



HAL
open science

The BMS-LM ontology for biomedical data reporting throughout the lifecycle of a research study: From data model to ontology

Amel Raboudi, Marianne Allanic, Daniel Balvay, Pierre-Yves Hervé, Thomas Viel, Thulaciga Yoganathan, Anais Certain, Jacques Hilbey, Jean Charlet, Alexandre Durupt, et al.

► To cite this version:

Amel Raboudi, Marianne Allanic, Daniel Balvay, Pierre-Yves Hervé, Thomas Viel, et al.. The BMS-LM ontology for biomedical data reporting throughout the lifecycle of a research study: From data model to ontology. *Journal of Biomedical Informatics*, 2022, pp.104007. 10.1016/j.jbi.2022.104007. hal-03560396

HAL Id: hal-03560396

<https://hal.science/hal-03560396v1>

Submitted on 1 Feb 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0 International License

The BMS-LM ontology for biomedical data reporting throughout the lifecycle of a research study: From data model to ontology

AUTHOR NAMES AND AFFILIATIONS

Amel Raboudi^{1,2,3} amel.raboudi@utc.fr

Marianne Allanic⁴ mallanic@althenas.com

Daniel Balvay² daniel.balvay@inserm.fr

Pierre-Yves Hervé¹ pyherve@fealinx.com

Thomas Viel² thomas.viel@inserm.fr

Thulaciga Yoganathan² thulaciga.yoganathan@inserm.fr

Anais Certain² anais.certain@inserm.fr

Jacques Hilbey⁵ jacques.hilbey@inserm.fr

Jean Charlet⁵ jean.charlet@sorbonne-universite.fr

Alexandre Durupt³ alexandre.durupt@utc.fr

Philippe Boutinaud¹ pboutinaud@fealinx.com

Benoît Eynard^{3,*} benoit.eynard@utc.fr

Bertrand Tavitian^{2,6,*} bertrand.tavitian@inserm.fr

* Co-last authors

¹ Fealinx, 37 rue Adam Ledoux 92400 Courbevoie, France

² Université de Paris, PARCC, INSERM, F-75006 Paris, France

³ Université de Technologie de Compiègne, Roberval, Compiègne, France

⁴ Althenas, Nantes, France

⁵ Laboratoire d'Informatique Médicale et d'Ingénierie des Connaissances en e-Santé (LIMICS)
UMRS_1142, Paris, France

⁶ APHP, Hôpital européen Georges Pompidou, Radiology department, 75015, Paris, France

Corresponding author

Amel Raboudi

Present/permanent address

Present: amel.raboudi@utc.fr

Permanent : araboudi@outlook.fr

1 INTRODUCTION

Biomedical research studies often require extensive collaborations between various institutions and disciplines, with contributions of many experts from an increasingly diverse modalities for data acquisition: e.g. the multiple “omics”, analytical and imaging techniques. The produced data need to be analyzed jointly. However, they are heterogeneous in their structure, semantics, formats (tables, images, proprietary formats) and have different levels of granularity (ranging from a checklist of physiological parameters to a thorough ECG follow up) [1]. As a consequence, biomedical researchers spend more and more time and efforts on data structuring and reporting [2]. Therefore, data are increasingly difficult to share and reuse, even worse, the reproducibility of published results is a serious issue [3], in spite of the worldwide open data initiatives.

Data reporting via rich metadata is critical to biomedical Research Data Management (RDM) and is encouraged in the open data context. Journals, research institutes, granting bodies and dataset repositories promote the respect of FAIR guidelines [4] and the use of Data Management Plans (DMP) [5]. FAIR means that scientific data, with (partial or full) public funding, must be “Findable” through its public registration with a dedicated identifier, “Accessible” to peers, “Interoperable” with other domain data and metadata and sufficiently-well annotated to be “Reusable”. DMPs ask researchers to identify relevant metadata standards before the beginning of a research study [5]. Both FAIR and DMP recommend the use of standard and published metadata sources to report and annotate data. FAIRSharing [6], a web resource of guidelines, standards, databases and tips for data management was created to this aim. However, in daily practice, most researchers are not equipped with a proper knowledge of the tools to fulfill these demands [7].

Several annotation templates and schemas, called Knowledge Organization Systems (KOSs), can be used for day-to-day data reporting with metadata. A KOS is defined as “any type of scheme for organizing information and promoting knowledge management” [8]. In the present article, all schemes of metadata are considered as KOSs. BioPortal [9] is a public repository for biomedical KOSs that provides access to published KOSs, which can be classified according to different criteria [10], e.g.

HTML glossaries have lower *semantic expressiveness* than OWL ontologies [11], and concept lists are less *structured* than data models or ontologies.

Because biomedical research studies are continuously evolving, and incorporate constantly new data types and formats, researchers often use local vernacular KOSs [12] in addition to the published KOSs for daily data annotation activities. Vernacular KOSs can take different forms: a standard modified to fit the needs of a study, a controlled vocabulary list shared among lab colleagues via excel files, terms from a domain ontology associated with other home-made and context-related terms, vernacular abbreviations, etc. In practice, a biomedical expert is living a dilemma. On the one hand, he is encouraged to use published KOSs in order to comply with the worldwide open data initiatives to share and reuse scientific data. On the other hand, local KOSs are more relevant to the researcher's own work. They can be enriched by terms from recent scientific advances, without the need to wait for updating of public KOSs, a slow process over which the researcher has no control. The use of abbreviated local terms is also easier for daily data annotation than the use of official terms. For instance, using "FDG" as a short term for "2'[18F]fluoro-2'-deoxy-D-glucose" is more convenient for oral and written communication. Therefore, sets of heterogeneous KOSs, both local and public, are generally used for data reporting. This further complicates biomedical Research Data Management (RDM) and generates difficulties in understanding shared data while hindering their reuse. In this context, biomedical researchers wonder how to make their data **understandable** by the scientific community, so that it can be reused with **confidence** long after the end of the study and without altering their research tasks?

The Biomedical Study Lifecycle Management BMS-LM data model made a step toward reuse with confidence by implementing a framework for traceability and data provenance annotation throughout the lifecycle of a research study [13]. Initially proposed for neuro-imaging data [14] it has been extended to the multi-modal data of preclinical research [13] such as histology, in vivo imaging with Positron Emission Tomography coupled to computed X-ray Tomography (PET-CT), and proteomics data. Provenance information describes the data and its original context, and redraws data history from their creation to their sharing [15]. As reported in [16], provenance is essential for trusting published scientific content. The more their sources are trusted, the more the data can be shared and the more reliable the research results are. The provenance of data acquisition and analysis includes the devices and their configuration, the algorithms used, the acquisition settings and processing parameters, the identification of the instrumentation, and others. The use of the BMS-LM data model prepares the data to be shared all the way to the end of the lifecycle, with all provenance metadata being collected all along the research study. Thus, **confidence** in data is privileged.

However, the BMS-LM data model do not include the biomedical researchers needs for flexible daily data annotation where the use of local terms and the non-conventional use of standard terms, a quasi-constant practice in any research team, is unavoidable. At the same time, it is an issue for data public sharing, because the local significance is not shared. How though to ensure the unbiased **understanding** of shared data that were annotated using local KOS? The proposition of this paper is to implement **semantic interoperability** between local and published KOSs, in the perspective to enable flexible annotation in the daily activities using local KOS as well as its public sharing using standard published KOSs. This paper details a methodology to implement the required semantic interoperability based on BMS-LM data model as a starting point to future industrialized software.

Semantic interoperability can be generally defined as “the ability of information systems to exchange information on the basis of shared, pre-established and negotiated meanings of terms and expressions.” [17] The semantic interoperability issue is treated in the literature from different perspectives. In ontology engineering, *OntoAnimal* tools and *eXtensible Ontology Development (XOD)* principles [18] enable the partial reuse of published ontologies, as well as the integration of local KOSs. In ontology matching, a compound alignment methodology was proposed [19], in order to boost inter-communication between ontologies. The Center for Extended Data Annotation and Retrieval (CEDAR) [20] has provided a researcher-friendly tool for metadata collaborative edition and standardization that is useful for terminologies management.

In this article, the semantic interoperability is handled from a knowledge organization point of view. It denotes the ability of different KOSs to be mapped to each other with preservation of the “intended meaning” according to the authors of [21]. This intended meaning could only be conveyed using explicit terms and knowledge while reducing implicit ones. Ontologies, defined as “explicit specifications of a shared conceptualization”, [22] are a good vector of explicit knowledge. Thus, the first objective of the present work was to convert the BMS-LM data model into an ontology. To design it, two recommendations from the state of the art were adopted: first, the use of a top foundational ontology encouraged by Gangemi et al. [23] to design a “conceptually more rigorous, cognitively transparent and efficiently exploitable” KOS, i.e. more semantically interoperable KOS, and the use of a three-level ontology construction approach, incorporating top, core and domain ontologies, that are believed to enhance semantic interoperability by Patel et al. [21].

A fourth level was added in our work to include local KOSs so that a maximum of local terms could be translated to their equivalent published terms when sharing the data. The paper presents methods and steps to build a four-level ontology that enables the semantic interoperability of local and

published KOSs. To the best of the authors knowledge, this is the first initiative to integrate institutional recommendations into laboratory members daily data annotations using ontologies.

The paper is structured as follows: the implementation methods using the top/core/domain/local levels are described. Then, the resulting 4-level semantic framework and BMS-LM core ontology are explained and illustrated with examples from the application context (small animal preclinical research). Finally, the BMS-LM core ontology is discussed and compared with other ontologies for semantic interoperability of heterogeneous KOS for biomedical data annotation. Conclusion and future work end the paper.

2 MATERIALS AND METHODS

2.1 THE BMS-LM DATA MODEL AS THE STARTING POINT OF THE PAPER

This section summarizes the major features of the BMS-LM data model described in [13] and converted later in the paper to an ontology. BMS-LM proposed two types of objects: Definition objects for provenance and Result objects for produced data (see 1-figures in the supplementary materials). Mainly, the BMS-LM data model covers all phases of a study lifecycle previously proposed in [14] as a set of four phases, inspired from [24]: (1) study specification (2) raw data acquisition, (3) derived data production, to (4) results publication. The data model is built with generic objects and traceability relations between them: a « Subject (SUB)» participates in a «Study (STU)» as a «Study Subject (SSU)». The latter undergoes «Exam Result (EXA)» defined by an «Exam Definition (EXD)» and composed of an «Acquisition Result (ACQ)». This is conducted using an «Acquisition Device (AQD)» that generates «Data Unit Result (DUR)». The DUR generated from the EXA is used in a «Processing Result (PCR)» composed of «Processing Unit Result (PUR)» and using a «Software Tool (STL)». All have their associated definition concepts (that not always end with the letter « D », «Processing Parameters (PCP)» and «workflow input (WFI)» are parts of the definition classes, for instance). At the end, the study results are published as a «Bibliography Reference (BBR)». The BMS-LM data model did not explicit all relations that exists between its classes (participates, undergoes, composed of, generated from, etc.). Only two types are used:

- Traceability relation: a link of data provenance traceability between two result classes (R) or two definition classes (D).
- Identification relation: a link between a result class and its corresponding definition class.

