

# The use of Foundational Ontologies in Bioinformatics\*

César Henrique Bernabé<sup>1</sup>[0000-0003-1795-5930], Núria Queralt-Rosinach<sup>1</sup>[0000-0003-0169-8159], Vítor E. Silva Souza<sup>3</sup>[0000-0003-1869-5704], Luiz Olavo Bonino da Silva Santos<sup>1,2</sup>[0000-0002-1164-1351], Annika Jacobsen<sup>1</sup>[0000-0003-4818-2360], Barend Mons<sup>1</sup>[0000-0003-3934-0072], and Marco Roos<sup>1</sup>[0000-0002-8691-772X]

<sup>1</sup> Leiden University Medical Center, Leiden, The Netherlands  
{c.h.bernabe, n.queralt\_rosinach, a.jacobsen, b.mons, m.roos}@lumc.nl

<sup>2</sup> University of Twente, Enschede, The Netherlands

l.o.boninodasilvasantos@utwente.nl

<sup>3</sup> Federal University of Espírito Santo

vitor.souza@ufes.br

**Abstract.** Ontologies have been used in biomedicine for several purposes, such as knowledge representation, data analysis and integration. The FAIR principles recommend the use of controlled vocabularies, such as ontologies, to define data terms precisely. However, ontologies are currently modelled following different approaches, sometimes defining overlapping concepts with conflicting definitions, which can harm data FAIRness. Foundational ontologies are defined as domain-independent ontologies that describe the most basic concepts of our world, and thus used to lessen the interoperability problem. With the aim to investigate how foundational ontologies can improve the benefits of FAIR data for life sciences research, we observed a need to assess how they are used in the area. To support our investigation, we conducted a systematic literature analysis, in which we selected appropriate works according to predefined criteria. From the selected works, we identified that, besides being used for several purposes, there is almost no empirical evidence testing the claims for or against the use of foundational ontologies. This indicates a need for evaluating the use of such artefacts in bioinformatics. Similarly, we observed a low adherence to formal ontology construction and evaluation methods during ontologies development, which can impact the quality and sustainability of ontologies, and thus the FAIRness of ontologized data.

**Keywords:** Systematic Literature Mapping · Foundational Ontologies · FAIR · Biomedical Ontologies · Ontology Engineering · Ontology Evaluation · Ontology Application.

---

\* This initiative has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N°825575 and the Trusted World of Corona (TWOC; LSH Health Holland).

## 1 Introduction

The biomedical field is facing an increasing growth in the volume of research data, which is impossible to analyse by human agents alone. To cope with this, several approaches have been proposed to make data and metadata (i.e. description of data) machine-readable and -actionable, such as to enable computers to understand and automatically process it. The FAIR principles [25] aim to make data Findable, Accessible, Interoperable and Reusable for both humans and machines, thus enabling efficient data analysis across multiple resources with minimal human intervention. The FAIR principles recommend the use of controlled vocabularies (e.g., ontologies) in the description of metadata and data terms, seeking to achieve an unambiguous understanding of concepts and properties.

In life sciences research, ontologies are used to support knowledge management and to improve data analysis, shareability and interoperability [14]. However, ontologies are currently modelled in different paradigms, sometimes describing overlapping concepts with conflicting definitions. Hence, there is a risk for the interoperability problem to happen to ontologies just as it does to data [13]. Moreover, misalignment between ontologies can impact the reliability of data analysis results due to the misunderstanding of terms, and also impact the (meta)data adherence to FAIR principles (FAIRness) [8].

Foundational ontologies are one of the approaches proposed to improve ontology interoperability, since they provide a general and domain-independent conceptual architecture that describes the most basic kinds of entities and relationships. Consequently, foundational ontologies support a more precise descriptions of the concepts and relations in a given universe of discourse [8].

We investigate how foundational ontologies are used in bioinformatics, what advantages they claim to have, and if these are validated for the biomedical domain, as well as what the gaps and opportunities are in this particular area of research. Our approach is based on a Systematic Literature Mapping (SLM), which is a method to analyse the state of the art on a particular topic [12]. To our knowledge, no other similar SLM has been conducted on this research question. Our overarching goal is that our findings will guide us to define how foundational ontologies can be used to improve biomedical data FAIRness.

