the definition as shown in Figure 7. In some cases, like "cabri", experts had to select relevant parts of both definitions and merge into a single concept, e.g. "chèvre".

Figure 7: Two examples of means to indicate synonymy in the Meat dictionary

Conversely, and in order to meet the meaning univocity constraint on SKOS concepts[1], some dictionary entries had to be split into several concepts and their definition modified accordingly. This is the case for "protein", the definition of which contained terms and definitions for various types of proteins. Concepts were created for each type of protein with its own definition.

Other concepts have been introduced to meet the needs of a meaningful hierarchy and to ensure a certain consistency within the thesaurus. In particular, the first level concepts (called "top concept" in SKOS) had to be generic enough to constitute thematic categories such as "breed" or "meat professions". The ontologies AHOL - Animal Health Ontology for Livestock (https://www.atol-ontology.com/en/a-ahol/) and ATOL - Animal Trait Ontology for Livestock (Le Bail et al., 2014) provided some concepts initially absent from the meat dictionary.

For this first version of the thesaurus, we chose not to introduce poly-hierarchy, an organisation in which a concept can have several generics. While this choice makes the classification work more difficult (because a non-obvious decision has to be made in case of difficult choice), it also makes it

[1] Know more about SKOS constraints : https://www.w3.org/2004/02/skos/core/guide/2004-11-25.html#secappii

more precise (because the best location has to be found). In order to organise the concepts in a hierarchical way, the experts relied on the generic terms already defined or mobilised their own expertise. In most of the cases where a concept could be located at several places in the hierarchy, the more generic solution was chosen. For instance, the branch “butchery animal” was chosen to encompass all animals slaughtered for their meat. A small number of cases were treated differently. For instance, “quarter” could be considered as a specific of “cutting” or of “half-carcass”. In this case, it was decided that it was more relevant to insert this concept under “half-carcass” in order to make its semantics more explicit.

This work was completed by the revision of some definitions more adapted in the context of the thesaurus, or that did not comply with definition writing good practices (Vézina, R. (2009)). Following the hierarchisation, some *skos:related* (“See”) relations present in the paper dictionary were removed in order to comply with the SKOS model that does not allow associative relations between concepts already in a hierarchical relation, i.e. belonging to the same branch of the thesaurus.

During this expertise work, some concepts were finally discarded from the thesaurus, like “VOF” (“Viande Ovine Française”), “VBF” (“Viande Bovine Française”) and “VPF” (“Viande Porcine Française”), keeping only official labels.

3. Results

The result of this work is a bilingual thesaurus with terms and definitions in French and English, which is made publicly available for reuse in the following link: http://agroportal.lirmm.fr/ontologies/MEAT-T. The Meat Thesaurus consists of 1519 concepts organised into 12 branches (or 1st level concepts) which are:

- breed
- butchery animal
- culinary preparation of meats
- cutting
- game
- health policy
- livestock production
- meat
- meat market
- meat professions
- professional organizations
- slaughtering

Each branch is divided in more specific sub-branches as shown in Figure 7, which allows navigating from generic to specific concepts.

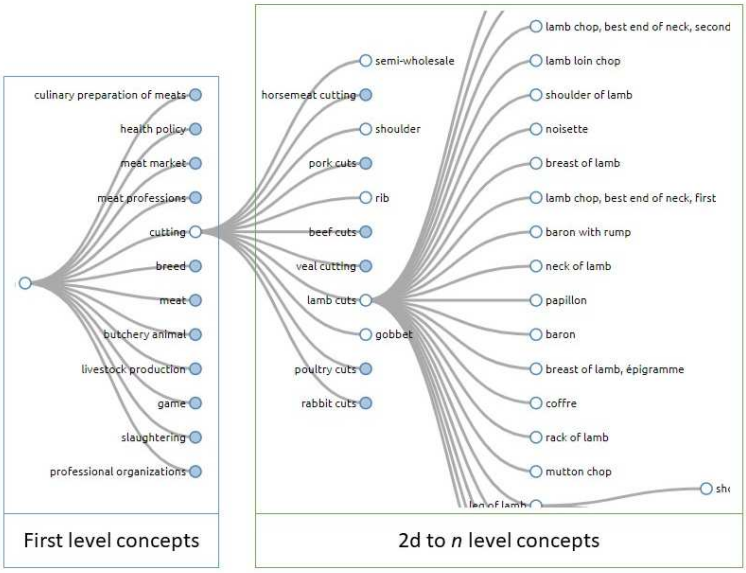

Figure 8: Declination of the "cutting" concept

3.1. A thesaurus covering the main areas of the meat industry

We present the first-level concepts in alphabetical order:

- "breed" includes the notions of hardy breed, standards as well as the different breeds of cattle, goats, horses, sheep and pigs.