The semantics of the BMS-LM data model were designed for machine use rather than for human understanding. For this reason, we decided to transform it into a core ontology (see the results section).

2.2 DEFINITION OF THE DIFFERENT KOS LEVELS REUSED IN THIS ARTICLE

In Patel et al.'s three-level approach, a **top ontology** is general and independent of a particular context or domain [25], and it reflects a consistent and formal view of the world, also called ontological commitment (for more on ontological commitment see [25] and [26]). In the biomedical research field, Descriptive Ontology for Linguistic and Cognitive Engineering (DOLCE) [23] [26] and Basic Formal Ontology (BFO) [27][28] are the two most well-known ones. Top ontologies describe general concepts such as entities, concepts, perdurants (DOLCE), continuants (BFO), physical entities, immaterial entities, processes, etc. (Throughout the paper, the Courier New 10px typeface will be used for concepts)

Core ontologies [25] provide the explicit semantics of a large domain; each can also be considered as a reference ontology for their large domain. **Core Reference ontologies** have been proposed in very different areas such as film production [29] and Business Process Analysis [30]. A core ontology is essential to establish a minimal ontological commitment for a large domain [31], as it guarantees consistency between all specific metadata schemes (KOSs) used within this domain.

The third level is about **domain ontologies** [25]. All ontologies and KOSs that gather specific domain knowledge belong to this level. The DICOM standard [32], the RADLEX lexicon [33], and the Quantitative Imaging Biomarkers Ontology (QIBO) [34] are examples of **domain KOSs** from radiology. Some published **domain ontologies** reuse a top ontology for better semantic interoperability. For instance, the Ontology of Biomedical Investigations (OBI) [35] and the Gene Ontology (GO) [36] specialize the BFO top ontology [27], whereas, the OntoNeuroLOG ontology [37] specializes the DOLCE [23] top ontology.

Patel et al. [21] also gives two examples from the multimedia domain, a core ontology [38] and a domain ontology [39] that were proposed according to the top/core/domain approach and which have been positively evaluated for their enhancing of semantic and syntactic interoperability. For the BMS-LM ontology, the top/core/domain method is used.

2.3 BUILDING THE FOUR-LEVEL SEMANTIC FRAMEWORK AND THE BMS-LM CORE ONTOLOGY

Having the BMS-LM data model as a starting point and the top/core/domain/local ontological levels as objective, several steps were followed to build the BMS-LM ontology. First, the top ontology was chosen. Then, the BMS-LM data model was converted to a core ontology. After that, we gathered local

terms from the application domain to build both the domain and local levels. The domain level contains published KOS that can replace local terms and the local level is made of the remaining gathered local terms.

The BMS-LM core ontology was developed using the Protégé 5.1.0 and 5.5.0 software (<http://protege.stanford.edu/>) and the OWL-Lite (Web Ontology Language – Lite of the W3 Consortium). It was designed with a maximum reuse of published KOS and in collaboration with both data management experts and biomedical research experts.

2.3.1 The choice of a top ontology

To choose the BMS-LM top ontology from the most well-known top ontologies in the biomedical research field, DOLCE and BFO, a list of criteria was defined and applied, as indicated in Tab1. To ensure better semantic interoperability, the following three criteria were considered: metrics for community activity, citations number, and the number of reusing ontologies in the biomedical field. The number of reusing ontologies from BioPortal [9], the Pubmed number of articles, and the citations from Google Scholar were of particular interest. The comparison between these ontologies from an ontological point of view is a common practice, it reveals that BFO has a realism-based approach, while DOLCE has a cognitive bias [40]. However, this observation is irrelevant to the objective of this paper, concerned about the semantic interoperability maximization from a pragmatic point of view.

Tab1. The choice of a top ontology: results of the comparison between DOLCE and BFO top ontologies according to ontology reuse, citations and community criteria

	Criterion	DOLCE	BFO	Source
1	Number of biomedical ontologies reusing the top ontology	34 modules of OntoNeuroLOG (ONLOG) ontology	220 ontologies in OBO Foundry	respective official resources http://www.obofoundry.org/ http://neurolog.unice.fr/ontoneurolog/v3.0/Documentation_OntoNeuroLOGv3.pdf
2	Number of BioPortal ontologies reusing the top ontology	3 ontologies (ONLOG: physical object)	210 ontologies (BFO: material entity)	BioPortal, calculating the number of the SAME_URIs in the “mappings tab > source column” of the (ONLOG: physical object) and the (BFO: material entity). Updated the 20/08/2021
3	Number of articles in Pubmed	13 (DOLCE ontology)	43 (BFO ontology)	Pubmed, using the search tool with the keywords (DOLCE ontology) and (BFO ontology) on 20/08/2021
4	Community activity	Not maintained	Maintained	FAIRsharing.org [6] visited on 20/08/2021
5	Number of citations	1857 (DOLCE+ WonderWeb)	47(BFO) 2522 (OBO Foundry)	Google scholar for the following articles: wonderweb [26], DOLCE [23], BFO [27], OBO Foundry [28] on 20/08/2021

According to the comparison table results, BFO and its corresponding Open Biological and Biomedical Ontology (OBO) foundry consortium were chosen, for the following three reasons:

1. BFO is reused by more biomedical ontologies than DOLCE (see line 1 and 2 from Tab1.)
2. BFO is cited in more articles than DOLCE in the biomedical field and in general (see line 3 and 5 in Tab1.).
3. The BFO community is more active in the biomedical field than the DOLCE community (see line 4 in Tab1.).

2.3.2 From the BMS-LM data model to the BMS-LM core ontology

The BFO top ontology come with a set of principles from the OBO foundry consortium. These principles were used to review the BMS-LM data model. First, its objects were **renamed** and **redefined** according to the OBO Foundry principles. The reuse of already available definitions in other ontologies was privileged. Then, the top-core relations between the BFO top ontology and the BMS-LM core ontology were studied and added. An algorithm was implemented to assist the mapping decision. It had analyzed published ontologies in BioPortal and retrieved the top-core links for concepts that are synonyms to the BMS-LM core concepts. For example, a list of synonyms to “Exam Results” was established and used to search in BioPortal for equivalent terms. We selected specially ontologies that has BFO as a top ontology. The used algorithm is described in Fig2. It was implemented using the Knime Analytics Platform¹.

```

Algorithm 1 Script-Bioportal-Core
Input: A list T of core Terms
For all each t in T do
  Send GET search request to Bioportal via: http://data.bioontology.org/search?q=t
  Get the list TS of available synonyms of t from Bioportal
  For all each ts in TS do
    Send GET request to Bioportal to retrieve the full ontology name source of ts
    Get all parents of ts till THING
    ADD the record to an Excel Spreadsheet E
Delete redundancies from E
Write the E file to the disc

```

Fig2. The algorithm used to search BioPortal for core concepts to use them in the BMS-LM core ontology

The resulting Excel file, listing the identified concepts from BioPortal, indicates in this case all BFO parents linked to an “exam”. Then, our choice was made manually by identifying the redundant BFO parent of a concept among the list provided by the Excel file. In our point of view, this redundancy reflects the community consensus about a top-level mapping that we tried to follow as much as possible. The list of input terms and a sample of output ones are provided as supplementary material (see 2-script1).

¹ <https://www.knime.com/>

Next, the BMS-LM relations were redefined. On the one hand, they were made explicit using the Relation Ontology (RO) from the OBO foundry consortium. Relations such as `: is about`, `has part`, `has member`, `is specified output of`, `were of` were of great use to make the BMS-LM relations explicit. On the other hand, relations of provenance must be elucidated considering that the original BMS-LM data model relations were mainly used for traceability and provenance. The PROV-O [41] ontology was identified as a reference ontology in the field. It is proposed by the World Wide Web (W3) consortium where provenance is defined “as a record that describes **people**, institutions, **entities**, and **activities** involved in producing a piece of data or thing in the world”[42]. An **entity** can be physical, digital, or conceptual. An **activity** occurs within a time period and acts on one or many entities. An **agent** is responsible for the execution of an activity. `Entities`, `activities`, and `agents` are the three main concepts of PROV-O and are linked by a set of relations such as `was derived from`, `was associated to`, `was attributed to`, `was started by`, `used`. All were reused in the BMS-LM ontology to make provenance semantics explicit.

2.3.3 Building the local and domain levels of the application context

The first design rule of the BMS-LM semantic interoperability framework is the maximum reuse of, and matching with, existing ontologies. A script implementing the algorithm in Fig3. was used for this purpose. It explores BioPortal ontologies and assist ontologists to build the domain ontological level for the application context (in this work, small animal preclinical research). To use it, local terms were collected using interviews, mind maps, and data collection from the application context (see the results section for details). These were provided as input to the algorithm, which provides as output the list of the identified domain concepts together with their direct parent(s), children, and definitions. A list of input terms and a sample output file for the local term “souris” (“mouse” in French) are provided as supplementary material (see 3-script2).

```

Algorithm 2 Script-Bioportal-Domain
Input : A list T of local Terms
For all each t in T do
  Send GET search request to Bioportal via : http://data.bioontology.org/search?q=t
  Get the list TS of available synonyms from Bioportal
  For all each ts in TS do
    Send GET request to Bioportal to retrieve the full ontology name source of ts
    Get the direct parent of ts, its children and its definitions
    ADD the record to an Excel Spreadsheet E
Delete redundancies from E
Write the E file to the disc

```

Fig3. The algorithm used to search BioPortal for domain concepts to use them in the BMS-LM domain level

Once the domain concepts identified, the decision to include them in the domain ontological level and their matching to the BMS-LM core concepts has obeyed a criteria list. Mainly, the criteria of inclusion were (1) a resource’s relevance to the BMS-LM ontological commitment and scope, (2) its reliance on

a top ontology, (3) the consistency of the resource itself, and (4) the “redundancy” of the concept’s definition and use. Redundancy is an indicator of consensus between different ontology providers. It raises the level of the domain-core-top matching confidence. Whenever inconclusive results were encountered, biomedical experts were consulted.

