Web Semantics: Science, Services and Agents on the World Wide Web 70 (2021) 100658

Contents lists available at ScienceDirect

Web Semantics: Science, Services and Agents on the World Wide Web

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

Beware of the hierarchy — An analysis of ontology evolution and the materialisation impact for biomedical ontologies

Romana Pernisch [a,*], Daniele Dell'Aglio [b,a], Abraham Bernstein [a]

[a] Department of Informatics, University of Zurich, Zurich, Switzerland
[b] Department of Computer Science, Aalborg University, Aalborg, Denmark

ARTICLE INFO

Article history:
Received 1 March 2021
Received in revised form 19 June 2021
Accepted 11 August 2021
Available online 17 August 2021

Keywords:
Ontology evolution
Materialisation
Evolution impact
Ontology change

ABSTRACT

Ontologies are becoming a key component of numerous applications and research fields. But knowledge captured within ontologies is not static. Some ontology updates potentially have a wide ranging impact; others only affect very localised parts of the ontology and their applications. Investigating the impact of the evolution gives us insight into the editing behaviour but also signals ontology engineers and users how the ontology evolution is affecting other applications. However, such research is in its infancy. Hence, we need to investigate the evolution itself and its impact on the simplest of applications: the materialisation.

In this work, we define impact measures that capture the effect of changes on the materialisation. In the future, the impact measures introduced in this work can be used to investigate how aware the ontology editors are about consequences of changes. By introducing five different measures, which focus either on the change in the materialisation with respect to the size or on the number of changes applied, we are able to quantify the consequences of ontology changes. To see these measures in action, we investigate the evolution and its impact on materialisation for nine open biomedical ontologies, most of which adhere to the $\mathcal{EL}^{++}$ description logic.

Our results show that these ontologies evolve at varying paces but no statistically significant difference between the ontologies with respect to their evolution could be identified. We identify three types of ontologies based on the types of complex changes which are applied to them throughout their evolution. The impact on the materialisation is the same for the investigated ontologies, bringing us to the conclusion that the effect of changes on the materialisation can be generalised to other similar ontologies. Further, we found that the materialised concept inclusion axioms experience most of the impact induced by changes to the class inheritance of the ontology and other changes only marginally touch the materialisation.

1. Introduction

Change are inevitable. Especially when capturing our growing knowledge, adaptation becomes necessary [1]. Therefore, ontologies like the Gene Ontology (GO) [2] and the National Cancer Institute Thesaurus (NCIT) [3] change over time to adapt the representation to the evolving knowledge. Experts or communities usually take care of the maintenance of these ontologies by adding new knowledge and removing or updating outdated and wrong information. For example, GO evolves as experts add new gene annotations of living organisms. Researchers then apply functions to this graph, such as functional enrichment analysis [4,5], mRNA expression, proteomics, genetic, or DNA methylation data analysis [6].

In this study, we focus on a general function—the materialisation. It is a deterministic logical entailment, which infers implicit statements. We investigate the materialisation because it is a common calculation used on ontologies to check for consistency but also for subsequent further tasks such as querying and recommendations. A small change in a class definition can have a notable impact on the ontology, significantly vary the number of materialised axioms, or even its consistency. However, not every change in the ontology leads to significant changes in the materialised graph. What are the consequences of ontology evolution on functions like the materialisation over time? Moreover, how does this change in materialisation further impact other tasks and applications?

A materialised ontology is larger than the original and its computation may consume considerable amounts of resources. As a consequence, indicating that a change may cause a potentially significant difference in materialisation would signal the

* Corresponding author.
E-mail addresses: pernisch@ifi.uzh.ch (R. Pernisch), dade@cs.aau.dk (D. Dell'Aglio), bernstein@ifi.uzh.ch (A. Bernstein).

https://doi.org/10.1016/j.websem.2021.100658

necessity for re-computation. This could lead not only to the re-computation of the materialisation but also to other results and calculations on top of the materialisation. Hence, we believe that curators and users need to be aware of the evolution and its consequences.

We first analyse ontology evolution and the applied changes using the following nine Open Biological and Biomedical Ontologies (OBO):

- NCIT [3],
- GO [2],
- Fission Yeast Phenotype Ontology (FYPO) [7],
- Uber-anatomy Ontology (UBER) [8],
- Human Disease Ontology (DOID) [9],
- Pathway Ontology (PWO) [10],
- Rat Strain Ontology (RSO) [11],
- Apollo Structural Vocabulary (ASV) [12], and
- Plant Trait Ontology (PTO) [13].

and investigate the following research questions:

RQ1.1. What is the relation between the number of changes applied to the ontology and its size for open biomedical ontologies, the evolution of which is available online?