## 2 Ontologies and their Abstraction Levels

Several definitions of the term “ontology” have been proposed in the last decades. Studer, Benjamins & Fensel [23] define ontologies as a “formal, explicit specification of a shared conceptualization”. ‘Formal’ means that the model is logically defined so it supports algorithmic reasoning. ‘Explicit’ refers to concepts being defined with unambiguous descriptions. Finally, ‘shared conceptualization’ refers to the consensual definition of domain concepts within the community of expected users.

Ontologies can be modelled in different levels of detail, and classified accordingly [7]. Application ontologies are built to address a specific use case, usually

constrained to a particular activity (e.g. orchestrate a machine learning workflow). Domain ontologies describe concepts related to a domain of discourse (e.g., rare diseases). Core ontologies provide an upper-level structural definition to a field that spans across different domains (e.g. biomedicine). Foundational ontologies define very general and domain-independent concepts (e.g., time, event and object). In the literature, foundational ontologies are also defined as “top” or “upper-level” ontologies, while core ontologies can also be described as “domain upper ontologies” or “middle-level ontologies”.

The Basic Formal Ontology (BFO) [5], the Descriptive Ontology for Linguistic and Cognitive Engineering (DOLCE) [15], the General Formal Ontology (GFO) [9] and the Suggested Upper Merged Ontology (SUMO) [16] are examples of foundational ontologies, while Biotop [2] is an example of core ontology. Currently, most biological and biomedical ontologies are registered in repositories such as the NCBO Bioportal [18] and the Open Biological and Biomedical Ontologies (OBO) Foundry [22].

### 3 Systematic Literature Mapping Method and Output

The SLM research method used for our literature study is defined by Kitchenham [12] as a secondary study designed to answer broad questions about a research area. Planning, Conducting and Reporting are the main steps of an SLM process, which are further divided into more specific tasks. In the planning step, the research questions, research sources, research query and selection criteria are defined. In the conducting step, papers are extracted from considered sources, deduplicated and selected according to inclusion and exclusion rules. In the reporting step, the results from the SLM are compiled and discussed in the form of answers to research questions.

**Planning.** Based on the literature (e.g., [11]), we observed that, despite being used for several purposes in bioinformatics, there are very few experiments testing the claims and drawbacks of foundational ontologies in the field. Additionally, a second and complementary observation is that there is a lack of methodological rigour in the development and evaluation of ontologies in bioinformatics. Consequently, we defined five research questions in the first task of the planning step. First of all, we would like to know “How are foundational ontologies used in bioinformatics?” (*RQ1*). Secondly, we would like to investigate the reason why foundational ontologies are or are not used, and hence we ask two questions: “What are the claimed advantages of using foundational ontologies?” (*RQ2*) and “What are the claimed drawbacks of using foundational ontologies?” (*RQ3*). Third, since the answers to *RQ2* and *RQ3* are based on perceptions by the extracted papers’ authors, we would like to find scientific support for the answers to the questions by asking “What is the empirical evidence for the advantages and drawbacks?” (*RQ4*). Finally, our second observation could be answered by asking “From the total number of papers that describe the development of a biomedical ontology, how many use existing formal development and evaluation methods?” (*RQ5*).

Based on our research questions, we selected papers that apply a foundational ontology in life science domains (Inclusion Criteria - IC). We excluded papers that did use a foundational ontology but are not related to life sciences (Exclusion Criteria 1 - EC1). For the sake of reproducibility of this study, we also excluded papers not written in English (EC2).

Sources were selected considering the biological and computational aspects of bioinformatics. We included one biosemantics focused source (Jane<sup>4</sup>), one biomedical (Pubmed<sup>5</sup>) and a third that also covers areas from computer science (Science Direct<sup>6</sup>). Finally, the search strategy was driven by the fact that different terms are used to describe foundational ontologies, as mentioned in Section 2. Thus, we included the “top level” and “upper level” synonyms in the search string, which can be generically described as: (“foundational ontology” OR “top-level ontology” OR “top level ontology” OR “upper-level ontology” OR “upper level ontology” OR “upper ontology”) AND (“biology” OR “biomedical” OR “biomedicine” OR “biological”). The specific search strings, the search results (and the date each one was performed), are available in the supplementary material at <https://doi.org/10.5281/zenodo.5793618>.

**Conducting.** In the extraction process, the search string was used in the mentioned sources and applied to the paper’s full text. The search result was downloaded from each source, merged, deduplicated and selected according to the criteria defined. The selection process is performed in three steps, where papers are firstly selected based on information from (i) the title and abstract. Then, the results from the first step are reanalysed, now by performing (ii) diagonal reading (title, abstract, introduction and conclusion sections, figures and tables). In the third step, papers from step two are finally selected/excluded based on (iii) full-text reading.

