



HAL
open science

Version control for interdependent ontologies: challenges and first propositions

Paul Fabry, Adrien Barton, Jean-François Ethier

► To cite this version:

Paul Fabry, Adrien Barton, Jean-François Ethier. Version control for interdependent ontologies: challenges and first propositions. International Conference on Biomedical Ontology (ICBO 2023), Aug 2023, Brasilia, Brazil. à paraître. hal-04301628

HAL Id: hal-04301628

<https://hal.science/hal-04301628>

Submitted on 23 Nov 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.

Version control for interdependent ontologies: challenges and first propositions

Paul Fabry¹, Adrien Barton^{2,1} and Jean-Francois Ethier¹

¹ *Groupe de Recherche Interdisciplinaire en Informatique de la Santé (GRIIS), Université de Sherbrooke, Quebec, Canada*

² *Institut de Recherche en Informatique de Toulouse (IRIT), CNRS, Université de Toulouse, France*

Abstract

The variety and quantity of health data have significantly increased due to healthcare improvements, making data interoperability an increasingly complex issue. Biomedical ontologies are a useful tool to support this interoperability. However, to provide an adequate coverage, ontological models can advantageously leverage a network of ontologies or part of ontologies, from various origins and strongly interdependent. While it fosters interoperability and minimizes duplication, this structure is very sensitive and any evolution of one of its parts requires more and more maintenance efforts. The challenge of tracking changes and assessing their effects is addressed in computer science through version control and while it has been applied to ontologies, applying this approach while importing parts of external ontologies and managing the potential impact of their evolution is a challenging task.

This article focuses on OWL ontologies and two questions regarding their versioning are addressed: what should be versioned in an ontology and how to determine its identity? The current processes rely on a versioning at the level of the ontology, which raises problems that are discussed. Versioning at a more fine-grained level of so-called “OWL components” could mitigate such problems. We identify two kinds of entities as potentially relevant to their identity: the associated cognitive representations and assertive statements; being able to track both separately would be an effective way to mitigate the aforementioned problems.

While further work is warranted to provide an operational definition of “OWL component”, an approach along the lines proposed here can support not only at a better change management through a cohesive, complete and fined-grained version control, but also a better import process while contributing to support ontology engineering methods.

Key words

OWL ontologies, interoperability, version control

1. Introduction

Advances in modern healthcare have significantly increased the variety and quantity of health data generated. Ontologies are well positioned to serve as mediation models to support interoperability between diverse clinical data sources as they provide formal, source-independent representations and do not rely on specific data models that might be tied to specific technologies or formats. One such example is PARS3 [1], a distributed data access platform that is currently being deployed to support activities of the Health Data Research Network Canada [2] and to enable the implementation of a mortality prediction tool in hospitals in Quebec. Implementing these projects implies a series of tasks from ontology development, to pivot relational model generation, to mappings with data sources. While this approach has been demonstrated and validated to enable the support of various tools like ReflexD [3], an audit-feedback tool for diabetes treatment in primary care, the next step is to support the evolution of users’ needs. Science evolves, new requirements emerge, mistakes are identified, ideas are refined, etc. This will often imply changes

in the knowledge representation generated at the inception of the project. This ontological evolution has a significant effect on the activities downstream of ontology development. It is critical to be able to identify what changes constitute a modification that can affect a user's understanding of what is represented or what should be instantiated in a table.

To follow on the questions above, if an ontology changes, how can one identify which of its parts might need to be reviewed for potential adjustments? If the answer is always "all of them", scalability is obviously at risk. What could be the consequences of this change on the applications which depend on it? In the context of PARS3, what are the impacts of a change in the ontology on the mappings between a data source and the pivot relational model? If the answer here again is to ask every source to review or redo every mapping between their assets and the pivot model, maintenance, scalability and as a result, acceptability and sustainability will be markedly reduced.

1.1. Background

This paper focuses on ontologies expressed in OWL2, as it is the most used Description-Logic language for biomedical ontologies. Good engineering principles have been proposed to mitigate the impact of ontology change. The Open Biomedical Ontologies Foundry [4] has proposed guidelines [5] including an emphasis on reusing the existing classes, a clearly defined scope for ontologies to ensure orthogonality of ontologies representing different domains and unique textual definitions following an Aristotelian form [6]. These principles were followed for the development of the ontologies supporting PARS3. A modular approach was chosen for the creation of the ontology model with several domain ontologies addressing specific topics such as drug prescriptions [7], laboratory results [8] or clinical forms [9]. All these ontologies are built upon the upper-level ontology Basic Formal Ontology (BFO) [10] and reuse classes and properties from reference ontologies in the biomedical domain like the Ontology of Biomedical Investigations (OBI) [11]. Finally, these ontologies are linked together via mid-level ontologies describing relevant parent categories pertaining to e.g. health procedures [12] or services [13].