RQ1.2. What are the most common complex changes which are applied to open biomedical ontologies, the evolution of which is available online?

We consider the growth of the ontologies in different aspects like number of triples or axioms, and the changes between consecutive versions themselves. We make use of the classification of complex changes provided by COnto-Diff [14] to also assess the evolution and the applied changes in more detail.

Due to the lack of measures to compare materialisations, we first define such an instrument. Therefore, our second research question is:

RQ2. What are measures that are simple to compute and allow to assess the impact of the ontology change on the ontology's materialisation?

We focus on measures that are meaningful in quantifying the impact, and at the same time they require a limited amount of calculation to compare two materialised ontologies. Our goal is to be able to compute these measures in an online environment like Protégé [15], without negatively affecting the usability of such tools. Impact measures can be added to the ChImp plugin [16] to widen the range of information displayed to the ontology editors while they are changing an ontology. We define two sets of impact measures to be able to capture different relations between the changes and the impact on the materialisation. The first set exhibits the consequences of ontology changes with respect to the size of the materialisation. The second set focuses on the changes to the materialisation with respect to the underlying changes to the ontology.

Then, we calculate the impact measures for the selected ontologies and compare them. Given the previous evolution analysis, we want to identify relevant aspects of changes and change types for the impact. Therefore, our last research question is:

RQ3. What aspects of changes and which complex change types have the largest share in the impact on the materialisation?

We found that a distinction between ontologies based purely on size is not statistically significant. Nonetheless, we observe a very high correlation between the number of changes and impact for all ontologies. Additions of leaves, moves, and changes of attribute values are the three most common change types in the evolution of the nine selected ontologies. Further, we find that concept inclusion axiom changes have the highest impact on the materialisation. Most ontologies only experience impact by changes to the class hierarchy. At the same time, the impact is observed on the materialised class hierarchy as well. Most other changes do not have a significant influence on the materialisation at all. Interestingly, complex change actions do not correlate with the impact on the materialisation. The location of the changes (subclass changes) is much more important to determine if changes will have a large impact or not.

Table 1
Definitions of symbols used in Fig. 1 and in the definitions of metrics in Section 3.1.

Symbol	Description
O_i	Ontology at time i
$\delta_i^+ = O_{i+1} \backslash O_i$	Added axioms
$\delta_i^- = O_i \backslash O_{i+1}$	Removed axioms
$M_i = mat(O_i)$	Materialisation of O_i, where M_i does not contain O_i
$\Delta_i^+ = M_{i+1} \backslash M_i$	New, inferred axioms part of M_{i+1} but not in M_i
$\Delta_i^- = M_i \backslash M_{i+1}$	Old, inferred axioms part of M_i but not in M_{i+1}
$impact(M_i, M_{i+1})$	Quantification of the difference between M_i and M_{i+1}

In conclusion, this study presents the following contributions:

- the definition of five novel impact measures, which capture the effect of changes on the materialisation, and
- an investigation of the evolution of nine OBO ontologies and their evolution's impact on the materialisation, where we found that:
  - the size of an ontology does not have a statistically significant effect on the number of changes applied,
  - addition of leaves, moves, and changes of attribute values are the most common complex change action types,
  - the impact on the materialisation is minimal compared to the size of the ontologies,
  - the impact correlates highly with the (absolute and relative) number of changes applied for all ontologies,
  - and, lastly, most changes are applied within the class hierarchy which also leads to the most impact being on the class hierarchy as well.

This paper is structured as follows. Section 2 explains and formalises the problem with the corresponding background information and related research. We define the ontology and impact measures in Section 3 and explain our calculation approach. In Section 4, we introduce the ontologies in detail. We analyse and discuss the evolution as well as the impact on the materialisation in Section 5 and address limitations and future work in Section 6. Lastly, Section 7 concludes this work.

2. Background and related research

This section formalises the problem of ontology evolution and its impact on downstream tasks. A visualisation based on our previous work [17] is shown in Fig. 1. We explain it in the remainder of this section.

We also introduce the background to the various aspects of Fig. 1, as well as the related research. First, we formally define ontologies. Then, we introduce the materialisation task and how it can be executed. The next step is the topic of ontology evolution as well as the formal definition of streams upon which the visualisation and definition from [17] are based. Lastly, we present the topic of impact of evolution and we discuss the related research.