**Reporting.** In this final step, the analysis conducted on the resulting selection of papers is compiled and reported. During the final iteration of the selection process, the individual answers were manually annotated on a reading sheet and then synthesized in mind maps. Due to space constraints, the mind maps and the information about the extraction process (which criteria was applied to each paper in each phase) are available as supplementary material.

## 4 Data Synthesis

The first step of the extraction process resulted in 404 records, which were then deduplicated, resulting in 346 papers. The final step of the extraction process resulted in 51 papers.

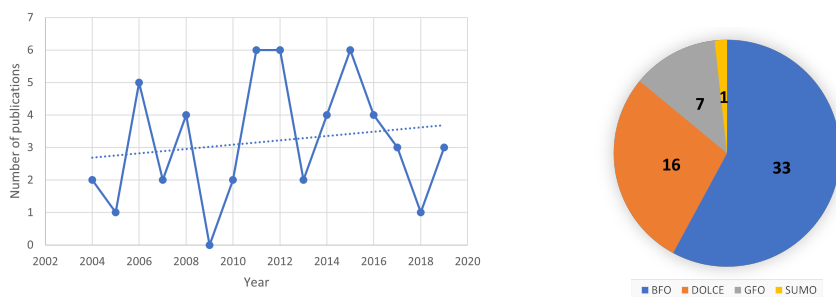
Figure 1a presents the number of selected papers per year of publication. The set comprehend works published from 2004 to 2019, with an average of 2 papers published per year from 2004 to 2010 and 4 papers published per year from 2011 to 2019, with a peak of 6 papers published in 2011, 2012 and 2015. Figure 1b

<sup>4</sup> [jane.biosemantics.org/](http://jane.biosemantics.org/)

<sup>5</sup> [pubmed.ncbi.nlm.nih.gov/](http://pubmed.ncbi.nlm.nih.gov/)

<sup>6</sup> [sciencedirect.com/](http://sciencedirect.com/)

shows the foundational ontologies popularity between the set of selected papers. It can be noticed that BFO is the most used foundational ontology, followed by DOLCE, GFO and SUMO, respectively. Figure 2 compares the use of the foundational ontologies per year (some works used more than one ontology). We consider that a foundational ontology is “used” when it is applied to one of the activities described in the answer to RQ1 (cf. Section 4.1).



(a) Number of publications per year among the set of selected papers in the SLM. (b) Popularity of each foundational ontology among the set of selected papers in the SLM.

Fig. 1: Overview of selected papers: number of publications per year (a) and number of uses of each foundational ontology (b).

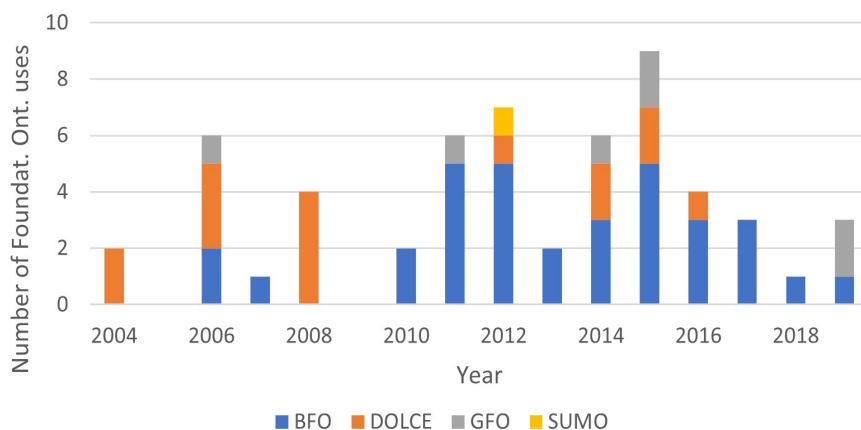


Fig. 2: The number of uses of each foundational ontology (Y axis) per year (X axis) among the set of selected papers in the SLM.

#### 4.1 Synthesized Responses to Research Questions

During data analysis, periodic meetings between co-authors were conducted to discuss the information extraction process in an attempt to reduce possible bias in our synthesis.