- "butchery animals" includes animals that are slaughtered for meat consumption. The animals are divided into their major families which are cattle, goats, asses, horses, sheep, pigs and poultry. "other animals for slaughter" lists animals whose meat is edible but not commonly consumed, such as ostriches, bison and peacocks;

- "culinary preparation of meat" includes concepts that describe meat-based preparations, preparations in which meat is an essential ingredient or some sauces. The second level concepts are: tableware, charcuterie, salting, cooking, sauce, charcuterie-salting and culinary preparations based on beef, game, pork, tripe products, veal, ovine meat, poultry, multi-meat.

- "cutting" describes all cuts for poultry, large cattle, lamb, equines, pork, rabbit and veal, in the French context. It also defines in a general way the cuts common to several animals such as the rib, the half-calf, the shoulder and the goblet. The rationale for this choice was to consider cuts at the butcher end

rather than muscle names as considered by scientists but the linkage between the two approaches remains to be developed.

- “game” includes all animals that are hunted for their meat and are not animals for slaughter, such as pheasants, kangaroos, hares, wild ducks and wood pigeons;

- “health policy” covers 7 second-level concepts including HACCP, diseases, veterinary inspection and traceability, which are themselves broken down into more specific concepts.

- “livestock production” includes both generic livestock farming practices (e.g. “zootechnics” or “animal welfare”) and practices specific to a type of livestock farming, for example “barn” (in “cattle breeding”) or “shepherd” (in “sheep-goat breeding”);

- “meat”; this is the core subject in this thesaurus. This branch covers the different types of meat such as beef, game, poultry, sheep and pigs. It also provides definitions of white and red meat and other notions such as packaging, preservation, processing of muscle into meat, meat industry and meat quality (which includes safety, sensory analysis, nutritional value, and official quality signs as well as beef grading at the consumer end).

- “meat market” describes the commercial exchanges within the meat sector, considering the French particularities. It contains 3 second-level concepts: wholesale meat market, livestock market and transport.

- “meat professions” includes all types of occupations in the sector as well as training and distinctions.

- “professional organizations” presents the different organisations divided into four categories: commercialisation-industry, distribution, interprofession, production.

- “slaughtering” contains the actions carried out in the slaughterhouse (stunning, killing, throat cutting, dehairing, evisceration, etc.), the tools used (plucking, hide, trocar, splitting, etc.), ritual slaughter (halal and kosher), concepts relating to the carcass (carcass yield, meat yield, atlas, fifth quarter, trapping, etc.) and carcass grading;

The Meat Thesaurus is therefore a resource that covers a wide range of concepts in the meat sector. The thematic organisation offers a navigation path that facilitates its discovery and use. The integration of definitions from the French Meat academy dictionary makes it a reliable source of knowledge that is representative of French practices. This can be extended to other global practices to provide connection and accurate relationships across global practices.

3.2. A FAIR resource

In order to allow its reuse by third parties, the Meat Thesaurus is made freely available to the community on a public portal, Agroportal, and in an open format. It is part of the FAIR approach where a set of principles provide a framework to ensure that a digital resource can be reused by third parties (humans and machines). Solutions proposed by the DipSO INRAE Open Vocabularies service have been combined to ensure that the meat thesaurus complies as far as possible with FAIR principles.

The analysis proposed in Table 1 was carried out in the framework of the ANR FooSIN project (https://foosin.fr/) and the questionnaire developed by the RDA SHARC working group (David et al., 2020).