When no shared domain concept was found using Algorithm 2, **local** terminologies were used. They formed the local ontological level for the application context. Local terms are valuable because they represent the vocabulary of the application context experts. Their inclusion could engage more biomedical experts in BMS-LM-guided data reporting. As a result, the specific domain ontology for this specific laboratory case contains two levels: a domain level where a maximum reuse of already published KOS is processed, and a local level where specific terms from the lab were added. The domain-local matching enables links between published terms, that are explicit for the community, and local terms used in practice in daily research data reporting.

2.4 THE BMS-LM ONTOLOGY PROPERTIES ENCODING

A two-phase strategy for properties’ encoding was adopted to enrich the BMS-LM domain and local ontological levels with relevant OWL properties. The first step was to form a collection of concept descriptors that answer the well-known Five Ws questions « *Who ? do What ?, Where ?, When ? and Why ?* » about each BMS-LM main concept. The Five Ws is also a reference tool for provenance information collection [43] [44]. Therefore, provenance descriptors are also collected using the Five Ws in this first step. The resulting output was a list of terms that describe the main concepts of BMS-LM. For instance, if the BMS-LM concept is “exam”, the list of terms would be: the operator doing the experiments, the examination protocol, the exam’s place, date, and prescription. A single element from that list is called a **descriptor** δ associated to a BMS-LM concept \mathcal{C} .

The second step was to encode the list of collected descriptors. This encoding must follow the OWL encoding format used in OBO Foundry ontologies. An analysis of mainly the OWL encodings in the IAO (Information Artifact Ontology), and OBI (the Ontology for Biomedical Investigation) [45] ontologies was therefore performed. It provided a list of possibilities to encode descriptors in an OBO-compliant manner, summarized in Fig4. The list was called **ADQIV** for **A**nnotation property, **D**ata property, **Q**uality, **d**ata Item, and **V**alue specification.

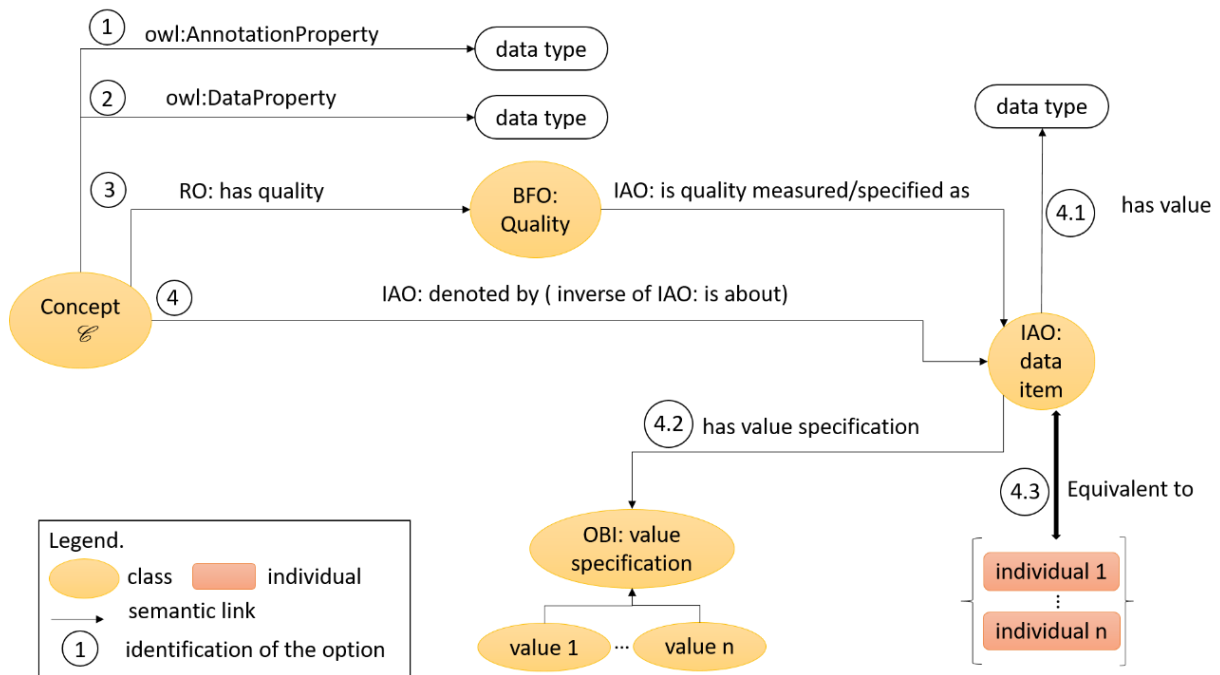


Fig4. The BMS-LM ontology encoding rules: the list of possibilities for the encoding of concepts descriptors in an OBO-compliant manner, called the ADQIV –Annotation property, Data property, Quality, data Item, and Value specification

With such a large choice, a strategy for a consistent encoding within the **ADQIV** list is required to simplify the process. The decision on how to choose whether a descriptor δ for a concept \mathcal{C} should be encoded as an `OWL:AnnotationProperty`, an `OWL:DataProperty`, a `BFO:quality`, or an `IAO:data item` depends on the following criteria:

- I- The specificity of the relation between the descriptor δ and the concept \mathcal{C} : if the descriptor δ is generic and not dependent to the concept \mathcal{C} , an existent `OWL:AnnotationProperty` could be used (option 1) or a new `OWL:DataProperty` could be added (option 2). For instance, the identifier, the name, a comment, etc.
- II- The nature of the first BFO parent of the concept \mathcal{C} (a `BFO:continuant` or a `BFO:occurent`): if the descriptor δ is about a `BFO:continuant` and is inherent to \mathcal{C} , the use of `BFO:quality` must be privileged (option 3) such as size, weight, height, otherwise the `BFO:data item` is directly used (option 4) for all `BFO:occurent` and the remaining `BFO:continuant`.

For data items, three sub-options are proposed. Either to give a simple value for the descriptor using the `BMSLM:has value` data property, or to define a list of its possible values. If the list is local and non-standard, its elements must have the type of an `OWL:individual`, otherwise they can be represented as an `OWL:class` using the concept `OBI:value specification`.

2.5 THE BMS-LM ONTOLOGY MATCHING ENCODING

Following the OBO Foundry principles, each BMS-LM entity, from the core, domain, and local levels, was matched, when applicable, with external entities, reused from other KOSs. This was done through formal matching using the `OWL:equivalentClass` and `OWL:equivalentProperty` in the case of strict equivalence. In other cases, the `skos:narrowMatch`, `skos:exactMatch` and `skos:broadMatch` from the SKOS (Simple Knowledge Organization System) W3C standard [46] were adapted, changed and used. Indeed, SKOS annotations are more suitable to relations between instances of KOSs concepts. As a consequence, `bmslm:broadMatch`, `bmslm:exactMatch` and `bmslm:narrowMatch` annotation properties were added to the BMS-LM ontology to express the same meaning but with less restrictions. For example, the `bmslm:data processing` concept is `bmslm:broadMatch` the `ncit:processing` concept. This means that the NCIT concept has a broader meaning than the BMS-LM one. This annotation prepares for future alignment between the BMS-LM ontology and the NCItthesurus. When an external entity from a published KOS was really reused (mainly for the domain level), the MIREOT (Minimum Information to Reference an External Ontology Term) [47] checklist was adopted as follows:

- the **source ontology URI** and the **source term URI** are reported using the `IAO:imported from` property; and
- the **superclass** is referenced using the `IAO:in branch` property.

In the case of an external definition reuse from published KOSs. The source was referenced using both annotations: `IAO:definition source` and `RDFS:isDefinedby`.

Regarding the interoperability of the BMS-LM ontology with previous developments (that might use the BMS-LM data model), some annotations were added to the core concepts to refer to the BM-LM data model old labels: `OWL:backwardCompatibleWith` was used for the previous data model names, and the `IAO:alternative term` was used to list the short and technical names of a data model concept.

3 RESULTS

Two main results are detailed in this section. First, the BMS-LM core ontology as an evolution of the BMS-LM data model. Second, the top/core/domain/local semantic interoperability framework made up with the BFO top level ontology and the BMS-LM core level ontology. It enabled the integration of local terms with published concepts using mapping relations at each level. As a preliminary test, examples from local terms of preclinical data were provided and discussed.

3.1 THE BMS-LM CORE ONTOLOGY AS AN EVOLUTION OF THE BMS-LM DATA MODEL

3.1.1 BMS-LM Concepts

A list of 29 concepts developed for BMS-LM is provided in Tab2. For each concept, its previous data model name and its new ontology name are specified as well as its parent in the BFO top ontology [62]. Possible matchings in other ontologies are indicated to inform about possible integrations for interoperability.

Tab2. Renaming and changes in generic concepts from the data model to the core ontology and their mapping to top-level concepts

	<i>BMS-LM previous name</i>	<i>BMS-LM new name</i>	<i>BFO parent</i>	<i>matchings in</i>
1.	Study	biomedical research study	planned process	DDI
2.	Subject	studied organism	OBI: organism	OBI,I OBC
3.	Sample Result	biological sample preparation	planned process	OBI
4.	Sample Result	biological sample	Object	NMR, OBI
5.	Sample Definition	sample preparation protocol	OBI: protocol	
6.	Agent Result	agent administration	planned process	
7.	Agent Result	agent product	Object	QIBO
8.	Agent Definition	agent administration protocol	OBI: protocol	
9.	Acquisition Device	device	OBI: device	OBI
10.	Exam Result	examination	planned process	Onto VIP
11.	Exam Definition	examination protocol	OBI: protocol	
12.	Acquisition Result	data acquisition	planned process	NCIT
13.	Acquisition Definition	data acquisition protocol	OBI: protocol	
14.	Data Unit Result	raw data unit	IAO: i-c-e*	IAO
15.	Data Unit Definition	expected data unit	IAO: i-c-e*	
16.	Processing Results	data processing	planned process	NCIT
17.	Processing Definition	data processing protocol	OBI: protocol	
18.	Processing Unit Result	derived data unit	IAO: i-c-e*	
19.	Processing Unit Result	data processing unit	planned process	
20.	Processing Unit Definition	processing unit protocol	OBI: protocol	
21.	Workflow Input	workflow input profile	process profile	
22.	Processing Parameters	processing parameters profile	process profile	
23.	Bibliographic Reference	published content	IAO: i-c-e*	IAO
24.	Intervention Result	biomedical intervention	planned process	ERO
25.	Intervention Definition	biomedical intervention protocol	OBI: protocol	
26.	Software Tool	software tool	IAO: software	
27.	Study Subject	study subject	role	NDD O
28.	Study Subject Group	study subject group	role	
29.	Reference Data	reference data	IAO: i-c-e*	

*:i-c-e = information content entity

During the conversion process of the BMS-LM data model to a core ontology, some concepts were split into two concepts to make their semantics more explicit, which explains why some previous names have become redundant (highlighted in **bold** in Tab2.). For instance, the `Processing Unit Result (PUR)` concept, defined as a process that analyzes input data and produces output results for another PUR, was expanded into the `derived data unit` concept to reference the produced data, and the `data processing unit` concept to describe the process executed.