Answers to *RQ1* can be classified into five main categories:

- **Ontology development:** in most cases, foundational ontologies were used as a starting point for ontology design, providing a set of basic categories in a top-down development strategy, where classes from the selected foundational ontology are used as a reference for deriving domain concepts. Foundational ontologies were also used in bottom-up (existing domain concepts are anchored in foundational ones), or middle-out (hybrid) approaches. Foundational ontologies also supported the development of ontology design patterns.
- **Ontology enrichment:** the axioms provided by foundational ontologies were added to domain concepts, thus improving inference and supporting the identification of new knowledge.
- **Ontology merging or alignment:** foundational ontologies were used as a common ground for the process of ontology merging, where different domain ontologies are combined to produce a new one. Similarly, foundational ontologies were used to develop mappings between different core/domain ontologies, usually with the aim of interoperability improvement.
- **Ontological analysis:** the ontological commitments (categories, relationships, constraints and axioms) defined by foundational ontologies were used to identify and repair inconsistencies in domain ontologies or other informational artefacts (i.e., information systems, databases, information flow processes or documents). To exemplify, in [19], DOLCE is used to address the polysemy of the term “inflammation”, which can be defined as a physiological function, a characteristic portion of a body part, clinical condition or as a diagnosis applicable to that condition.
- **Ontology based data analysis:** domain ontologies developed using a foundational ontology were used to perform data integration and analysis. Here, the usefulness of ontologies can be noticed in two ways. First, by grounding the data on an ontology, researchers are able to identify errors, organize the data and connect it with external sources. Second, by being efficiently curated and by using reasoning, the ontology-grounded data can undergo a more significant data analysis (e.g., using machine learning algorithms), which can present useful results that support clinical decisions [14].

Answers to the question related to the claimed advantages and motivations (*RQ2*) can be grouped in two main categories: improvement of ontologies and improvement of data. Regarding the former category, foundational ontologies are claimed to improve the semantic understanding of terms and avoid ambiguity (as in the “inflammation” example above), enhance reasoning and prevent errors (by the axioms added by foundational ontologies), speed up ontology development (through the reuse of top-level categories and other ontologies grounded on the same foundational one), improve interoperability (based on the idea that ontologies that use the same foundational ontology tend to interoperate easier), and facilitate ontology maintainability (by reusing categories from foundational ontologies). Advantages related to data are: to enhance data consistency and interoperability (by grounding data into a human-readable semantic artefact),

and improve queriability (ontologized data can be queried using human-readable terms).

A convergence to a small set of similar answers was observed when asking *RQ3*. Several works mentioned the excessive complexity brought up by the use of foundational ontologies, manifested in the time spent understanding class descriptions and the high level of familiarity needed with background philosophical theories. Additionally, the reduced usability in domain ontologies caused by the added complexity from foundational ones is described as another disadvantage.

Unfortunately, just one paper specifically presented an empirical assessment to test the claimed advantages and drawbacks (*RQ4*). Boeker *et al.* [3] conducted a controlled trial to test the hypothesis that “students who received training on top-level ontologies and design patterns perform better than those who only received training in the basic principles of formal ontology engineering.” In the assessment phase, students were asked to solve problems related to different topics, producing a set of ontological models that were compared to a gold standard. However, “the experiment showed no significant effect of the guideline-based training on the performance of ontology developers”.

The last question aims to assess methodological rigour in the process of building and evaluating ontologies (*RQ5*). A subset of 33 of the 51 selected papers developed a new domain ontology (using foundational ontologies). When analysing this subset, we investigated how papers addressed ontology engineering [1] and evaluation [4]. The information about the methods employed by each paper is available in the supplementary material.

**Ontology Engineering.** Five papers stated that ontologies were built following the OBO principles [22], which suggest the use of BFO as a foundational ontology for ontology development. From the 33 papers that developed ontologies, only four explicitly mentioned the use of an ontology engineering method proposed in the literature (i.e., Ontology Development 101 (OD101) [17] or OntoSpec [10]).

**Ontology Evaluation.** Only two papers used Competency Questions [6] as verification activities. On the other hand, sixteen works applied the developed ontology to use case scenarios as means of validation, more specifically: in data integration experiments, to support the development of an information system, in data classification algorithms, in ontology mapping experiments, and in querying and text mining tasks. Two papers performed instantiation [1] as a validation step and four papers validated the ontologies with domain experts.