FAIR Objectives	Solutions to meet the FAIR principles
Findable	• The thesaurus is uniquely and globally identified by a DOI (Digital Object Identifier) which links it to its metadata accessible on Data INRAE (https://data.inrae.fr/). The DOI also provides a mean to precisely cite the thesaurus; • The long-term preservation of DOI and URI identifiers are guaranteed by Data Cite and INRAE respectively; • The thesaurus is referenced in Data INRAE and displayed on Agroportal (http://agroportal.lirmm.fr/), its metadata and content are indexed by Internet search engines;
Accessible	• The thesaurus is downloadable from Agroportal in various open formats (SKOS/XML, RDF/XML and CSV); • The description and content of the thesaurus are accessible via web services or API (http://data.agroportal.lirmm.fr/); • The description of a concept is accessible via its URI, which is dereferenced (dereferencing service provided by the DipSO INRAE); • Accessibility is based on an institutional service operated by the DipSO.
Interoperable	• The content of the thesaurus is represented using the W3C SKOS standard, in RDF. A CSV version is also available; • Its metadata is represented using elements from standard schemas that can be interpreted by many information systems and search engines and are widely adopted by the community;
Reusable	• The thesaurus is displayable and usable using free and open source tools such as VocBench or Skosmos; • The conformity of the thesaurus with the SKOS model has been tested with the Skos Play tester (Sparna, https://skos-play.sparna.fr/skos-testing-tool/) which is based on the qSKOS rules (https://github.com/cmader/qSKOS/wiki/Quality-Issues); • The editorial choices and intentions of the authors are explained in this article

Table 1 : Description of the FAIR objectives

Since this evaluation, the FooSIN project has contributed to a new service on Agroportal which computes and displays the FAIRness score of each resource it hosts. The methodology and criteria are explained by Amdouni and Jonquet (2021). The service provides score explanation and tips to improve the findability, accessibility, interoperability and reusability, which will guide us in enhancing the content, documentation and metadata of the Meat Thesaurus.

4. Discussion

Having performant information retrieval systems is important in any field including in the meat sector. A thesaurus provides a precise and controlled vocabulary which serves coordinating indexing and information retrieval (Clarke, 2017). Unfortunately, many terminological resources in the meat sector are either in paper format, in PDF or in tabulated format and, consequently, are not suitable. For example, the UNECE (United Nations Economic Commission for Europe) provides standardised descriptions of cuts for different butchery animals for the European Union (www.unece.org/trade/wp7/Meat-Standards) which is potential useful for meat trade but it is only available in PDF format. In addition, the University of Nebraska-Lincoln provides the physical and chemical characteristics of beef and pork muscles (https://bovine.unl.edu/ ; https://porcine.unl.edu/). Nevertheless, this resource is only limited to cuts and not the meat sector as a whole. A digital tool with the possibility of doing linkages will therefore be of great benefit to integrate trade-based or cut-based standards with more specific scientific and more generalised industry terms.

The meat thesaurus is presented as a useful resource respecting the FAIR principles of open science for indexing specific databases in the meat sector including contextual and national characteristics and for information retrieval from these databases. This will also help to provide a semantic base for the construction of a meat quality ontology that will be further discussed.

4.1. Indexing databases

The meat thesaurus can be classically used in bibliographic databases to index journal articles, books or any document on meat, its processing, trade and related disciplines.

Technological advances over the last decades have dramatically increased the rate of collecting scientific information and data (Hughes et al., 2008). This has led to a proliferation of large biological databases that manipulate concepts like units, scales, or laboratory methods (Chriki et al., 2013) that are defined in different ways, or not defined at all. This is a huge limitation for meta-analysis and modelling approaches (Hocquette et al., 2012). It is therefore advantageous for researchers to use already defined concepts and add keywords to the metadata when feeding databases. This will also help in the reusability and interoperability of research databases. According to ISO 25964-I (International Organisation for Standardization 2011, Clause 4.1), a thesaurus has as prime function to support information retrieval by guiding the choice of terms for data indexing and information searching. This will enable an indexer and a searcher to choose the same term for the same concept. A thesaurus gives a unified and formalized representation of information in the information retrieval system, reflecting paradigmatic relations between terms. It is an effective tool for thematic retrieval as it provides search precision on specific subjects.

As an example, in 2002, the National Agricultural Library Thesaurus (NALT) was established to select controlled vocabulary terms for subjects indexing in databases such as AGRICOLA or PubAg (https://agclass.nal.usda.gov/). With the emergence of repositories and catalogues for other research products like data, software, and protocols, it has become crucial to use reference concepts to be able to query them in an efficient manner. Yet, thematic wide vocabularies like Agrovoc or NALT are generally not suited to precisely describe data, which often requires finer grained or specific vocabularies for a given scientific field.