3.1.2 BMS-LM Relations

The BMS-LM data model relations were extended in the ontology. Tab3. shows the new BMS-LM relations. Their OBO Foundry source ontologies are indicated and some use examples from the BMS-LM ontology are given. This table also gives information about the different relations categories in the BMS-LM core ontology.

Tab3. The list of BMS-LM relations, their source ontologies from the OBO Foundry, examples of their use from the BMS-LM core ontology, and their categories. min = minimum cardinality / exactly = exact cardinality

	<i>relation name</i>	<i>source ontology</i>	<i>BMS-LM example of use</i>	<i>category</i>
1	has member	RO	study subject group → min 1 study subject	aggregation
2	member of	RO	Inverse of previous	aggregation
3	has part	BFO	examination → min 1 data acquisition	composition
4	part of	BFO	Inverse of previous data processing unit → exactly 1 data processing	composition
5	has participant	RO	biomedical research study → min 1 study subject	contribution
6	participates in	RO	Inverse of previous	contribution
7	has role	RO	studied organism → min 1 study subject	qualification
8	role of	RO	Inverse of previous	qualification
9	has specified input	OBI	agent administration → min 1 agent product	ingestion
10	is specified input of	OBI	Inverse of previous raw data unit → min 0 data processing unit	ingestion
11	has specified output	OBI	data processing unit → min 1 derived data unit	production
12	is specified output of	OBI	Inverse of previous biological sample → exactly 1 biological sample preparation	production
13	realized in	BFO	study subject → min 1 biomedical research study	realization
14	Realizes	BFO	Inverse of previous	realization
15	derives from	RO	biological sample → min 1 studied organism	transformation

In Tab3., relations are categorized to make their role explicit in the BMS-LM core ontology. To be specific, **aggregation** and **composition** correspond to the classical UML relations modeled respectively with \diamond and \blacklozenge . **contribution** is when a BFO:continuant participates in a BFO:occurent. **production**, and its inverse category **ingestion**, designate the output and the input of a BFO:process respectively. **transformation** is when a BFO:continuant is a derived result of another BFO:continuant. **qualification** is a feature relating to an inherent property of a BFO:continuant. A **realization** is when a BFO:realizable entity is realized in a BFO:process.

3.1.3 Provenance and lifecycle concepts and relations

This section contains an explanation of how the BMS-LM core ontology makes provenance and lifecycle semantics explicit. In fact, the information coverage of provenance and all lifecycle phases is the key to end-to-end traceability of biomedical data. It is an old feature of the BMS-LM data model that was preserved in the ontology as explained bellow.

3.1.3.1 Provenance

The BMS-LM ontology was designed to be interoperable with the general ontology PROV-O. This makes the BMS-LM provenance semantics explicit and allows semantic interoperability with PROV-O-

compatible ontologies. In practice, PROV-O has, in its initial version, a list of three main concepts: `Agent` (PROV-O), `Entity`, and `Activity` (see the Methods section). Tab4. provides a classification of the BMS-LM concepts according to these general concepts.

Tab4. Classification of the BMS-LM concepts according to PROV-O main concepts

Agent (PROV-O)	Entity	Activity	Not compatible
software tool	expected data unit	examination	examination protocol
agent product	derived data unit	data acquisition	data acquisition protocol
study subject	raw data unit	data processing	data processing protocol
device	reference data	data processing unit	data processing unit protocol
principal investigator	published content	agent administration	agent administration protocol
operator	biological sample	biomedical intervention	biomedical intervention protocol
technical referent	study subject group	biological sample preparation	biological sample preparation protocol
	studied organism	biomedical research study	workflow input profile
			processing parameters profile

Some BMS-LM concepts do not fit the three-concept model of PROV-O. Indeed, the concepts whose parents are `OBI:protocol` and `BFO:process profile` (see Tab2.) are not compatible with the definitions of `Agent` (PROV-O), `Entity`, or `Activity`. However, they complete the provenance information about a `BFO:process` in the BMS-LM core ontology.

PROV-O relations use the past tense, which is not compliant with the OBO foundry Relation Ontology (RO) who uses the present tense. This motivated their renaming from “was” to “is”. Tab5 details the BMS-LM core ontology relations for provenance. It includes a list of native BMS-LM relations as well as reused relations from the PROV-O ontology and from the OBO Foundry consortium..

Tab5. The list of the provenance relations in the BMS-LM core ontology. For each relation, its owl:domain and owl:range were indicated, an example of its use and its source ontology (PROV-O from the W3C, RO from OBO Foundry, or BMS-LM core ontology) are provided. min = minimum cardinality / exactly = exact cardinality

relation name	Domain/Range	example of use	source
<i>acts on behalf of</i>	Agent→Agent	operator →min 1 operator	PROV-O
<i>has member</i>	Entity→Entity (extended to *→*)	study subject group →min 1 study subject	PROV-O RO
<i>is associated with</i>	Activity→Agent (extended to *→*)	examination →min 1 study subject	PROV-O
<i>is attributed to</i>	Entity→Agent	examination protocol →min 1 operator	PROV-O
<i>is derived from</i>	Entity→Entity	biological sample →min 1 studied organism	PROV-O RO
<i>is generated by</i>	Entity→Activity	biological sample → exactly 1 biological sample preparation	PROV-O
<i>is informed by</i>	Activity→Activity	raw data unit →min 1 data acquisition	PROV-O
<i>Uses</i>	Activity→Entity	data processing unit →min 1 data processing	PROV-O
<i>is influenced by</i>	*→*	examination →min 1 device	PROV-O
<i>has for protocol</i>	process→protocol	*→*	PROV-O
<i>has for process profile</i>	process→process profile	examination →exactly 1 examination protocol	BMSLM
<i>has for parameters</i>	process→processing parameters profile	*→*	BMSLM
		data processing unit →min 1 processing parameters profile	BMSLM

3.1.3.2 Lifecycle

Although built to cover all lifecycle steps, the BMS-LM data model did not have explicit concepts for study lifecycle. In the BMS-LM core ontology, it has been decided to add the lifecycle steps of the BMSLM data model: a `study's lifecycle phase` concept and its four children: (1) `study specification`, (2) `study data collection`, (3) `study data analysis`, and (4) `study results publishing` concepts. However, we did not find relevant materials to help us identify its parent in the BFO ontology except in the UBERON [48] ontology. It defines a lifecycle as: “UBERON:0000105 - `life cycle stage` -: A spatiotemporal region encompassing some part of the life cycle of an organism”. The spatiotemporal dimension is relevant to an organism because it has a material and physical existence in space, whereas, for a study, only the time dimension seems to be relevant. The BMS-LM `study lifecycle phase` concept can thus be set as a subclass of the `BFO:one-dimensional temporal region`.

In terms of relations, each concept of BMS-LM ontology was associated with one or more study lifecycle phases to ensure traceability. All concepts can be related to study lifecycle phases, with the relation `PROV-O:is associated with` originally from the PROV-O ontology but adapted to BMS-LM requirements.

3.1.4 Accessibility

The OWL-Lite encoding of the BMS-LM ontology is available in the following URL (<http://www.fealinx-biomedical.com/ontologies/bmslm/1.0/bmslm.owl>) and as supplementary material of the present paper (see 4-BMSLM_ontology).

3.2 THE BMS-LM CORE ONTOLOGY AS PART OF THE FOUR LEVELS SEMANTIC INTEROPERABILITY FRAMEWORK

The BMS-LM semantic interoperability framework is composed of four levels: top, core, domain, local (see Fig5.).

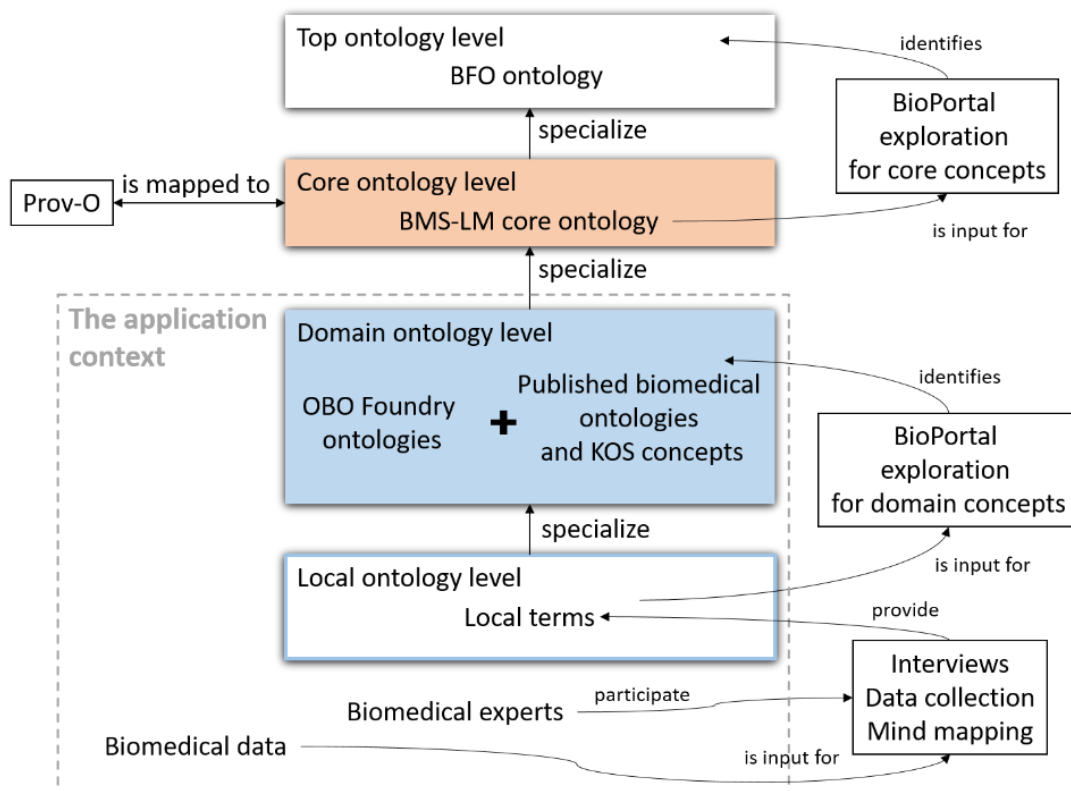


Fig 5. A summary of the BMS-LM four levels semantic interoperability framework

To comment on Fig.5, the local level was added to the state-of-the-art three levels, to take into consideration the specificity of the laboratory and the research team that produced the data. After data collection, interviews and mindmapping with biomedical expert, local terms were collected and used as input to BioPortal exploration script for the domain concepts search (see the Methods section, Algorithm2). The BioPortal exploration help identifying related concepts from published KOSs and OBO foundry ontologies. Those concepts were reused and aggregated in the domain ontological level. Their reuse process respects the MIREOT principles [47]. When local terms could not be replaced by domain ones, they were kept in the local ontological level. Thus, the local level specialized the domain one and the latter specialized the core level where the BMS-LM core ontology is used. The BMS-LM core ontology, mapped to the Prov-O ontology, provided a framework to enable biomedical data annotation with provenance information throughout the study lifecycle. The use of the BFO top ontology has made the resulting ontology more interoperable with published ontologies. In fact, BFO has an interoperability common ground with more than 200 biomedical ontologies via the OBO Foundry consortium.