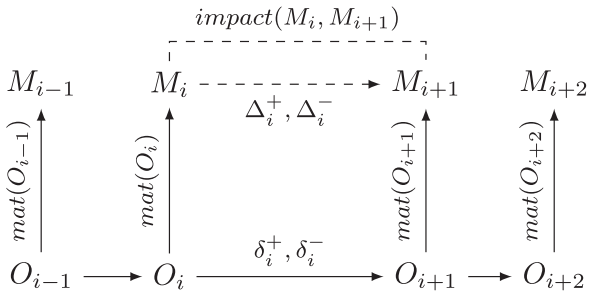


Fig. 1. General model of the problem setting with O_i being the ontology at time t, δ_i the changes on O_i which lead to O_{i+1}, $mat(\cdot)$ the operation executed on the ontology, in our case the materialisation, M_i the result, and $impact(M_i, M_{i+1})$ the impact. All symbols are defined in Table 1.

2.1. Ontology and $\mathcal{EL}^{++}$

We use the definition of ontology proposed by Baader, Brandt, and Lutz in [18]. Such ontologies are expressed through the description logic $\mathcal{EL}^{++}$. OWL 2 EL, which adheres to $\mathcal{EL}^{++}$, is a OWL 2 profile suited for ontologies with a very large number of concepts organised in complex structures. In $\mathcal{EL}^{++}$, the consistency checking and class expression subsumption tasks can be executed in polynomial time. As such, it is often used in biological and biomedical ontology engineering. When talking about ontologies in this work we refer to such $\mathcal{EL}^{++}$ ontologies and denote to one ontology with O.

An ontology is made of two sets—the TBox and the ABox. A TBox consists of four types of axioms: general concept inclusions (GCIs), role inclusions, domain restrictions, and role restrictions. The ABox includes concept and role assertions. We use $\mathcal{T}$ and $\mathcal{A}$ to refer to the TBox and ABox, respectively.

As in other description logics, $\mathcal{EL}^{++}$ distinguishes between classes, individuals, and properties. The set of classes is composed of classes defined in the TBox and used as types in the ABox. Individuals are entities belonging to classes. The set of properties includes those defined in the TBox, and used in the ABox to relate individuals.

2.2. Materialisation and reasoners

Materialisation is the process of calculating the implicit statements in an ontology. This is done by taking into account both the ontological language used to define the ontology as well as the axioms and assertions stored in the ontology. We define the materialisation $mat(\cdot)$ as a function that is applicable to an ontology and produces the result M, as shown in Fig. 1. In this work, the materialisation M does not include O.

There are various *reasoners* that can perform a materialisation and they focus on different ontological languages, and implement distinct algorithms. HermiT [19] and FaCT++ [20] are two examples of general purpose reasoners that handle OWL 2 DL. The HermiT reasoner can check for consistency as well as identify subsumption relationships [19]. However, the reasoning becomes incomplete if the ontology contains property chains or transitivity axioms. FaCT++ is another reasoner for expressive description logic [20]. FaCT++ and HermiT deploy different reasoning algorithms to obtain the materialisation but are expected to return the same results. There are also reasoners which are specifically developed to support ontologies in $\mathcal{EL}^{++}$ logic, such as ELK [21], CEL [22], and TrOWL [23]

Lastly, there are incremental reasoners, such as RDFox and Pellet, especially when dealing with ontology evolution. Incremental reasoning is enabled by reusing previous results when dealing with dynamic knowledge bases. RDFox is a triple store which supports parallel datalog reasoning and incremental updates to the materialisation [24]. Pellet is a sound and complete OWL-DL reasoner that supports reasoning with individuals, custom data types, and other unique features [25].

2.3. Ontology evolution

We define an *evolving ontology* $\mathcal{O}$ as a sequence $(O_1, O_2, \ldots, O_i, O_{i+1}, \ldots)$, where O_i denotes the ontology version i. This definition of evolving ontology is similar to the one of *ontology stream* proposed by Ren and Pan in [26]. Let O_i and O_{i+1} be two consecutive versions in $\mathcal{O}$. The update of $\mathcal{O}$ between i and $i+1$ is described by a set of changes δ_i. δ indicates a set of edits that are authored by one or more agents, such as ontology engineers, curators, or maintenance bots.

Ontology evolution is a well studied and understood topic. Zablith et al. [27] survey various evolution processes. In addition, Hartung et al. [28] show the different tools for managing, exploring, and propagating changes on ontologies. Both focus on how ontologies are maintained. Our study is orthogonal to the mentioned ones, since they do not consider the consequences of the evolution and exclusively look at the need for updates as well as how these are conducted. Rashid et al. [29] use the evolution of a knowledge graph to assess its quality by examining consistency, completeness, persistence, and historic persistence. Quesada-Martínez et al. [30] use the OQuaRE framework [31] to investigate the evolution and quality of eight OBO Foundry ontologies.