Some experts who participated in this work are members of the International Meat Research 3G Foundation (https://imr3g.org/). The main aim of this foundation is to strengthen the link between the different actors in the beef sector (farmers, producers, slaughterhouses, wholesalers and retailers). To achieve this, the foundation is establishing an international database with a large number of consumers' scores related to beef palatability and related animal, carcass, cut and muscle data. These scores are obtained according to standard protocols of the most advanced beef grading system: The Meat Standards Australia methodology (MSA). This system has been undergoing development since the 1990s, always with the same protocols to record the most powerful determinants of beef eating quality on a large scale first in Australia and then across countries such as France, Ireland, Poland, South-Africa, South Korea or USA (Hocquette et al., 2020, Bonny et al., 2018). Using the MSA standard methodology already contributes to the strategy of the International Meat Research 3G Foundation but is not sufficient to achieve data interoperability and reusability at scale. These international collaborations emphasised the desirability of establishing common standards, description and protocols for data collection and physical collection of cuts, fabrication of consumer samples and

untrained consumer testing protocols and control software to ensure data were fully compatible and able to be aggregated. As a result, agreed descriptive terms and protocols reflecting global data have been endorsed for use in conjunction with the UNECE Bovine Language allowing free universal use in conjunction with IMR3GF accredited assessment training and providing the ability to develop 3G consumer prediction models from appropriate pooled data. To go in that direction, the Meat Research 3G Foundation will ensure that the future database will allow its users to work using a shared language. The main objective of publishing the meat thesaurus was to provide a semantic base specific to meat quality that would sustain the future database.

Beyond its use in this database, the meat thesaurus is made available and thus expected to be adopted by the meat supply chain actors who need accurate definitions of specific concepts. Thanks to the access services provided by Agroportal, it can be incorporated directly in digital applications used by agricultural and consumer devices connected to the Internet. Such applications could then benefit from a standard food vocabulary with a global scope also including regional specificities, French ones in our case. Parts of the meat thesaurus can also be integrated into or mapped with other semantic resources. For instance, the Chinese Agricultural Thesaurus was mapped to AGROVOC to allow users to access the vast repositories of Chinese agricultural knowledge, previously inaccessible to non-Chinese speakers, by allowing them to use vocabulary in languages with which they are familiar as an entry point to the indexed resources (Liang and Sini, 2006).

To summarise, the meat thesaurus described here, freely accessible and following the open science principles is likely to be a useful resource for indexing specific databases in the meat sector. This thesaurus has the potential to include regional characteristics such as those of the French meat production up to now. On the other hand, its genericity and its detailed description of the meat sector make it potentially useful at the international level, especially in the area of meat quality.

4.2. Using the thesaurus to build an ontology

Scientific research on meat eating quality has to answer complex problems, which can require working on heterogeneous data sometimes coming from various sources. One of the main obstacles to data integration is the variability with which producers name the objects studied, variables, units of measurement, etc. Integrating domain knowledge encoded in an ontology to data processes can be a solution to facilitate data integration and reason on these data. An ontology provides formal descriptions of domain knowledge and allows reasoning on data while a thesaurus focuses on how to name things and organises them thematically.

The authors are interested in developing an ontology dedicated to predict meat quality. Meat quality can be defined as a set of properties that together identify what we appreciate about meat when we purchase it, eat it, or select it for use as a raw material for processing into meat products (Purslow, 2017). Traditionally, meat quality traits are grouped in two sets: intrinsic and extrinsic traits. The intrinsic traits are those associated with the product itself which include appearance, colour, water-holding capacity, and odour, nutritional value (fatty acid composition, iron and vitamin contents, etc), sensory perception (tenderness, flavour, juiciness). Other traditional quality factors normally expressed as freshness or wholesomeness, or safety in terms of lack of pathogens, parasites, infections agents, or toxins are also major intrinsic quality traits. The extrinsic quality traits are related to the way meat is produced and therefore include issues related to the well-being of meat animals and the sustainability of production systems in terms of environmental and economic performances. Starting from the Meat Thesaurus will facilitate the construction of the ontology as it largely covers the notions that are directly or indirectly associated to meat quality. These notions can be found within the meat thesaurus under the following main concepts:

- nutritional values of meat	- health policy
- sensory analysis	- veterinary inspection
- conservation	- livestock production

Ontological classes can be derived from these concepts by adding formal properties and relation as well as logical rules. In order to predict quality, i.e. to infer the level of quality as “unsatisfactory”, “Good everyday quality”, “better than everyday quality” and “premium quality”, the ontology will have to include formal definitions and specific rules, i.e. classes, for the animal type, the name of the cuts and the cooking method. For example, a butcher may want to predict the level of quality of a beef cut, or individual muscle portions, e.g. the sirloin from a famous beef breed. He may also want to know what type of cooking method will be the best to maximise eating quality in order to provide the best advice to consumers. Concepts of the ontology like the breed of the animal, the results of veterinary inspection obtained from the farmer, the meat colour and marbling assessment (the ontology can give precision on the assessment method) can be used. Finally using this information, the ontology can provide recommendations for the cooking method and help to provide a potential level of quality.