Consequently, the clinical model structure consists of a network of ontologies or part of ontologies, which are either imported or created within our group, and are strongly interdependent. On the one hand it adds great flexibility and allows us to add representations of various domains as needed. In addition, ontology reuse enables us to leverage the domain expertise of other groups. On the other hand this structure is very sensitive and any change in its parts requires more and more maintenance efforts as the model grows and evolves, which may compromise its sustainability in the long run. Interestingly, the challenge of tracking change, evaluating its significance, and identifying the cascading effects is also an issue in computer science and is often addressed through version control.

Version control has already been discussed in the context of ontologies and is even part of the OBO Foundry principles [14] as well as FAIR [15]. However, in this context it consists in adding a version date in the metadata of the ontology. The change of the version date is at the discretion of the authors. Yet, different types of modifications in the ontology, such as the addition or removal of a class or the correction of an obvious typo in annotations such as a comment, can have very different impacts and it would be useful to differentiate between them. This is especially true since a version date refers to the whole ontology and does not in itself reveal which classes have been modified, which is all the more challenging in the case where a user imports only a few from a source ontology that may have several thousands.

The objective of this article is to lay the foundation for a versioning methodology that would allow the safe and efficient identification of what is impacted (or not) by a change in a context of highly interconnected ontologies. The core principles concerning both ontologies and versioning upon which this work is based and the issues at stake will be outlined, and some recommendations

will be proposed and discussed. The resulting methodology and its implementation will be addressed in subsequent works.

2. Versioning challenges for ontologies

Version control as it pertains to computer science encompasses all the tools and methods for tracking and providing control over changes to a given project [16]. This control is performed at a granularity level deemed pertinent to manage the project most efficiently without creating an undue maintenance burden. To achieve this, artefacts of interest that are relevant to the project will be tracked individually. In many software projects, these artefacts are text files parts of the project's source code or documentation for example.

To ensure the tracking of an artefact, it is necessary to endorse a diachronic identity criterion on artefacts that allow us to identify some of them as evolutions of the same entity. An artefact like a computer file is identified (that is, uniquely denoted in a given context) by a resource identifier composed by the location in the file tree and the name of the file. This resource identifier provides a criterion of diachronic identity of the file it refers to, i.e. two files are considered as two states (versions) of the same informational artefact if they are identified by the same resource identifier. From a versioning perspective, even if the content of a file completely changes over time, the different versions will be considered as the same file at different states over time as long as they are identified by the same resource identifier. When trying to apply this methodology to ontologies, challenges emerge.

2.1. Ontology as the artefact of interest for versioning

One option could be to choose the ontology as the artefact of interest for the versioning. This calls for the characterization of what an ontology is. The word “ontology” can refer, depending on the context, to artefacts of various kinds, belonging to the fields of philosophy and computer science, which have as a common goal the representation of entities of the world or a particular domain of it, the categories to which they belong, their properties and the relations that hold between them.

In knowledge engineering, ontologies have been classically defined [17,18] as follows :

An ontology is a formal explicit specification of a shared conceptualization.

A definition grounded in a realist interpretation of the world is proposed by Smith et al. [10]:

Ontology = def. “A representational artefact, comprising a taxonomy as proper part, whose representations are intended to designate some combination of universals, defined classes, and certain relations between them.”

The definition from [10] above assumes the existence of universals independently of the conceptualizations that we can make of the reality. In addition, an important element of Gruber and Guarino's definitions is the notion of shared conceptualization, that is an ontology is meant to be shared, usually between its authors and a community that is using it [19].

Many formalisms have been proposed for ontologies; among them, OWL is a widely used and recognized standard [20]. An OWL ontology is uniquely denoted with an internationalized resource identifier (IRI), which can be used for versioning purposes. For example, OBI is associated with the IRI: “<http://purl.obolibrary.org/obo/obi.owl>” that is specific to this ontology. While the ontology will undergo modifications over time, e.g. by adding and deleting classes or properties, it

will be considered as the same ontology as long as its IRI remains the same. However, using the ontology as the versioning artefact may not be the best suited method.

Firstly, this approach considers an ontology as a monolithic artefact, one that is therefore used as a whole. This is a questionable assumption, especially in a context of interconnected ontologies where sometimes only parts of an ontology might be reused in another. Let's consider for example an ontology A that imports some part of an ontology B. If B changes, should a new version of A be declared even if not a single class declared directly in ontology A has been modified? If a new version of B is proposed, how to know if the classes imported in A are affected? If they are not, should this entail the creation of a new version of A?