3.2.1 Link with the top level

The BMS-LM core ontology was built and mapped to the BFO top ontology following the OBO Foundry guidelines in the Arp et al. book [49], its compliance with the OBO 11 principles [50] is explained in Tab6.

Tab6. BMS-LM core ontology construction rules: their compliance with the OBO Foundry 11 principles

	OBO principle	application in BMS-LM
1.1	Availability	Available at http://www.fealinx-biomedical.com/ontologies/bmslm/bmslm.owl
1.2	Licensing	Available under CC-BY 3.0 license
1.3	Other ontologies' reuse: credit and annotations	BMS-LM follows the MIREOT (Minimum Information to Reference an External Ontology Term) checklist [47]
2	Common format	The official release is in RDF/XML syntax.
3	Identifier in <IDSPACE>_<NUMBER> format	All elements of BMS-LM ontology are identified following a BMSLM_0000000 format.
4	Versioning	The versions of the BMS-LM ontology follow the M.m format. The follow up of the development process is done using SubVersion (svn).
5	A clear scope	The BMSLM ontology scope is indicated in the OWL file, it states that it is a core ontology that covers biomedical studies lifecycles and provenance management with maximum reuse of existing ontologies.
6	Textual definition	All core concepts are defined using the <IAO: textual definition> annotation.
7	Relations: reuse of the Relation Ontology (RO)	The RO ontology relations are all reused for BMS-LM ontology.
8	Documentation	Documentation of the BMS-LM ontology is done through embedded comments in situ when necessary.
9	Documented Plurality of Users	The BMS-LM ontology is used in the DRIVE-SPC project (2015REMP9HE085 grant from Paris University) and will be used in the PACIFIC project (BPI n°2018- PSPC- 07) and PsyCare project (ANR 18-RHUS-0014). All of them are in France. Information about reuse can be found in the list of OWL ontology annotations.
10	Commitment To Collaboration	The BMS-LM ontology construction is mainly based on maximum reuse and matching with existing ontologies, which reinforces collaboration between biomedical KOSs.
11	Locus of Authority	In the BMS-LM ontology annotations, the name and email address of the corresponding author are given. This author is the reference person for the ontology.

3.2.2 Link with the domain and local level of preclinical research

The BMS-LM core ontology was applied to the preclinical research studies conducted on animals, with a therapeutic or diagnostic objective of output to human beings. Small animal preclinical research data were produced by the IVIR (In-Vivo Imaging Research) laboratory of the Paris Cardiovascular Research Center (PARCC) in the context of the DRIVE-SPC project between the IVIR laboratory and the Fealinx company (2015REMP9HE085 grant from Paris University). We used mainly data from [50], together with data from ongoing biomedical studies. An exploration work was realized using data collection, recurrent data explanation meetings during the 6 first months of 2017, and structured interviews with identified key persons among the IVIR laboratory members (5 users, 5 structured interviews, 7 specification meetings, 9 follow-up meetings) during four months from February 2018 to June 2018. In the supplementary material (FigSM2 in 1-figures), a detailed interview structure is given (collected answers are confidential, which is the reason why we could not provide all the interviews content). A mind map of the local terms collected from the PET-CT data is also given as supplementary material (see 5-local_terms). As a first run, 137 local input terms were used. They resulted in 3220 rows (concepts from BioPortal) and were analyzed to build the domain and local levels.

In the following figures, examples are given to show how the local and domain levels were constructed from sample data.

In Fig6., an image of histology from the pilot laboratory was annotated with local terms. “Staining”, “DAPI”, “Lectin”, “FOV” are some of collected local terms from interviews. Using Algorithm2 described in the method section, domain concepts were identified from published KOSs in Bioportal. Concepts from PATIT (Placental Investigative Technique) ontology, OBI ontology, FBbi structured controlled vocabulary, NCIT (National Cancer Institute Thesaurus) were identified (see blue boxes in Fig6.). The reused domain level concepts are standard terminologies that are accepted by data publication repositories. When no standard domain term was found, a concept was added to the local level. For instance, the local term “Apotome” which references the microscope used to acquire the histology image was not found in Bioportal. Thus, a new concept `DRIVE: Axioimager Apotome` was added to the local level. This new concept enables the reporting of the microscope’s different configurations used when acquiring the histology data.

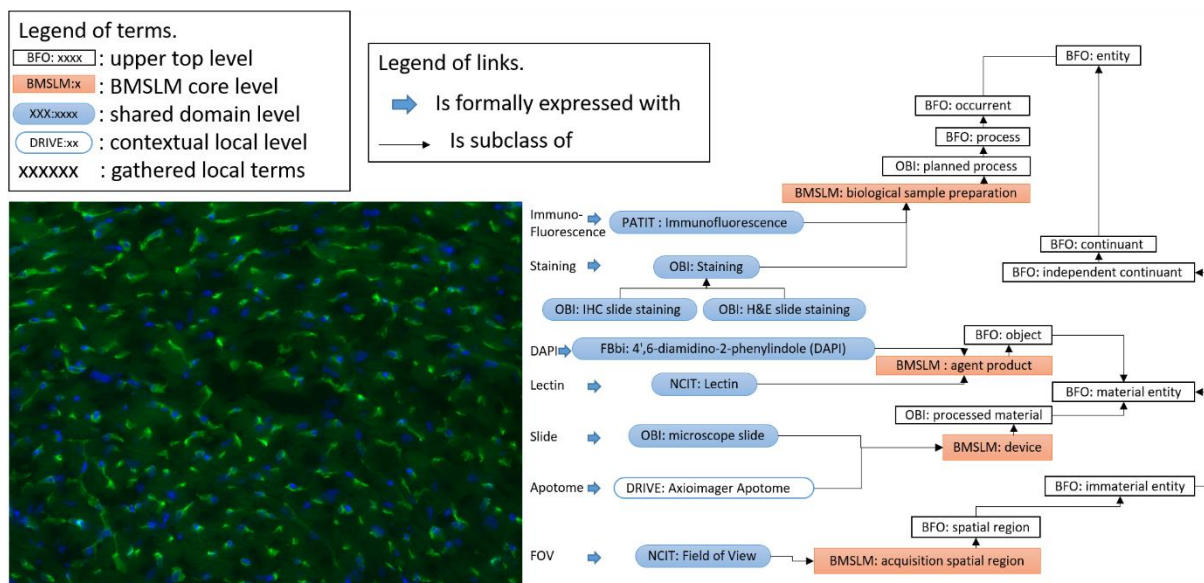


Fig6. Use of the BMS-LM core ontology with data from Histology domain.

Each application-specific concept (local and domain levels) was linked to the generic core level of the BMS-LM semantic interoperability framework. For example, the `DRIVE: Axioimager Apotome` was linked to the core concept `BMS-LM:device`, which was linked in the more abstract level to an `OBI: processed material` classified as a `BFO: material entity`. These four levels enabled the understanding of local terms by creating a structured tree of abstract (or generic) concepts around it. This way, even a non-expert person will be able to understand the general meaning of the local term.

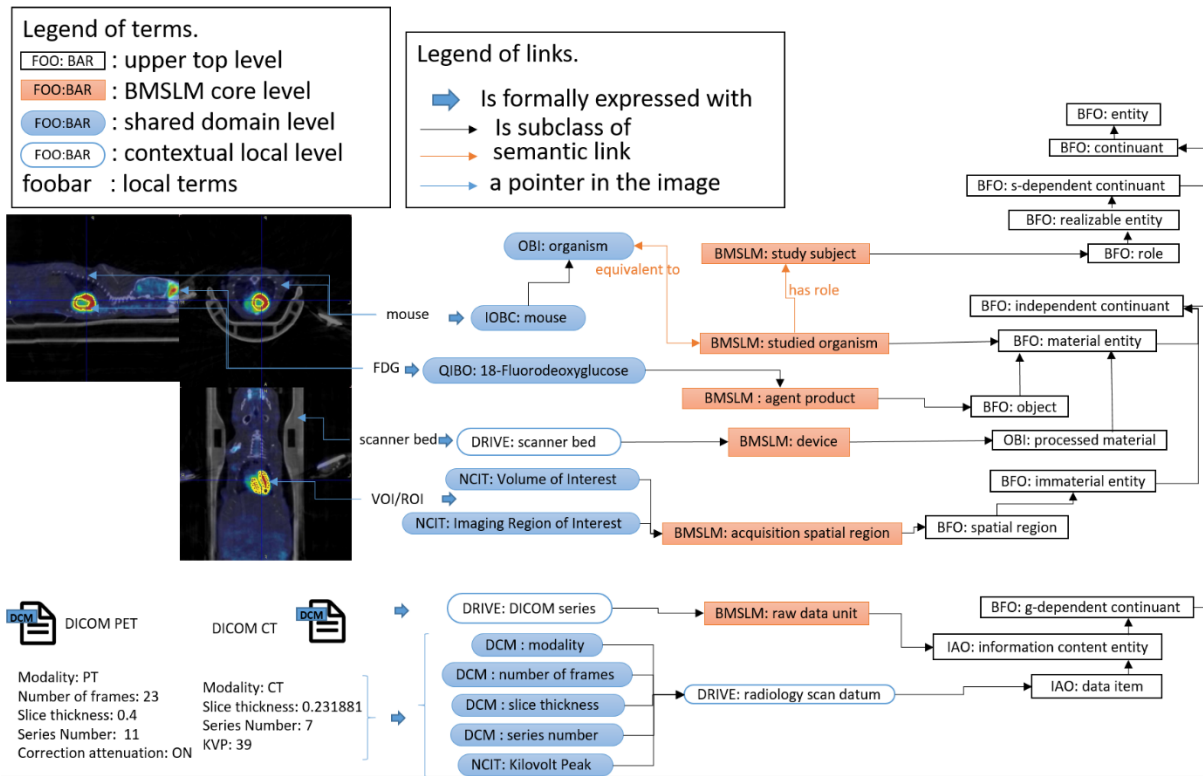


Fig7. Use of the BMS-LM core ontology with data from PET-CT imaging domain.