Deriving ontologies from thesauri has been proven. Indeed, previous examples in which thesauri have helped in the establishment of ontologies do exist. For example, the FoodOn ontology was constructed using, among others, the Langual thesaurus which was published in 2017 (Møller and Ireland, 2018). This work was carried out by a consortium to build a comprehensive open source ontology, consisting

of hierarchies of terms that cover raw foods, process conditions for packaging, cooking and preservation, and a variety of product type schemes under which food products can be classified. The description of each food is based on descriptive qualities, its components as well as the associated processes. Other thesauri, including AGROVOC and the Aquatic Sciences and Fisheries Abstracts Thesaurus (ASFA) have also been converted into ontologies in the same way, in order to improve their expressiveness and take advantage of the tools made available by the semantic web community (Lauser et al., 2008).

The future ontology will not only benefit from the Meat thesaurus but also from semantic resources including ontologies and thesauri developed by other communities and shared in repositories like Agroportal that offer facilities for reusing them. For example, when we search for the definition of marbling (a key quality trait of beef) in AgroPortal, we have access to several definitions considering different specificities of different regions of the World as shown in Table 2 For each class of the future ontology, this feature will allow to pick-up the definition and terms most suited to our needs. For instance, for the measurement of marbling, the definitions from ICAR or USDA differ and might not be relevant to some users like French users for example. The thesaurus will therefore enable users to access locally relevant definitions and provide clarity on how definitions compare across systems.

ATOL	Any measurable or observable characteristic of the degree of infiltration of intramuscular fat
ICAR **ICAR Recording Guidelines May 2014**	Marbling can be defined as the flecks of fat in the lean. Marbling is usually evaluated visually in the rib-eye muscle, which is exposed between the 12th and 13th ribs.
	Intramuscular fat (marbling) is the intermingling or dispersion of fat within the lean
MSA **www.mla.com.au**	It is the last fat to be deposited and the first to be utilised by the animal as an energy source ; marbling is the fat that is deposited between muscle fibres.
	Marbling is assessed from the 5th to 13th rib on the carcase and seen as intramuscular deposits of fat within the muscle.
	The term marbling refers to the small flecks of fat scattered throughout the muscle
USDA **www.meat.tamu.edu**	Marbling (intramuscular fat) is the intermingling or dispersion of fat within the lean. Graders evaluate the amount and distribution of marbling in the ribeye muscle at the cut surface after the carcass has been ribbed between the 12th and 13th ribs.
	The types and amounts of fat in the muscle the intermingling of fat deposits in muscle.
JMGA (beef marbling standards) **www.jmga.or.jp**	The amount and distribution of intramuscular fat.
	Marbling is flecks or thin strips of fat in beef
CBGA **www.beefgradingagency.ca**	Marbling is the fine white flecks of fat running through the lean meat. The amount of marbling in the rib eye defines the minimum standard for the top-quality grades
	Amount, size and distribution of intramuscular fat deposits in the Longissimus muscles of a livestock carcass that have been exposed by knife-ribbing
Sistema Brasileiro de Classificação de Carcaças de bovinos **www.uel.br**	This assessment of the quantity and of the intramuscular fat distribution is made visually in the muscle Longissimus dorsi, between the 12th and 13th ribs. For the classification of marbling, photographic standards produced by the USDA

Table 2 : Results of the query with the word "marbling" on AgroPortal

To summarise, addressing complex biological problems in meat science may be helped by using ontologies, which provide formal descriptions of domain knowledge and allows reasoning on data. The Meat Thesaurus provides names and definitions for concepts within the meat sector and organises them thematically. Among other uses, it is likely to help building a powerful ontology for the meat sector in combination with other existing resources in the same field.