Secondly, according to the W3C OWL2 reference documents [20], an ontology is defined as: “[...] a certain kind of computational artefact – i.e., something akin to a program, an XML schema, or a web page – generally presented as a document.” While this allows a document-based approach to versioning, facilitated by tools such as Git [21], this might lead to counterintuitive consequences. For example, two classes with different namespaces located in the same computational artefact would be part of the same ontology according to the W3C definition above; however, in a different conception of ontology that would emphasize namespaces, they would belong to different ontologies.

Thirdly, the size of an ontology may represent a significant challenge for its versioning. The Gene Ontology [22] for example has more than 40 000 classes. If the modification of one of them leads to a new version of the whole ontology, this implies the tedious task for all its users of checking for possible changes in all the classes at each new version. Therefore, using the whole ontology as the most fine-grained artefact of interest for versioning is not an ideal solution.

2.2. Parts of ontology as the artefact of interest for versioning

An ontology proposes a representation of a domain by one or several authors. Such a representation takes place initially in the form of a collection of cognitive representations¹ of its author on the basis of which they create publicly available informational content entities (ICE) [23]. In the context of OWL2 ontologies, such ICEs will be called “OWL component” (OC). At least some of those OCs are associated with an identifier, such as an International Resource Identifier (IRI) that consists of a permanent location, a namespace and a unique string (for example: “http://purl.obolibrary.org/obo/OBI_0000011” is an IRI from OBI), which is of particular interest for versioning. Also at least² some of those OCs refer to a portion of reality that is part of the domain, in a way that can be understood by other users. These can be analyzed using a three level framework inspired by Ceusters and Smith [24]:

- Level L1: a portion of reality;
- Level L2: the cognitive representation of this portion of reality;
- Level L3: a publicly accessible ICE, the OWL component, that emanates³ from the L2 cognitive representation and is also about the same portion of reality.

What is named in the literature as a ‘class’ or ‘term’ is an important type of such OCs. OCs can also include annotations such as definitions, or elucidations. Neuhaus [25] characterizes some of the annotations as “assertive annotations”, i.e. “...the kind of annotations that are intended to assert a true proposition about the domain of the ontology.” Neuhaus also regroup assertive annotations and logical axioms under the term of “assertive statements”. A definition would specify in natural language a way to group certain particulars, while logical axioms can be automatically processed by a reasoner to classify the OWL components.

¹ Cognitive representations are also concretizations of ICEs according to IAO [23].

² “at least” because it remains a complex open question whether all OCs are about a portion of reality (e.g. object properties).

³ The precise nature of the relation of emanation is outside the scope of this work.

An OC can be associated with some logical axioms. For example, a class might be mentioned by some logical axioms.

For example, the OC with the IRI “http://purl.obolibrary.org/obo/OBI_0000011” (that has for label “planned process”) is associated to the following assertive statements (among others):

- Definition: “A process that realizes a plan which is the concretization of a plan specification.”
- *planned process* SubclassOf *process*

Two kinds of relations between a class and a portion of reality are proposed. Firstly, a relation of aboutness that is determined by the cognitive acts of its author: in the IAO framework, inspired by Chisholm [26], the “aboutness of those of our representations formulated in speech or writing [...] is to be understood by reference to the cognitive acts with which they are or can in principle be associated” [23]. Secondly, a relation that is determined by the interpretation of the assertive statements by another competent reader who will make their own cognitive representation of a portion of reality on the basis of this interpretation. In this case, the assertive statements can be considered as describing that portion of reality (we will presume here that any competent external user will have the same interpretation). The goal when creating a class is that its assertive statements describe a portion of reality identical to the one pointed to by its aboutness.

2.3. Identity criteria for ontological components

One may propose two identity criteria of OC based on cognitive representations and assertive statements, depending on whether one relies on what we will call (inspired by theories in philosophy of language [27]) the “intentionalist” or “descriptivist” conceptions of the identity of an OC:

- R₁: In the intentionalist conception, the identity is determined by its author’s related cognitive representation.
- R₂: In the descriptivist conception, the identity is determined by its associated assertive statements.

Hence, in the case of R₁, even though the assertive statements might change significantly, as long as the OC emanates from the same cognitive representation of the author, the identity of the OC does not change and so the IRI should remain the same. Meanwhile in R₂, even though the author’s cognitive representation associated with an OC might change, as long as the assertive statements are not significantly modified⁴, the identity of the OC does not change and so the IRI should remain the same.

The distinction between these conceptions is very rarely made explicit when creating and sharing an ontology, and this may negatively affect its evolution or its external reuse. The cognitive representation of the author (level 2) and the assertive statements are therefore two aspects that need to be managed appropriately to evaluate an OC as a potential artefact of interest for versioning, and as described above, two possible approaches, namely R₁ and R₂, need to be considered.