Fig7. presents the four levels structure for PET-CT imaging data. The PET-CT image shows a mouse in the scanner bed. It was injected with a radioactive substance (coded from red to yellow in the image). By examining the four-level structure starting from local terms to top concepts, one can identify the radioactive substance used for the experiment. The four levels were constructed as follows: from the local term 'FDG', its full name 'QIBO:18-Fluoredeoxyglucose' was identified in the domain ontology QIBO [34], which was linked to the core level as a 'BMSLM: agent product' which is classified as an 'object', a 'material entity' and an 'independent continuant' in the top level. By reading the structure from top to local, the "FDG" must be identified as the active product, or more precisely the radioactive tracer, used in this PET-CT experiment. This four level structure is not only understandable by domain experts working with FDG, who will concentrate on the domain-specific levels (domain + local), but it is also accessible to the non-specialist that will manage to understand using the remaining two generic levels (core + top).

Fig6. and Fig7. were shown to three users from the IVIR laboratory in order to evaluate the new ontology and its associated framework. The questionnaires used for the test are provided as supplementary material (see 6-questionnaires). Briefly, a first set of 11 questions was asked to assess their ease with the original BMS-LM data model. Then, both figures were explained, starting from local terms up to top terms. Naturally, the domain and local levels were more familiar to the users. After explanation, another set of 6 questions was asked to determine whether the core level (the BMS-LM

ontology) was better understood. Results are shown in the table below. It is noteworthy that all the users did progress and two managed to reach a score of 100%. Furthermore, two users confirmed just after the test that the ontology was clearer than the data model.

Key users (From the less “used to the data model” to the better used to it)	user1	user2	user3
Starting level	1.75/11 = 15.9%	6/11 = 54.55%	6.75/11 = 61.36%
After test level	4/7 = 57.14%	7/7 = 100%	7/7 = 100%
% of progress	41.24%	45.45%	38.64%

In the same spirit, Fig8. shows (from bottom to top) a list of local terms collected in the IVIR laboratory, their equivalent standard domain concept published in Bioportal, and their core and top corresponding concepts. Terms were collected from different domains: proteomics, histology, and in vivo imaging. This shows how the 4-layer semantic framework of the BMS-LM core ontology is inclusive and can easily integrate different modalities and heterogeneous data and KOS.

4 DISCUSSION

The present study describes the BMS-LM core ontology and its semantic interoperability framework applied to preclinical research data. It shows that the use of ontologies rather than data models offers a better solution towards interoperability and explicit semantics. As an example, the BMS-LM data model name for an examination protocol is `Exam Definition` which is less explicit than its equivalent in the BMS-LM ontology `BMSLM: examination protocol`. Likewise, the BMS-LM data model concept `Sample Result` could be understood with two different meanings: is `Result` the object output of the sampling process or the documentation of the sampling process? To make it clear, the `Sample Result` concept was split in two separate concepts in the BMS-LM ontology, the `biological sample preparation` describing the process and the `biological sample` describing the resulting object. The renaming of the BMS-LM data model concepts was of great use to explicit the core level semantics to biomedical experts. For instance, explaining that a STU (short name for a `Study`) has a list of SSU (short name of a `Study Subject`) which is linked to its EXA (short name for an `Exam`) and PCR (short name for a `Processing`) is far more complicated then explaining that a `biomedical research study` has a `study subject` which is linked to its `examination` and `data processing`. Concerning relations, the BMS-LM data model proposed two types of relation: identification and traceability. In practice, however, relations between BMS-LM concepts are richer and convey a more specific meaning, such as composition, aggregation, etc. Thus, to facilitate the understanding of their meaning, their semantics were made explicit with the use of the OBO Foundry ontology of relations (RO), BFO, and OBI.

Actually, the BMS-LM ontology is based on the BFO top ontology to ensure explicit and shared semantics. It reuses a set of OBO Foundry domain ontologies: RO, IAO, and OBI for concepts and relations. Potential overlaps between OBI and BMS-LM were identified. OBI is directly based on the BFO, hence it has both core and domain concepts for biomedical investigations. OBI domain concepts may be reused when relevant, like any other biomedical KOS. However, some OBI core concepts are equivalent to BMS-LM core concepts, such as `OBI:protocol` and `OBI:planned process` (for more explicit semantics, the latter's preferred name in the BMS-LM ontology is "`BMSLM:plan realization process`"). Nevertheless, the scope of each is different. The OBI concepts do not cover the provenance or the study lifecycle. In addition, OBI is not a core ontology, even though it contains core concepts. A discussion with the BFO and OBI communities may help remove the overlaps identified in the next versions of both ontologies, given that BMS-LM submission to the OBO Foundry consortium and publication in BioPortal is in process.

The BMS-LM core ontology was also mapped to the provenance ontology PROV-O from the W3C. Since the concepts in PROV-O do not cover all the complex provenance information of the BMS-LM, other provenance concepts were added. For the current BMS-LM ontology version (1.0), only the *Agent*, *Activity*, and *Entity* main concepts of PROV-O were found relevant. Other PROV-O concepts may be integrated in future versions.

The BMS-LM core ontology was designed to incorporate the maximum reuse of, and matching with, heterogeneous KOSs. The use of the BMS-LM semantic interoperability framework and methods enables the semantic interoperability in the application context by facilitating data understanding and sharing between humans and by reusing published KOS. Four KOS were reused in the domain level shown in Fig6. and five for the domain level shown in Fig7. The semantic interoperability was then enabled by the conversion from the BMS-LM data model to the BMS-LM ontology. Further implementation of the top/core/domain/local ontology using data from a specific biomedical research study (including clinical study) is beyond the scope of this paper and is the focus of ongoing studies in cardiology (the PACIFIC Consortium to redefine heart failure with preserved ejection fraction (HFpEF)²) and psychology (The PsyCARE French consortium aiming to improve early intervention in psychosis³). Nevertheless, examples from the preclinical domain were provided and validated with a three-user test to show the relevance, and utility for data understanding, of local terms' inclusion in the ontology construction process.

In the literature, a list of works have been identified to deal with heterogeneous KOS and data. Brahaj et al. [51] proposed the Core of Scientific Metadata Ontology (CSMO) to report contextual information about scientific data such as hardware, institutions, people, software, studies, investigations or experiments, resulting data, and related publications. However, no consideration was given to the local terms and the semantic interoperability between KOS.

The ISA software suite [52] proposes a three-concept model (Investigation, Study, Assay) together with a variety of tools (ISACreator, ISAConfigurator, ISAValidator...) for the reporting of biological and molecular investigations by researchers. Recently, the ISA original format, ISA-Tab, was converted to RDF according to the LinkedISA approach using semantic web standards [53]. This promising set of tools for the annotation of biomedical data, built upon BFO to ensure more semantic interoperability, does not cover all the lifecycle of a research study.

The EXACT (EXperiment ACTions) ontology was proposed [54], after a review of published protocols, to describe experimental protocols efficiently and unambiguously, and thereby reinforce their

² ClinicalTrials.gov Identifier : NCT04189029

³ <https://psy-care.fr/>

reliability and reproducibility. A more recent work provides a framework for protocol conversion from a natural language format to a semantically-defined format [55] to enable machine and human protocol reuse. A similar effort, focused on simplifying and making protocol semantics explicit, is the SMART Protocols (SP) ontology [56] and its associated “minimal information” SIRO (Sample, Instrument, Reagent, and Objective) model. Both are promising initiatives but only cover the data acquisition part and do not integrate local terms in their models.

A simple domain-independent methodology and core ontology called SECO (Scientific Experiments Core Ontology) for LIMS (Laboratory Information Management System) data annotation was proposed recently [57]. SECO reuses domain ontologies and is based on BFO top ontology. SECO has the same objectives as the BMS-LM core ontology, create a domain-independent ontology, but it is limited to LIMS data and does not consider local terms and provenance annotations.

Thus, the advantage of using the 4-level framework to implement specific-domain ontologies is that it takes into account the local specific terms, the worldwide recommendation of published KOS reuse, the study lifecycle coverage, and the semantic clarity of the chosen labels. The 4-level framework thus ensures the semantic interoperability with published KOS, especially those reusing BFO and mapped to PROV-O. It helps the biomedical researcher manage KOS heterogeneity through the use of a four-level ontology construction method. In practice, a biomedical researcher annotates his/her data using his/her local terms to ensure smooth and easy intra-lab exchanges. The conversion of local annotations to published standard ones can be supported by a translation process that automatically converts local terms into published ones using the four levels of the BMS-LM based ontology. The data could then be shared with standard annotations that facilitates its understanding and reuse. It is to be stressed however that ideally, the collection of local terms and their mapping to other levels should be done by a referent person (ontologist, data manager) interacting with the research team.

Promoting reuse releases data providers from the burden of having to implement their guideline-compliant procedures from scratch. Indeed, the BMS-LM core ontology plays a mediator role between the top-down incentives of institutions and governments to improve Research Data Management (RDM), and the bottom-up initiatives of data providers to structure their local terminologies. The FAIR guidelines compliance (top-down) and the local terms' inclusion (bottom-up) are two features of the 4-level semantic framework that tend to integrate both approaches (top-down and bottom-up), in an operational way.

5 CONCLUSION AND FUTURE WORK

The BMS-LM core ontology was built according to an OBO-compliant and FAIR-compliant approach. It was proposed together with its semantic interoperability framework in order to enable the semantic interoperability with heterogeneous KOS used for data annotation in a specific context. The objective was not to eradicate KOS heterogeneity, as it is inherent to scientific research. Instead, this work adopted an inclusive strategy, including published KOSs as much as possible via their partial reuse, and including different profiles of expertise with the use of local, domain, core, and top ontological levels, with the aim to foster understanding between biomedical experts, biomedical KOS experts, and data managers.

From a bottom-up perspective, each biomedical researcher uses his own vernacular terms. During data sharing or publishing, annotated data is re-annotated with more standard-compliant metadata (i.e. a set of MESH annotations, or RADLEX lexicon terms) through matchings from one level to another. Future work will focus on the automation of reuse and matching of KOSs to enable instant interoperability, and on 'in situ' knowledge capture to collect more relevant local terms (also called interface terms in this context). The latter aspect will enable knowledge capture at the local level via ergonomic forms, GUIs (Graphical User Interfaces), and other HMIs (Human Machine Interfaces). The collected local terms will be mapped from level to level, up to the top ontological level. Naturally, this is fully dependent on further research in ontology matching, mapping, and evolution.