5. Conclusion

Terminological, syntactic and semantic discrepancies between data sources need to be overcome so that it will be possible for researchers to have transparent access to different disparate data sources, and possibly integrate them to produce new knowledge. There is thus a growing awareness of the need for ontologies in life science. There is also a need to share a wealth of knowledge among different researchers and research or professional groups. As a result, a growing number of biological ontologies are being built in life science as information management is of paramount importance. Unfortunately, very few resources could be associated with this initiative in the field of meat. We propose a new resource as a thesaurus specific to the meat sector to fill this gap. As shown in this article, transforming textual content to a structured, machine actionable and organised resource requires time and specific competencies. To avoid data loss and expensive post-processing, it is important to adopt a FAIR-by design approach when developing dictionaries or any other terminological or semantic resource. This means to, right from the beginning of a project, adopt tools allowing data structuring, even a simple spreadsheet, and whenever possible including features for standard representation, modification tracking, and quality check (section 5 of Aubin et al. 2017 presents some of the most used tools in the agricultural community working with semantic resources). Making semantic resources easily findable and accessible is also a key for reducing duplicated efforts and costs, enhance reusability and thus allow better semantic interoperability of data in the domain.

The meat thesaurus described here represents all aspects of the meat supply chain from production to consumption and this has many advantages for all those wishing to use it, in particular professionals, consumers and researchers. It not only provides precise definitions of the different concepts specific of the meat sector, but also highlights relationships between them through categories and hierarchical and explicit associated links. This will contribute to a better knowledge of the sector, its organisations, its products, its policies, etc. The advantage of this thesaurus is that it can also be used to index databases and bibliographic resources.

This thesaurus covers various aspects of meat: meat-producing animals, slaughter and cutting, marketing, culinary preparations and consumption. The users of this thesaurus may therefore have different profiles: professionals of the sector, researchers, journalists, students or simple consumers who wish to better know the specificities of the meat sector, in particular in the French context. This thesaurus does exist in English while retaining the specificities of the French meat industry and having the potential to incorporate other regional specificities. In addition, the thesaurus will further be expanded into different languages. It has also the potential to be regularly updated. Indeed, it could happen that the thesaurus does not cover all the requirements of any database which may contain additional variables not defined yet or defined in a different way. In this case, it will be possible to request additions or modifications of concepts in the thesaurus.

In combination with other existing and freely available resources in the meat sector, this thesaurus will also serve as the basis for an ontology through specification of different classes that will be more formal and more precise. The meat ontology will make it possible to deepen a particular aspect of meat, such as its quality and will also enable reasoning for computers for a more automatic use. This is in line with the strategy of international initiatives such as the aims of the International Meat Research 3G Foundation. This is also in agreement with the open science policy promoted for instance in EU-funded research projects.

Funding

This project was partly funded by the French Meat Academy on one hand and the ANR project FoodSIN (number ANR-19-DATA-0019-01) on the other hand.

Meat Thesaurus Access: http://agroportal.lirmm.fr/ontologies/MEAT-T

CRediT authorship contribution statement

Moise KOMBOLO: Data curation, Formal analysis, Investigation, Methodology, Software, Visualization, Writing – original draft. Jérémy YON: Data curation, Formal analysis, Investigation, Methodology, Software, Writing – review & editing. François LANDRIEU and Brigitte RICHON: Data curation, Formal analysis, Investigation, Writing – review & editing. Sophie AUBIN and Jean-François HOCQUETTE: Conceptualization, Funding acquisition, Methodology, Project administration, Resources, Supervision, Validation, Writing – review & editing.

Declaration of Competing Interest

None.

Acknowledgments

The authors thank all members of the French Meat Academy for their work which provided definitions for all concepts in the original Meat dictionary.

References

Ahol – Livestock Ontologies, https://www.atol-ontology.com/en/a-ahol/ (accessed 1.27.21).

Aitchison, J., Gilchrist, A., & Bawden, D. (2004). Thesaurus construction and use a practical manual. Europa Publications, New York; London.

Amdouni, E., & Jonquet, C. (2021). FAIR or FAIRer? An integrated quantitative FAIRness assessment grid for semantic resources and ontologies. MTSR 2021 - 15th International Conference on Metadata and Semantics Research, Nov 2021, Madrid, Spain. In press, ⟨10.13140/RG.2.2.23265.84321⟩. ⟨lirmm-03208544v3⟩

Aubin, S., Caracciolo, C., & Zervas, P. (2017). Landscaping the Use of Semantics to Enhance the Interoperability of Agricultural Data. RDA Agrisemantics Working Group. https://doi.org/10.5281/zenodo.3697332

Baker, T., Whitehead, B., Musker, R., & Keizer, J. (2019). Global agricultural concept space: lightweight semantics for pragmatic interoperability. *npj Science of Food, 3,* 16. https://doi.org/10.1038/s41538-019-0048-6