Let’s consider the following example about a hypothetical ontology about fruits, the Fruit Ontology (FO). At time t₁ the author of this ontology adds an OWL component that refers to its own cognitive representation of the class of oranges (*Citrus sinensis*) in the reality. This OC (let’s call it “OC1”) is associated with the IRI: “FO_004” and is associated with the following assertive statements:

⁴ We are talking about changes that might affect the understanding of the OC’s meaning, not merely a typo fix or a logically equivalent reformulation of a statement.

- Definition D1: “A citrus fruit which is the edible fruit of the orange tree and that is orange in color.”
- *orange* subClassOf *citrus fruit*

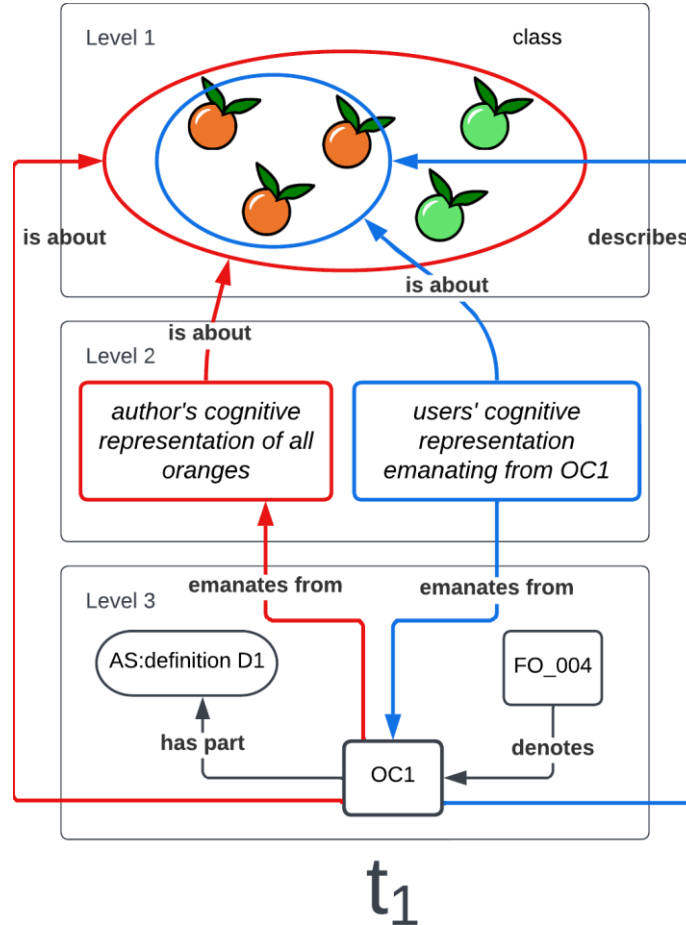


Figure 1: Relations between the OWL component (OC1) from the Fruit Ontology, the portion of reality it is about, and the associated author's and users' cognitive representations at time t_1 .

At time t_2 , the author is made aware that the definition D1 does not explicitly mention unripe oranges, which are green, and therefore that some of the ontology's users assume OC1 is about only ripe oranges (Figure 1).

The author is now faced with two options: either **A**) he considers that his initial cognitive representation is still the relevant one and so the assertive statements need to be updated accordingly, or **B**) he considers that what needs to be represented is instead ripe oranges and so he is now working with a new cognitive representation but as it turns out, the assertive statements derived from this new representation are similar to those initially derived.

Each of these options may have an impact on the identity of OC1 that must be evaluated according to the **R1** and **R2** conceptions discussed above, we thus have four distinct scenarios:

- Scenario **R1A**: The cognitive representation of the class of oranges is not modified between t_1 and t_2 and a new definition D2 (“A citrus fruit which is the edible fruit of the orange tree and that is orange in color when ripe, and green when unripe.”) is proposed at t_2 for clarification. In this scenario, OC1 is still conformant to the same

cognitive representation, but one of the associated assertive statements has been modified. As the identity of OC1 is determined by the cognitive representation in R1 and this representation didn't change, what comes out of this scenario is a new version of OC1, which is thus associated with the same IRI "FO_004". (Figure 2).

- Scenario **R1B**: A new cognitive representation of interest (about ripe oranges) has been identified by the author of OC1. Therefore, a new OWL component (OC2) and thus a new IRI, is created by reusing assertive statements of OC1 (in this case the definition). (Figure 2).