Two key lessons could be taken from this work. First, top-down recommendations must be instantiated in the bottom-up level to enable their day-to-day application. Second, data reporting with interoperable ontologies by biomedical experts will likely help the process.

List of main abbreviations

ADQIV: **A**nnotation property, **D**ata property, **Q**uality, data Item, and **V**alue specification

BMS-LM: BioMedical Study – Lifecycle Management

DICOM: Digital Imaging and Communications in Medicine

DMP: Data Management Plan

FAIR: Findable Accessible Interoperable Reusable

KOS: Knowledge Organization System

NCBO Bioportal: National Center for Biomedical Ontology Bioportal (a repository for ontologies)

OBO: Open Biological and Biomedical Ontology

OWL: Ontology Web Language (W3C standard)

PROV-DM: PROVenance Data Model

PET-CT: Positron Emission Tomography – Computed Tomography

RDF: Resource Description Framework

RDM: Research Data Management

SKOS: Simple Knowledge Organization System (W3C standard)

UML: Unified Modeling Language

Ontologies

BFO: Basic Formal Ontology

DCM: DICOM Controlled Terminology

DOLCE: Descriptive Ontology for Linguistic and Cognitive Engineering

FBbi: Biological Imaging Methods Ontology

IAO: Information Artifact Ontology

IOBC: Interlinking Ontology for Biological Concepts

MESH: Medical Subject Heading

MS: Mass Spectrometry Ontology

NCIT: National Cancer Institute Thesaurus

OBI: Ontology of Biomedical Investigations

PATIT: PAT placental Investigative Technique

PROV-O: PROVenance Ontology

QIBO: Quantitative Imaging Biomarker Ontology

RO: Relation Ontology

CONFLICT OF INTEREST

The authors declare no conflict of interest. ARA, MAL, PYH, PBO are employees of Fealinx SAS.

FUNDING

Amel Raboudi is supported by the ANRT CIFRE scholarship n°216/1649. The DRIVE-SPC project is supported by the COMUE Sorbonne Paris City University and the Fealinx Company. The work presented in this paper was partially supported by the PACIFIC (BPI n°2018- PSPC- 07) and PsyCARE (ANR 18-RHUS-0014) projects.

AUTHOR'S CONTRIBUTION

ARA, PYH, JCA, JHI, and MAL contributed actively to the ontology encoding. BEY participated in the lifecycle semantics aspect. DBA, TVI, and BTA reviewed the application examples. ADU, DBA, BEY, and PBO reviewed the structure and content of the paper from its earlier incarnation. TVI, TYO, and ACE contributed to the interviews and data collection phase. ARA drafted the manuscript. ARA, DBA, MAL, PYH, BEY, TVI, ADU, PBO, BTA revised it. All authors read and approved the final version of this paper.

ACKNOWLEDGMENT

The authors wish to express their sincere appreciation to the first key users of the BMS-LM early versions: Drs. Gwennael Autret, Caterina Facchin, Anikitos Garofalakis, Maily Perez-Liva, and Joevin Sourdon for their feedback and time.

6 REFERENCES

1. Rance, B., Canuel, V., Countouris, H., Laurent-Puig, P., Burgun, A.: Integrating Heterogeneous Biomedical Data for Cancer Research: the CARPEM infrastructure. *Applied Clinical Informatics*. 7, 260–274 (2016). <https://doi.org/10.4338/ACI-2015-09-RA-0125>.
2. Anderson, N.R., Lee, S., Brockenbrough, J.S., Minie, M.E., Fuller, S., Brinkley, J., Tarczy-Hornoch, P.: Issues in biomedical research data management and analysis: Needs and barriers. *Journal of the American Medical Informatics Association*. 14, 478–488 (2007). <https://doi.org/10.1197/jamia.M2114>.
3. Borgman, C.L.: The conundrum of sharing research data. *Journal of the American Society for Information Science and Technology*. 63, 1059–1078 (2012). <https://doi.org/10.1002/asi.22634>.
4. Wilkinson, M.D., Dumontier, M., Aalbersberg, I.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.: The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*. 3, 160018 (2016). <https://doi.org/10.1038/sdata.2016.18>.

5. Miksa, T., Simms, S., Mietchen, D., Jones, S.: Ten principles for machine-actionable data management plans. *PLOS Computational Biology*. 15, e1006750 (2019). <https://doi.org/10.1371/journal.pcbi.1006750>.
6. Sansone, S.-A., McQuilton, P., Rocca-Serra, P., Gonzalez-Beltran, A., Izzo, M., Lister, A.L., Thurston, M.: FAIRsharing as a community approach to standards, repositories and policies. *Nature Biotechnology*. 37, 358–367 (2019). <https://doi.org/10.1038/s41587-019-0080-8>.
7. Biocuration: Distilling data into knowledge. *PLoS Biol*. 16, (2018). <https://doi.org/10.1371/journal.pbio.2002846>.
8. Hodge, G.: *Systems of Knowledge Organization for Digital Libraries: Beyond Traditional Authority Files*. Digital Library Federation, Washington, USA (2000).
9. Whetzell, P.L., Noy, N.F., Shah, N.H., Alexander, P.R., Nyulas, C., Tudorache, T., Musen, M.A.: BioPortal: enhanced functionality via new Web services from the National Center for Biomedical Ontology to access and use ontologies in software applications. *Nucleic Acids Research*. 39, W541–W545 (2011). <https://doi.org/10.1093/nar/gkr469>.
10. Souza, R.R., Tudhope, D., Almeida, M.B.: Towards a taxonomy of KOS: Dimensions for classifying Knowledge Organization Systems. *Knowledge organization*. 39, 179–192 (2012).
11. Bergman, M.: An Intrepid Guide to Ontologies, <https://www.mkbergman.com/374/an-intrepid-guide-to-ontologies>, last accessed 2020/05/23.
12. Daniel-Le Bozec, C., Steichen, O., Dart, T., Jaulent, M.-C.: The role of local terminologies in electronic health records. The HEGP experience. *Studies in Health Technology and Informatics*. 129, 780–784 (2007).
13. Raboudi, A., Allanic, M., Hervé, P.-Y., Balvay, D., Durupt, A., Boutinaud, P., Tavitian, B., Eynard, B.: Implementation of a Product Lifecycle Management framework for Biomedical Research. In: *Proceedings of the IFIP International Conference on Product Lifecycle Management 11-14 July 2021*. , Curitiba, Brazil (2021).
14. Allanic, M., Hervé, P.-Y., Pham, C.-C., Lekkal, M., Durupt, A., Brial, T., Grioche, A., Matta, N., Boutinaud, P., Eynard, B., Joliot, M.: BIOMIST: A Platform for Biomedical Data Lifecycle Management of Neuroimaging Cohorts. *Frontiers in ICT*. 3, (2017). <https://doi.org/10.3389/fict.2016.00035>.
15. Simmhan, Y.L., Plale, B., Gannon, D.: A survey of data provenance in e-science. *Association for Computing (ACM) Machinery Special Interest Group On Management of Data (SIGMOD) Record*. 34, 31 (2005). <https://doi.org/10.1145/1084805.1084812>.
16. Ciccamese, P., Soiland-Reyes, S., Belhajjame, K., Gray, A.J., Goble, C., Clark, T.: PAV ontology: provenance, authoring and versioning. *Journal of Biomedical Semantics*. 4, 37 (2013). <https://doi.org/10.1186/2041-1480-4-37>.
17. Garde, S., Knaup, P., Hovenga, E.J.S., Heard, S.: Towards Semantic Interoperability for Electronic Health Records. *Methods Inf Med*. 46, 332–343 (2007). <https://doi.org/10.1160/ME5001>.
18. He, Y., Xiang, Z., Zheng, J., Lin, Y., Overton, J.A., Ong, E.: The eXtensible ontology development (XOD) principles and tool implementation to support ontology interoperability. *Journal of Biomedical Semantics*. 9, 3 (2018). <https://doi.org/10.1186/s13326-017-0169-2>.
19. Oliveira, D., Pesquita, C.: Improving the interoperability of biomedical ontologies with compound alignments. *Journal of Biomedical Semantics*. 9, 1 (2018). <https://doi.org/10.1186/s13326-017-0171-8>.
20. Martínez-Romero, M., O’Connor, M.J., Dorf, M., Vendetti, J., Willrett, D., Egyedi, A.L., Graybeal, J., Musen, M.A.: Supporting Ontology-Based Standardization of Biomedical Metadata in the CEDAR Workbench. In: Warrender, J.D. and Lord, P. (eds.) *Proceedings of the 8th International Conference on Biomedical Ontology ICBO 2017*. pp. 1–6. CEUR-WP, Newcastle-upon-Tyne, UK (2017).
21. Patel, M., Koch, T., Doerr, M., Tsinarakis, C., Gioldasis, N., Golub, K., Tudhope, D.: Semantic interoperability in digital library systems. *DELOS Network of Excellence on Digital Libraries and UKOLN University of Bath, Bath, UK, University of Bath, Bath, UK* (2005).
22. Studer, R., Benjamins, V.R., Fensel, D.: Knowledge engineering: Principles and methods. *Data & Knowledge Engineering*. 25, 161–197 (1998). [https://doi.org/10.1016/S0169-023X\(97\)00056-6](https://doi.org/10.1016/S0169-023X(97)00056-6).