## 5 Discussion

The results of our review corroborates our observations, as we could identify only one paper that ran an empirical experiment to test the claims for the use of foundational ontologies. It is important to note that this does not mean that these claimed (dis)advantages are under- or overrated. Actually, it indicates that there is a clear need for evaluating these claims within the bioinformatics domain. In other fields, there already exist several works that performed evaluations in

an attempt to provide empirical evidence for the use of foundational ontologies, and their results might be generalizable to bioinformatics. For instance, Keet [11] conducted an experiment where participants, after receiving training in foundational ontologies, had to choose between developing a “Computer Ontology” from scratch or by reusing an OWL version of DOLCE or BFO. The study concluded that advantages brought up by the use of foundational ontologies make up for the time spent on getting acquainted with them. Verdonck *et al.* [24] ran an experiment to test the differences between traditional conceptual modelling and ontology-driven conceptual modelling, in which foundational ontologies were used. The authors found that very few differences (e.g., number of ambiguities and inconsistencies) are perceived when experimenting with simple and small models. However, significant improvements are perceived when modelling bigger or more sophisticated domains.

We also investigated how ontologies were built and evaluated. As mentioned in the previous section, only 4 of 33 papers used a systematic ontology engineering method. Similarly, only 2 used a formal evaluation method (Competency Questions) despite testing the ontology through application cases. Although the use of foundational ontologies is claimed to impose a certain level of rigour during ontology development and evaluation, these processes need to be supported by additional techniques [1]. We argue that the use of ontology engineering and evaluation methods should be an important concern in the research and development of ontologies in bioinformatics, since these methods guide ontology designers in defining aspects related to sustainability of ontologies (e.g., continuous integration, maintainability, documentation), which consequently impacts the long-term realization of the FAIR principles. In addition, ontology evaluation intends to identify inconsistencies in the developed ontologies, thus improving interoperability.

From the results overview, we observed that BFO’s popularity between the set of selected papers can be related to the OBO principles proposed by the OBO Foundry. These principles require the use of BFO as the standard foundational ontology for the ontologies registered in the repository. We also noticed that the boundaries between foundational and core ontologies are not clearly defined. For instance, while some papers described Biotop as a core ontology for bioinformatics (e.g., [20]), others classified it in the same level as BFO and DOLCE (e.g., [3]). In our work, Biotop was considered a core ontology.

Our findings also identified a need to define the best practices that can improve ontologies for bioinformatics. Simon *et al.* [21] mentions that there are understandable reasons for the *ad hoc* features of many biomedical ontologies. Given the urgency to move from paper based to digital systems, ontologists were forced “to make a series of uninformed decisions about complex ontological issues”, which can be understood in the context of our work as the lack of formal rigour in ontology development. The author also mentions that ontologists have been tempted to seek immediate solutions to particular problems, but such behaviour cannot be accepted in a semantically interoperable world, since “*ad hoc* solutions foster further *ad hoc* problems.”



Finally, we recognize that some studies may not have been considered in our analysis due to two reasons: (i) a paper may have used a foundational ontology without explicitly mentioning it, or (ii) used an ontology that is in a grey area between foundational and core ontologies. These can be seen as a trade-off in using a method such as an SLM, since definitions (e.g., whether a foundational ontology was used or not) should be clearly stated so the process can be systematically repeated. As a next step, we plan to manually include and discuss other important studies (not captured by the SLM, but suggested by experts).

## 6 Conclusion

This paper described an SLM that was conducted to understand how foundational ontologies are used in bioinformatics, as well as to identify the empirical evidence in favour or against claimed advantages. Additionally, we investigated the level of methodological adherence in papers that used foundational ontologies to construct domain ones. The understanding of how foundational ontologies are used in bioinformatics can better drive future research towards the improvement of ontology interoperability, and consequently the FAIRness of research data. Our findings can imply two main conclusions. First, there is a lack of empirical evidence for or against the use of foundational ontologies, and future research should tackle this need. Second, this particular area of bioinformatics should deal with ontology development and evaluation more formally and systematically. Consequently, we recommend that research in bio-ontologies should study the creation or reuse of methods for ontology engineering (considering phases from ontology requirements elicitation to testing and sustainability) and ontology evaluation (encompassing both evaluation techniques and procedures for application-based evaluation) supported by foundational ontologies. Thus, as future work, we plan to investigate how foundational ontologies are used in other fields and what can be reused to improve research in ontologies for bioinformatics.