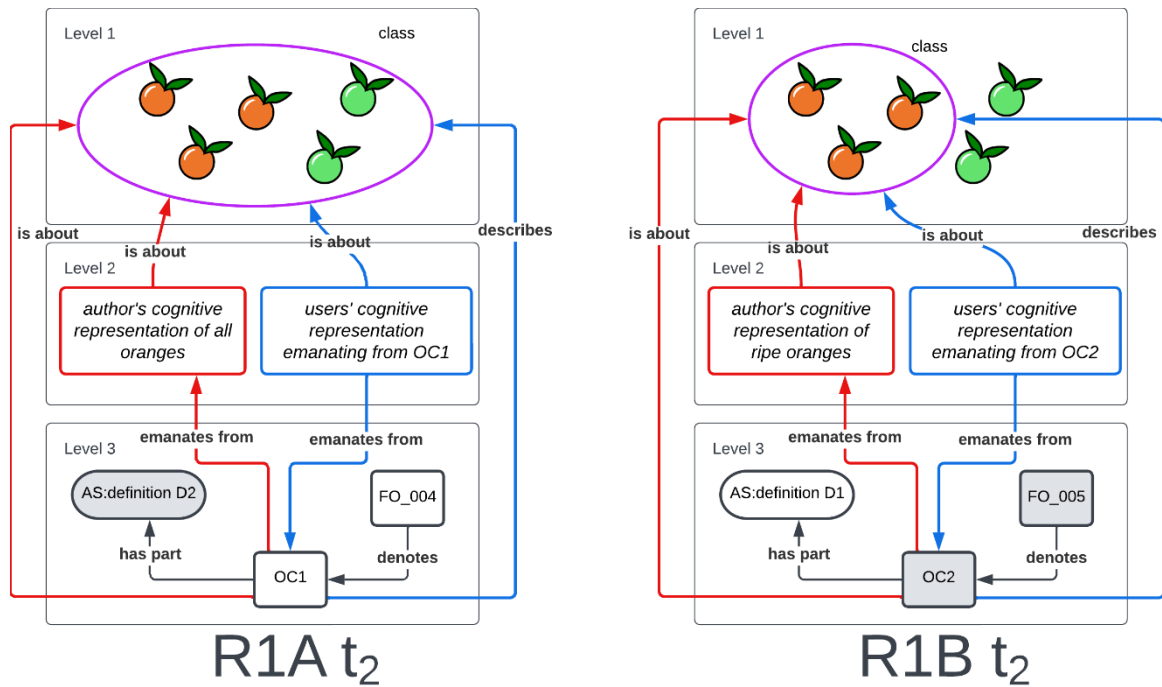


Figure 21: Scenarios R1A and R1B at t_2 . In R1A the cognitive representation of the class of oranges remains unchanged between t_1 and t_2 and a new definition is proposed at t_2 for clarification. In R1B a new cognitive representation of interest has been identified by the author and therefore, a new OWL component (OC2) is created.

- Scenario **R2A**: R2A describes the same reality as R1A but endorses the R2 conception rather than the R1 conception. A new definition D2 has been formulated at t_2 , and consequently the assertive statements have changed. In this case, a new OC (OC3), and thus a new IRI, is created along the new assertive statements (Figure 3).
- Scenario **R2B**: Here also, R2B describes the same reality as R1B, but endorses the R2 conception rather than the R1 conception. The assertive statements have not been modified and thus the identity of OC1 has not changed even though OC1 is now related to another cognitive representation (Figure 3).