23. Gangemi, A., Guarino, N., Masolo, C., Oltramari, A., Schneider, L.: Sweetening Ontologies with DOLCE. In: Gómez-Pérez, A. and Benjamins, V.R. (eds.) *Knowledge Engineering and Knowledge Management: Ontologies and the Semantic Web*. pp. 166–181. Springer Berlin Heidelberg, Berlin, Heidelberg (2002). https://doi.org/10.1007/3-540-45810-7_18.
24. Poline, J.-B., Breeze, J.L., Ghosh, S.S., Gorgolewski, K., Halchenko, Y.O., Hanke, M., Helmer, K.G., Marcus, D.S., Poldrack, R.A., Schwartz, Y., Ashburner, J., Kennedy, D.N.: Data sharing in neuroimaging research. *Front. Neuroinform.* 6, (2012). <https://doi.org/10.3389/fninf.2012.00009>.
25. Roussey, C., Pinet, F., Kang, M.A., Corcho, O.: An Introduction to Ontologies and Ontology Engineering. In: Falquet, G., Métral, C., Teller, J., and Tweed, C. (eds.) *Ontologies in Urban Development Projects*. pp. 9–38. Springer London, London (2011). https://doi.org/10.1007/978-0-85729-724-2_2.
26. Masolo, C., Borgo, S., Gangemi, A., Guarino, N., Oltramari, A.: WonderWeb deliverable D18 ontology library (final). Laboratory For Applied Ontology and ISTC-CNR, Trento, Italy (2003).
27. Smith, B.P., Kumar, A., Bittner, T.: Basic Formal Ontology for bioinformatics. IFOMIS Reports (2005).
28. Smith, B., Ashburner, M., Rosse, C., Bard, J., Bug, W., Ceusters, W., Goldberg, L.J., Eilbeck, K., Ireland, A., Mungall, C.J., Leontis, N., Rocca-Serra, P., Ruttenberg, A., Sansone, S.-A., Scheuermann, R.H., Shah, N., Whetzel, P.L., Lewis, S.: The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration. *Nature Biotechnology.* 25, 1251–1255 (2007). <https://doi.org/10.1038/nbt1346>.
29. Chakravarthy, A., Beales, R., Matskanis, N., Yang, X.: OntoFilm: A Core Ontology for Film Production. In: Chua, T.S., Kompatsiaris, Y., Merialdo, B., Haas, W., Thallinger, G., and Bailer, W. (eds.) *International Conference on Semantic and Digital Media Technologies*. pp. 177–181. Springer Berlin Heidelberg, Berlin (2009).
30. Pedrinaci, C., Domingue, J., de Medeiros, A.K.A.: A core ontology for Business Process Analysis. In: Bechhofer, S., Hauswirth, M., Hoffmann, J., and Koubarakis, M. (eds.) *European Semantic Web Conference*. pp. 49–64. Springer Berlin Heidelberg, Berlin (2008).
31. Gruber, T.R.: Toward principles for the design of ontologies used for knowledge sharing? *International Journal of Human-Computer Studies.* 43, 907–928 (1995). <https://doi.org/10.1006/ijhc.1995.1081>.
32. DICOM Standard, <https://www.dicomstandard.org/>, last accessed 2020/03/16.
33. Langlotz, C.P.: RadLex: A New Method for Indexing Online Educational Materials. *RadioGraphics.* 26, 1595–1597 (2006). <https://doi.org/10.1148/rg.266065168>.
34. Buckler, A.J., Ouellette, M., Danagoulian, J., Wernsing, G., Liu, T.T., Savig, E., Suzek, B.E., Rubin, D.L., Paik, D.: Quantitative Imaging Biomarker Ontology (QIBO) for Knowledge Representation of Biomedical Imaging Biomarkers. *Journal of Digital Imaging.* 26, 630–641 (2013). <https://doi.org/10.1007/s10278-013-9599-2>.
35. Bandrowski, A., Brinkman, R., Brochhausen, M., Brush, M.H., Bug, B., Chibucos, M.C., Clancy, K., Courtot, M., Derom, D., Dumontier, M., Fan, L., Fostel, J., Fragoso, G., Gibson, F., Gonzalez-Beltran, A., Haendel, M.A., He, Y., Heiskanen, M., Hernandez-Boussard, T., Jensen, M., Lin, Y., Lister, A.L., Lord, P., Malone, J., Manduchi, E., McGee, M., Morrison, N., Overton, J.A., Parkinson, H., Peters, B., Rocca-Serra, P., Ruttenberg, A., Sansone, S.-A., Scheuermann, R.H., Schober, D., Smith, B., Soldatova, L.N., Stoeckert, C.J., Taylor, C.F., Torniai, C., Turner, J.A., Vita, R., Whetzel, P.L., Zheng, J.: The Ontology for Biomedical Investigations. *PLoS One.* 11, (2016). <https://doi.org/10.1371/journal.pone.0154556>.
36. Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T., Harris, M.A., Hill, D.P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J.C., Richardson, J.E., Ringwald, M., Rubin, G.M., Sherlock, G.: Gene Ontology: tool for the unification of biology. *Nat Genet.* 25, 25–29 (2000). <https://doi.org/10.1038/75556>.
37. Gibaud, B., Forestier, G., Benoit-Cattin, H., Cervenansky, F., Clarysse, P., Friboulet, D., Gaignard, A., Hugonnard, P., Lartizien, C., Liebgott, H., Montagnat, J., Tabary, J., Glatard, T.: OntoVIP: An

- ontology for the annotation of object models used for medical image simulation. *Journal of Biomedical Informatics*. 52, 279–292 (2014). <https://doi.org/10.1016/j.jbi.2014.07.008>.
38. Hunter, J.: Enhancing the semantic interoperability of multimedia through a core ontology. *IEEE Trans. Circuits Syst. Video Technol.* 13, 49–58 (2003). <https://doi.org/10.1109/TCSVT.2002.808088>.
 39. Tsinaraki, C., Polydoros, P., Christodoulakis, S.: Interoperability Support between MPEG-7/21 and OWL in DS-MIRF. *IEEE Transactions on Knowledge and Data Engineering*. 19, 219–232 (2007). <https://doi.org/10.1109/TKDE.2007.33>.
 40. Seyed, A.P.: BFO/DOLCE Primitive Relation Comparison. *Nat Prec.* 1–1 (2009). <https://doi.org/10.1038/npre.2009.3481.1>.
 41. Lebo, T., Sahoo, S., McGuinness, D., Belhajjame, K., Cheney, J., Corsar, D., Garijo, D., Soiland-Reyes, S., Zednik, S., Zhao, J.: Prov-o: The prov ontology, <http://www.w3.org/TR/2013/REC-prov-o-20130430/>.
 42. Belhajjame, K., B'Far, R., Cheney, J., Coppens, S., Cresswell, S., Gil, Y., Groth, P., Klyne, G., Lebo, T., McCusker, J., others: Prov-dm: The prov data model, <http://www.w3.org/TR/2013/REC-prov-dm-20130430/>.
 43. Sahoo, S.S., Sheth, A., Henson, C.: Semantic provenance for eScience - Managing the deluge of scientific data. *IEEE Internet Computing*. 12, 46–54 (2008). <https://doi.org/10.1109/MIC.2008.86>.
 44. Ding, L., Bao, J., Michaelis, J.R., Zhao, J., McGuinness, D.L.: Reflections on Provenance Ontology Encodings. In: McGuinness, D.L., Michaelis, J.R., and Moreau, L. (eds.) *Provenance and Annotation of Data and Processes*. pp. 198–205. Springer Berlin Heidelberg, Troy, MI, USA (2010). https://doi.org/10.1007/978-3-642-17819-1_22.
 45. Brinkman, R.R., Courtot, M., Derom, D., Fostel, J.M., He, Y., Lord, P., Malone, J., Parkinson, H., Peters, B., Rocca-Serra, P., Ruttenberg, A., Sansone, S.-A., Soldatova, L.N., Stoeckert, C.J., Turner, J.A., Zheng, J., the OBI consortium: Modeling biomedical experimental processes with OBI. *Journal of Biomedical Semantics*. 1, S7 (2010). <https://doi.org/10.1186/2041-1480-1-S1-S7>.
 46. SKOS Simple Knowledge Organization System Reference, <https://www.w3.org/TR/2009/REC-skos-reference-20090818/>, last accessed 2020/05/12.
 47. Courtot, M., Gibson, F., Lister, A.L., Malone, J., Schober, D., Brinkman, R.R., Ruttenberg, A.: MIREOT: the Minimum Information to Reference an External Ontology Term. In: *Nature Precedings*. , Buffalo, NY (2009). <https://doi.org/10.1038/npre.2009.3574.1>.
 48. Mungall, C.J., Torniai, C., Gkoutos, G.V., Lewis, S.E., Haendel, M.A.: Uberon, an integrative multi-species anatomy ontology. *Genome Biology*. 13, R5 (2012). <https://doi.org/10.1186/gb-2012-13-1-r5>.
 49. Arp, R., Smith, B., Spear, A.D.: *Building Ontologies with Basic Formal Ontology*. MIT Press, Cambridge, MA, USA (2015).
 50. OBO principles, <http://www.obofoundry.org/principles/fp-001-open.html>, last accessed 2020/04/09.
 51. Brahaj, A., Razum, M., Schwichtenberg, F.: Ontological Formalization of Scientific Experiments Based on Core Scientific Metadata Model. In: Zaphiris, P., Buchanan, G., Rasmussen, E., and Loizides, F. (eds.) *Theory and Practice of Digital Libraries*. pp. 273–279. Springer Berlin Heidelberg, Berlin, Heidelberg (2012). https://doi.org/10.1007/978-3-642-33290-6_29.
 52. Rocca-Serra, P., Brandizi, M., Maguire, E., Sklyar, N., Taylor, C., Begley, K., Field, D., Harris, S., Hide, W., Hofmann, O., Neumann, S., Sterk, P., Tong, W., Sansone, S.-A.: ISA software suite: supporting standards-compliant experimental annotation and enabling curation at the community level. *Bioinformatics*. 26, 2354–2356 (2010). <https://doi.org/10.1093/bioinformatics/btq415>.
 53. González-Beltrán, A., Maguire, E., Sansone, S.-A., Rocca-Serra, P.: linkedISA: semantic representation of ISA-Tab experimental metadata. *BMC Bioinformatics*. 15, S4 (2014). <https://doi.org/10.1186/1471-2105-15-S14-S4>.
 54. Soldatova, L.N., Aubrey, W., King, R.D., Clare, A.: The EXACT description of biomedical protocols. *Bioinformatics*. 24, i295–i303 (2008). <https://doi.org/10.1093/bioinformatics/btn156>.

55. Soldatova, L.N., Nadis, D., King, R.D., Basu, P.S., Haddi, E., Baumle, V., Saunders, N.J., Marwan, W., Rudkin, B.B.: EXACT2: the semantics of biomedical protocols. *BMC Bioinformatics*. 15, S5 (2014). <https://doi.org/10.1186/1471-2105-15-S14-S5>.
56. Giraldo, O., García, A., López, F., Corcho, O.: Using semantics for representing experimental protocols. *Journal of Biomedical Semantics*. 8, 52 (2017). <https://doi.org/10.1186/s13326-017-0160-y>.
57. Aloulen, Z., Belhajjame, K., Grigori, D., Acker, R.: A Domain-Independent Ontology for Capturing Scientific Experiments. In: Kotzinos, D., Laurent, D., Spyratos, N., Tanaka, Y., and Taniguchi, R. (eds.) *Information Search, Integration, and Personalization*. pp. 53–68. Springer, Cham, Cham (2019). https://doi.org/10.1007/978-3-030-30284-9_4.