## References

1. de Almeida Falbo, R.: Sabio: Systematic approach for building ontologies. In: ONTO. COM/ODISE@ FOIS (2014)
2. Beisswanger, E., Schulz, S., Stenzhorn, H., Hahn, U.: Biotop: An upper domain ontology for the life sciences. *Applied Ontology* **3**(4), 205–212 (2008)
3. Boeker, M., Jansen, L., Grewe, N., Röhl, J., Schober, D., Seddig-Raufe, D., Schulz, S.: Effects of guideline-based training on the quality of formal ontologies: A randomized controlled trial. *Plos one* (2013)
4. Brank, J., Grobelnik, M., Mladenic, D.: A survey of ontology evaluation techniques. In: *Proceedings of the conference on data mining and data warehouses (SiKDD 2005)*. Citeseer Ljubljana, Slovenia (2005)
5. Grenon, P., Smith, B.: Snap and span: Towards dynamic spatial ontology. *Spatial cognition and computation* **4**(1), 69–104 (2004)

6. Grüninger, M., Fox, M.S.: The role of competency questions in enterprise engineering. In: *Benchmarking—Theory and practice*. Springer (1995)
7. Guarino, N.: Semantic matching: Formal ontological distinctions for information organization, extraction, and integration. In: *International Summer School on Information Extraction*. Springer (1997)
8. Guizzardi, G.: Ontology, ontologies and the “i” of fair. *Data Intelligence* (2020)
9. Herre, H.: General formal ontology (gfo): A foundational ontology for conceptual modelling. In: *Theory and applications of ontology: computer applications*. Springer (2010)
10. Kassel, G.: Integration of the dolce top-level ontology into the ontospec methodology. *arXiv preprint cs/0510050* (2005)
11. Keet, C.M.: The use of foundational ontologies in ontology development: an empirical assessment. In: *Extended Semantic Web Conference*. Springer (2011)
12. Kitchenham, B., Charters, S.: Guidelines for performing systematic literature reviews in software engineering (2007)
13. Kusnierczyk, W.: Nontological engineering. *Frontiers in Artificial Intelligence and Applications* (2006)
14. Machado, C.M., Rebholz-Schuhmann, D., Freitas, A.T., Couto, F.M.: The semantic web in translational medicine: current applications and future directions. *Briefings in bioinformatics* (2015)
15. Masolo, C., Borgo, S., Gangemi, A., Guarino, N., Oltramari, A.: *Ontology library. wonderweb deliverable d18 (ver. 1.0, 31-12-2003)* (2003)
16. Niles, I., Pease, A.: Towards a standard upper ontology. In: *Proceedings of the international conference on Formal Ontology in Information Systems* (2001)
17. Noy, N.F., McGuinness, D.L., et al.: *Ontology development 101: A guide to creating your first ontology* (2001)
18. Noy, N.F., Shah, N.H., Whetzel, P.L., Dai, B., Dorf, M., Griffith, N., Jonquet, C., Rubin, D.L., Storey, M.A., Chute, C.G., et al.: *Biportal: ontologies and integrated data resources at the click of a mouse. Nucleic acids research* (2009)
19. Pisanelli, D.M., Gangemi, A., Battaglia, M., Catenacci, C.: *Coping with medical polysemy in the semantic web: the role of ontologies. IOS Press* (2004)
20. Schulz, S., Martínez-Costa, C.: *Harmonizing snomed ct with biotoplite: an exercise in principled ontology alignment. IOS Press* (2015)
21. Simon, J., Dos Santos, M., Fielding, J., Smith, B.: *Formal ontology for natural language processing and the integration of biomedical databases. International journal of medical informatics* (2006)
22. Smith, B., Ashburner, M., Rosse, C., Bard, J., Bug, W., Ceusters, W., Goldberg, L.J., Eilbeck, K., Ireland, A., Mungall, C.J., et al.: *The obo foundry: coordinated evolution of ontologies to support biomedical data integration. Nature biotechnology* (2007)
23. Studer, R., Benjamins, V.R., Fensel, D.: *Knowledge engineering: Principles and methods. Data & knowledge engineering* (1998)
24. Verdonck, M., Gailly, F., Pergl, R., Guizzardi, G., Martins, B., Pastor, O.: *Comparing traditional conceptual modeling with ontology-driven conceptual modeling: An empirical study. Information Systems* (2019)
25. 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., et al.: *The fair guiding principles for scientific data management and stewardship. Scientific data* (2016)