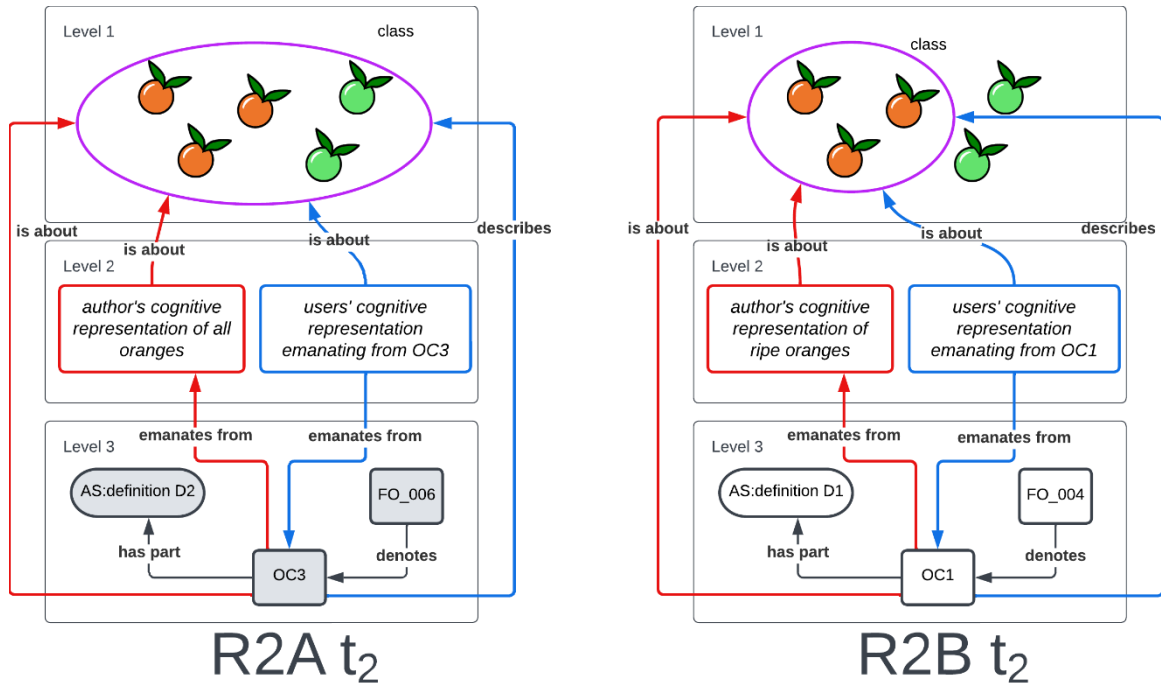


Figure 3: Scenarios R2A and R2B at t_2 . In R2A, the assertive statements have changed, thus a new OC and a new IRI are created along them. In R2B, the assertive statements have not changed. OC's identity has not changed even though it is associated with a new cognitive representation of the author.

Let's consider that OC1 is imported at time t_1 in two fictive distinct ontologies:

- The Plant Parasites ontology (PPO), created by the author of OC1, imported it in the axiom: *medfly* subClassOf (is_parasite_of some *orange*). As a matter of fact, the author of PPO considered that FO_004 was about oranges at any stage of development.
- The Juice Ontology (JO) imported it in the axiom: *orange juice* subClassOf (is_made_of some *orange*). The authors of JO have initially imported "FO_004" as they understood the assertive statements as describing ripe oranges.

For these two ontologies, the previously mentioned scenarios will have various implications. For example, in the scenario R1A, as OC1 is now about oranges at any stage of development, JO needs to modify its axiom and include another OC while PPO doesn't need to. On the contrary, in the scenario R2B, as OC1 is now about ripe orange, JO could keep the axiom as is while PPO needs to modify it. In context of a direct import, this will lead to incompatible OCs that might be difficult to identify, and the "live" change might affect users before the problem can be identified and fixed.

These scenarios are found in practice without being explicitly identified as such, which leads to many ambiguities when reusing ontologies. However, each conception has its advantages and disadvantages depending on whether one favors the author's cognitive representation importance in contributing to the coherence of the OC over the users' perspective which is based on their understanding of the assertive statements.

The examples above illustrate that OCs can be good candidates as artefacts of interest for versioning, provided that we are able to track both the cognitive representation of the author (via the operationalization of R1) and the assertive statements (via the operationalization of R2).

3. Discussion

Despite emerging tools that provide an invaluable help [28], managing the evolution of interconnected ontologies is a tedious and error-prone task. Versioning processes are a way forward to support the management of ontology evolution. In this work, two issues at the heart of this question are addressed.

3.1. Artefact of interest for ontology versioning

The first issue concerns the artefacts of interest, whose evolution needs to be tracked individually. As mentioned above, the current processes imply a versioning at the level of the ontology, but this presents some important problems, if only because it is not simple to outline the target of versioning in a context of interconnected modular ontologies where sometimes only parts of an ontology might be reused in another. Therefore, ontologies should no longer be considered as standalone and monolithic but rather as highly interconnected informational constructs, and a focus on a more fine-grained artefact is required, leading to the introduction of so-called “OWL components” as the versioned artefacts.

In the current work, we only discussed one kind of OC, namely classes. However, other components of an ontology such as object properties, data properties or even ontological instances are possible candidates as OCs of interest for versioning. Indeed, all these elements can be associated with assertive statements that can change over time. Defining OCs in the least ambiguous and most comprehensive way possible is of the utmost importance and will be the subject of subsequent work. In addition, not all parts of an ontology are of interest for versioning and therefore it will be important to determine whether certain OCs are not relevant for versioning.

It seems also relevant to consider what an ontology is in relation to OCs. On the one hand, ontologies are conceptually envisioned through high-level definitions such as the one previously mentioned [10,18]. On the other hand, ontologies as defined by W3C are described as computerized constructs and manipulated via files or namespaces. As it has been discussed in this work, the relationship between the two is not so clear cut and this calls for approaching OWL ontologies according to principles more in line with those of software engineering.

3.2. Identity of the artefact of interest

The next issue is the determination of the identity of the artefact of interest. Even though the OCs relevant for versioning have only been characterized through their role for versioning, two possibly relevant aspects for their identity have been identified: their cognitive representations and their associated assertive statements (logical axioms and assertive annotations).