Bonny, S.P.F., Hocquette, J.-F., Pethick, D.W., Legrand, I., Wierzbicki, J., Allen, P., Farmer, L.J., Polkinghorne, R.J., & Gardner, G.E. (2018). Review: The variability of the eating quality of beef can be reduced by predicting consumer satisfaction. *Animal, 12,* 2434–2442. https://doi.org/10.1017/S1751731118000605

Chaudieu G. (1970). Le petit dictionnaire de boucherie et de boucherie-charcuterie. Ed. J. Peyronnet, 1970. 193 p

Chriki, S., Picard, B., Faulconnier, Y., Micol, D., Brun, J.-P., Reichstadt, M., Jurie, C., Durand, D., Renand, G., Journaux, L., & Hocquette, J.-F. (2013). A Data Warehouse of Muscle Characteristics and Beef Quality in France and A Demonstration of Potential Applications*. Italian Journal of Animal Science, 12*, e41. https://doi.org/10.4081/ijas.2013.e41

Chriki, S., Picard, B., Jurie, C., Reichstadt, M., Micol, D., Brun, J.-P., Journaux, L., & Hocquette, J.-F. (2012). Meta-analysis of the comparison of the metabolic and contractile characteristics of two bovine muscles: Longissimus thoracis and semitendinosus. *Meat Science, 91*, 423–429. https://doi.org/10.1016/j.meatsci.2012.02.026

555 Clarke, S.D. (2017). Thesaurus (for information retrieval). URL https://www.isko.org/cyclo/thesaurus
556 (accessed 2.3.22).
557 David, R., Mabile, L., Specht, A., Stryeck, S., Thomsen, M., Yahia, M., Jonquet, C., Dollé, L., Jacob, D.,
558 Bailo, D., Bravo, H., Gachet, S., Gunderman, H., Hollebecq, J.-E., Ioannidis, V., Le Bras, Y., Lerigoleur,
559 E., Cambon-Thomsen, A., & Community, S. (2020). Templates for FAIRness evaluation criteria -
560 RDA-SHARC ig. https://doi.org/10.5281/zenodo.3922069
561 Dictionnaire de la viande (2012). Académie de la viande (France). Publisher, Autres voix. ISBN,
562 2918237086, 9782918237082
563 Dikeman, M., Devine, C. (Eds.), (2014). Encyclopedia of meat sciences, 2. ed. ed. Elsevier, Amsterdam.
564 Dooley, D.M., Griffiths, E.J., Gosal, G.S., Buttigieg, P.L., Hoehndorf, R., Lange, M.C., Schriml, L.M.,
565 Brinkman, F.S.L., Hsiao, W.W.L., 2018. FoodOn: a harmonized food ontology to increase global food
566 traceability, quality control and data integration. *npj Science of Food, 2,* 23.
567 https://doi.org/10.1038/s41538-018-0032-6
568 FoodOn - Summary | LIRMM AgroPortal [WWW Document], n.d. URL
569 http://agroportal.lirmm.fr/ontologies/FOODON (accessed 6.22.21).
570 Glatre, E., 2008. Dictionnaire de la viande. BPI, Clichy.
571 Health, T.L.P. (2019). More than a diet. *Lancet Planet. Health, 3*, e48. https://doi.org/10.1016/S2542-
572 5196(19)30023-3
573 Hocquette, J.-F., & Chatellier, V. (2011). Prospects for the European beef sector over the next 30 years.
574 *Animal Frontiers, 1*, 20–28. https://doi.org/10.2527/af.2011-0014
575 https://imr3g.org/, n.d. URL https://imr3g.org/ (accessed 6.19.21).
576 Jonquet, C., Toulet, A., Arnaud, E., Aubin, S., Dzalé Yeumo, E., Emonet, V., Graybeal, J., Laporte, M.-A.,
577 Musen, M.A., Pesce, V., & Larmande, P. (2018). AgroPortal: A vocabulary and ontology repository
578 for agronomy. *Computers and Electronics in Agriculture, 144,* 126–143.
579 https://doi.org/10.1016/j.compag.2017.10.012
580 LanguaLTM - an international framework for food description - Thesaurus 2014 [WWW Document],
581 n.d. URL
582 https://www.langual.org/langual_thesaurus.asp?termid=C0120&haschildren=False&owner=C01
583 73&openstr=00000_A0361_A0356_C0116_C0164_C0127_C0173 (accessed 5.10.21).
584 Lauser, B., Johannsen, G., Caracciolo, C., Keizer, J., van Hage, W.R., & Mayr, P. (2008). Comparing
585 human and automatic thesaurus mapping approaches in the agricultural domain 10.
586 Liang, A.C., & Sini, M. (2006). Mapping AGROVOC and the Chinese Agricultural Thesaurus: Definitions,
587 tools, procedures. *New Review of Hypermedia and Multimedia, 12,* 51–62.
588 https://doi.org/10.1080/13614560600774396

maladie | Animal Health Ontology for Livestock, URL https://opendata.inra.fr/AHOL/page/AHOL_0005001 (accessed 6.20.21).