While the formal aspect of an ontology allows the use of reasoners and query tools to automatically process an ontology, an ontology is not limited to its formal elements. An ontology is also a way to share knowledge between communities of humans and for this goal, assertive annotations such as natural language definitions may have a role as significant as the logical axioms [19].

However, not all annotations are assertive and one needs to distinguish assertive annotations from annotations concerning the OC itself, such as its author or the date of creation. An annotation that will require additional exploration to arbitrate its inclusion or not as an assertive statement is the label of an OC. Given the uncertainty around its status, it is not included in the examples above. A natural language label in itself may be too ambiguous for playing an assertive role: experience shows that a given label could be associated with several distinct OC from distinct ontologies.

However, labels are not randomly chosen either and the eventual import that the semantics of natural language labels should have into ontologies will need to be analyzed.

In an ideal world, one could envision a one-to-one relation between cognitive representations and assertive statements, as the cognitive representation of the author of an OC would be exactly translated in its assertive statements, which in turn would be exactly interpreted in a similar cognitive representation by the users that have only access to its assertive statements. However, OCs may retain the same IRI while their assertive statements undergo significant changes through time. Despite everyone's best effort, discrepancies exist and will likely remain and therefore the chosen versioning process must be able to handle these situations.

However, regardless of the quality of the assertive statements, an evolution of knowledge may imply a modification of the cognitive representation associated with an OC. Since OCs are developed in interlinked coherent groups (ontologies), the cognitive representation of the author is likely to play an important role in keeping this group coherent. Being able to track the cognitive representation of the author would allow the user community to know when something fundamental changed in the way the author structures the knowledge representation (e.g. because of a shift in science) versus when they attempt to improve the assertive statements to better match their cognitive representation.

To this end, being able to individually track the changes of both the cognitive representations and assertive statements would be a significant advantage. As it stands though only one IRI is associated with an OC and it is sometimes used to track the author's cognitive representation (R1) and sometimes the assertive statements (R2). However, it cannot do both in parallel. Associating two IRIs, one that tracks OC identity according to conception R1, and another one that tracks OC identity according to conception R2, would allow the tracking of both the cognitive representation and the assertive statements. Further work should elaborate the details of such a proposition.

4. Conclusion and Future Work

The work presented here aims at laying the foundation for a safe handling of changes to OWL components when they are reused. Being able to track changes in the cognitive representations of the authors or the assertive statements is the first step to further scalability and sustainability while diminishing the incentive to build "yet another model".

Other challenges lie ahead. It will likely be desirable to identify various types of changes without semantic implications including:

1. Changes to the computational representations e.g. changes of OWL syntax or in the order of elements;
2. Changes in non-assertive annotations e.g. additions of contact information annotations;
3. Changes in natural language assertive statements e.g. correcting a typographical error or replacing a word by a perfect synonym;
4. Changes in logical axioms producing a logically equivalent result e.g. replacing "OC_A OR OC_B" by "OC_B OR OC_A".

Automatically detecting changes of type 3 or 4 is not trivial in all cases.

Finally, when the changes include the cognitive representation of the author or assertive statements, it will be greatly beneficial to be able to have an approach to navigate through OCs of interest and identify those not impacted by the changes, leaving only a subset for manual review and adjustments as needed.

Overall, an approach based on the principles presented here will ensure not only a better change management through a cohesive, complete, and fine-grained version control, but also a better import process while contributing to supporting ontology engineering methods.

5. References

- [1] PARS3 • Solutions • GRIIS. GRIIS n.d. <https://griis.ca/en/solutions/pars3/> (accessed June 16, 2023).
- [2] McGrail K, Diverty B, Lix L. Introducing Health Data Research Network Canada (HDRN Canada): A New Organization to Advance Canadian And International Population Data Science. *IJPDS* 2020;5. <https://doi.org/10.23889/ijpds.v5i5.1493>.
- [3] ReflexD • Primary health care and follow-up for people with diabetes • Solutions • GRIIS. GRIIS n.d. <https://griis.ca/en/solutions/reflexd/> (accessed June 16, 2023).
- [4] Smith B, Ashburner M, Rosse C, Bard J, Bug W, Ceusters W, et al. The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration. *Nat Biotechnol* 2007;25:1251. <https://doi.org/10.1038/nbt1346>.
- [5] OBO Foundry - Principles overview n.d. <http://obofoundry.org/principles/fp-000-summary.html> (accessed June 16, 2023).
- [6] Seppälä S, Ruttenberg A, Smith B. The Functions of Definitions in Ontologies. *Formal Ontology in Information Systems* 2016:37–50. <https://doi.org/10.3233/978-1-61499-660-6-37>.
- [7] Ethier J-F, Goyer F, Fabry P, Barton A. The prescription of drug ontology 2.0 (PDRO): More than the sum of its parts. *International Journal of Environmental Research and Public Health* 2021;18. <https://doi.org/10.3390/ijerph182212025>.
- [8] Barton A, Fabry P, Lavoie L, Ethier J-F. LABO: An ontology for laboratory test prescription and reporting. 2019 Joint Ontology Workshops Episode V: The Styrian Autumn of Ontology, JOWO 2019, vol. 2518, Graz; Austria: CEUR-WS.org; 2019.
- [9] Fabry P, Barton A, Ethier J-F. QUESTO – An ontology for questionnaire. *ICBO|ODLS 2020 International Conference on Biomedical Ontologies 2020*, vol. 2807, Bolzano, Italy: CEUR-WS.org; 2020, p. B.1-12.
- [10] Arp R, Smith B, Spear AD. Building ontologies with Basic Formal Ontology. Cambridge, Massachusetts: Massachusetts Institute of Technology; 2015.
- [11] Bandrowski A, Brinkman R, Brochhausen M, Brush MH, Bug B, Chibucos MC, et al. The Ontology for Biomedical Investigations. *PLOS ONE* 2016;11:e0154556. <https://doi.org/10.1371/journal.pone.0154556>.
- [12] Fabry P, Goyer F, Barton A, Ethier J-F. An Ontological Analysis of Health Procedure Information. vol. 3073, 2021, p. 36–47.
- [13] Fabry P, Goyer F, Barton A, Ethier J-F. An informational perspective on the ontology of services. vol. CEUR Workshop Proceedings, CEUR-WS.org; 2022, p. paper 5.
- [14] OBO Foundry n.d. <https://obofoundry.org/principles/fp-004-versioning.html> (accessed March 30, 2023).
- [15] Wilkinson MD, Dumontier M, Aalbersberg IjJ, Appleton G, Axton M, Baak A, et al. The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data* 2016;3:160018. <https://doi.org/10.1038/sdata.2016.18>.
- [16] Plaice J, Wadge WW. A New Approach to Version Control. *IEEE Transactions on Software Engineering* 1993;19:268–76. <https://doi.org/10.1109/32.221137>.
- [17] Gruber TR. A translation approach to portable ontology specifications. *Knowledge Acquisition* 1993;5:199–220. <https://doi.org/10.1006/knac.1993.1008>.
- [18] Guarino N, Oberle D, Staab S. What Is an Ontology? In: Staab S, Studer R, editors. *Handbook on Ontologies*, Berlin, Heidelberg: Springer; 2009, p. 1–17. https://doi.org/10.1007/978-3-540-92673-3_0.
- [19] Neuhaus F, Hastings J. Ontology development is consensus creation, not (merely) representation. *AO* 2022;17:495–513. <https://doi.org/10.3233/AO-220273>.

- [20] OWL 2 Web Ontology Language Primer (Second Edition) n.d. <https://www.w3.org/TR/2012/REC-owl2-primer-20121211/> (accessed June 16, 2023).
- [21] Chacon S, Straub B. Pro Git. Apress; 2014.
- [22] Carbon S, Douglass E, Good BM, Unni DR, Harris NL, Mungall CJ, et al. The Gene Ontology resource: Enriching a GOld mine. *Nucleic Acids Research* 2021;49:D325–34. <https://doi.org/10.1093/nar/gkaa1113>.
- [23] Ceusters W, Smith B. Aboutness: Towards Foundations for the Information Artifact Ontology. *Proceedings of the Sixth International Conference on Biomedical Ontology (ICBO)*, CEUR vol. 1515; 2015, p. 1–5.
- [24] Ceusters W, Smith B. Foundations for a realist ontology of mental disease. *Journal of Biomedical Semantics* 2010;1:10. <https://doi.org/10.1186/2041-1480-1-10>.
- [25] Neuhaus F. What is an Ontology? 2018. <https://doi.org/10.48550/arXiv.1810.09171>.
- [26] Chisholm RM. The primacy of the intentional. *Synthese* 1984;61:89–109. <https://doi.org/10.1007/BF00485490>.
- [27] Michaelson E, Reimer M. Reference. *The Stanford Encyclopedia of Philosophy* 2022. <https://plato.stanford.edu/archives/sum2022/entries/reference/> (accessed March 31, 2023).
- [28] Matentzoglou N, Goutte-Gattat D, Tan SZK, Balhoff JP, Carbon S, Caron AR, et al. Ontology Development Kit: a toolkit for building, maintaining and standardizing biomedical ontologies. *Database* 2022;2022:baac087. <https://doi.org/10.1093/database/baac087>.