Meat | UNECE [WWW Document], n.d. URL https://unece.org/trade/wp7/Meat-Standards (accessed 2.6.22).

Meyer C., ed. sc. (2022), Dictionnaire des Sciences Animales. [On line]. Montpellier, France, Cirad. [date of access 19/01/2022]. <URL : http://dico-sciences-animales.cirad.fr/ >

Møller, A., & Ireland, J. (2018). LanguaLTM 2017 - Multilingual Thesaurus. https://doi.org/10.13140/RG.2.2.13274.64964

Nair, M.N., Canto, A.C.V.C.S., Rentfrow, G., & Suman, S.P. (2019). Muscle-specific effect of aging on beef tenderness. *LWT Food Science and Technology, 100,* 250–252. https://doi.org/10.1016/j.lwt.2018.10.038

Purslow, P. (2017). New Aspects of Meat Quality - 1st Edition. URL https://www.elsevier.com/books/new-aspects-of-meat-quality/purslow/978-0-08-100593-4 (accessed 1.27.22).

Ontology - an overview | ScienceDirect Topics, URL https://www.sciencedirect.com/topics/computer-science/ontology (accessed 11.26.20).

Page d'accueil | AGROVOC, URL http://www.fao.org/agrovoc/fr (accessed 5.26.21).

Salaun, M.-C., Yon, J., Le Bail, P.-Y., & Reichstadt, M. (2020). Animal Health Ontology for Livestock. https://doi.org/10.15454/KKZ3TS (accessed 5.26.21).

Seman, D.L., Boler, D.D., Carr, C.C., Dikeman, M.E., Owens, C.M., Keeton, J.T., Pringle, T.D., Sindelar, J.J., Woerner, D.R., de Mello, A.S., & Powell, T.H. (2018). Meat Science Lexicon. *Meat and Muscle Biology, 2,* 1. https://doi.org/10.22175/mmb2017.12.0059

Shrestha, R., Matteis, L., Skofic, M., Portugal, A., McLaren, G., Hyman, G., & Arnaud, E. (2012). Bridging the phenotypic and genetic data useful for integrated breeding through a data annotation using the Crop Ontology developed by the crop communities of practice. *Frontiers in Physiology, 3*. https://doi.org/10.3389/fphys.2012.00326

Vézina, R., Darras, X., Bédard, J., & Lapointe-Giguère, M. (2009). La rédaction de définitions terminologiques. Office québécois de la langue française, Québec.

Welcome to the Thesaurus and Glossary | NAL Agricultural Thesaurus and Glossary, URL https://agclass.nal.usda.gov/ (accessed 6.20.21a).

Welcome to the Thesaurus and Glossary | NAL Agricultural Thesaurus and Glossary, URL https://agclass.nal.usda.gov/ (accessed 6.24.21b).

Wilkinson, M.D., Dumontier, M., Aalbersberg, Ij.J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L.B., Bourne, P.E., Bouwman, J., Brookes, A.J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C.T., Finkers, R., Gonzalez-Beltran, A., Gray, A.J.G., Groth,

P., Goble, C., Grethe, J.S., Heringa, J., ’t Hoen, P.A.C., Hooft, R., Kuhn, T., Kok, R., Kok, J., Lusher, S.J., Martone, M.E., Mons, A., Packer, A.L., Persson, B., Rocca-Serra, P., Roos, M., van Schaik, R., Sansone, S.-A., Schultes, E., Sengstag, T., Slater, T., Strawn, G., Swertz, M.A., Thompson, M., van der Lei, J., van Mulligen, E., Velterop, J., Waagmeester, A., Wittenburg, P., Wolstencroft, K., Zhao, J., & Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data, 3*, 160018. https://doi.org/10.1038/sdata.2016.18

World Wide Web Consortium (W3C), https://www.w3.org/ (accessed 6.20.21).