The study of ontology changes and change classification is orthogonal to ontology evolution research. OntoDiff [32] is a tool that enables the user to detect changes between two versions of the same graph. It works by identifying semantically equivalent elements between the ontologies. Klein and Noy [33] developed an ontology describing 80 basic changes. They also introduce a notion of complex changes, showing how they help in the interpretation of consequences for data and entities. Papavasileiou et al. [34] propose and evaluate a new approach based on a language to express changes, together with an algorithm to compute changes between versions. They require the ontology to be in RDF(s), which is usually not the case for biomedical ontologies. COnto-Diff [14] and the integrated CODEX [35] both detect changes and group low level changes into high level change actions. They provide a simple classification and a rich action semantics. In our analysis, we will use COnto-Diff to classify changes into complex changes, because it is specifically targeted towards the biomedical domain.

Flouris et al. [36] distinguish between evolution, debugging, and other aspect of ontology change and provide an overview of research in each of the areas. Noy and Klein in [1] already made clear that ontology evolution is not the same as database schema evolution. They point out that an evolution's consequences are generally unknown because of the decentralisation of ontologies. With this work, we aim to lessen the huge gap consisting of the unknown consequences of ontology evolution by quantifying the impact of evolution. Potentially, our research could lead to better informing both sides – the developers and the users – on the issues of ontology evolution and its consequences.

There is also research that focuses on the prediction of ontology evolution [37–39]. Pesquita et al. [37] focus on GO and learn a model predicting which parts of GO will undergo annotation change in the next version. Similarly, Cardoso et al. [38] exploit the evolution of biomedical ontologies and various features to build predictive model to identify concepts that will change in the future as well as the type of changes. Also, Meroño et al. [39] go into this direction of predicting changes and expand the work of Pesquita et al. [37] with semantic drift for the detection and

prediction of changes. These three studies differ from our research since they do not consider consequences of the changes, but they predict where changes will happen in the next version of an ontology. Their models could be used to help ontology editors with their maintenance task, whereas our research would inform them about consequences of the applied changes.

2.4. Evolution impact

Given an ontology O_i, $mat(O_i)$ creates the result M_i. Applied to an evolving ontology $\mathcal{O}$, we have an evolving sequence of results $\mathcal{M} = (M_1, M_2, \ldots, M_i, \ldots)$. Given O_i and O_{i+1}, the respective results M_i and M_{i+1} can be the same when the changes δ_i do not affect the result of $mat(\cdot)$, or they can differ to a varying degree. We model this comparison through the function $impact(M_i, M_{i+1})$, which represents the impact that the evolution had on the results of $mat(\cdot)$.

One form of impact from ontology evolution is *semantic drift*. As an ontology evolves, the meaning of names within the ontology can shift, causing new alignments between concepts and real world things. Therefore, semantic drift is a form of impact of the ontology evolution.

Originally, this type of impact is used in linguistics, where words drift semantically over time. Wegmann et al. [40] recently investigated different forms of semantic drift and how to measure it using word embeddings and neighbourhood comparisons. For ontologies, SemaDrift [41] is a tool to calculate various semantic drift measures between versions of ontologies. It applies different methods of calculation and distinguishes between an exact, inexact, and hybrid ontology matching approach. OntoDrift [42] builds on top of SemaDrift, addressing some of their shortcomings by including more aspects in their semantic drift calculations. A notion of semantic drift has also been investigated in the context of code repositories and bug fixing [43]. The defect prediction quality decreased as the software evolved. The authors found that this worsening can be accredited to a semantic drift of the classes which need to be taken into account when predicting defects. Therefore, the evolution of the repository showed an impact on the prediction. Our work answers similar questions in a different context and focuses on the structural changes.

Different studies focus on the *impact of ontology evolution*. Chen et al. [44] discuss how learned models become less accurate as a stream evolves semantically. Their work is directly related to our approach: they study machine learning as their task, where we focus on materialisation. They measure impact with accuracy loss and use concept drift [45] as the underlying change which causes the impact. Know-Evolve [46] is a model that enables deep temporal reasoning over dynamic knowledge bases. The authors apply machine learning over the graph and predict re-occurrence of events. The time component directly affects the results of the reasoning from which an impact could be derived. Gonçalves et al. [47] define a categorisation of changes based on a logical impact. They investigate if changes affect the set of entailed axioms in the next version and distinguish between effectual and ineffectual changes. Using this categorisation, they analyse NCIT. Gross et al. [4] examine how the changes in an ontology impact previously conducted functional analysis. They propose the stability of individual concepts as their impact measure. Gottron and Gottron [48] also investigate the impact of knowledge base evolution using Linked Open Data. They implement twelve different indexing methods and evaluate how the index is affected by the evolution of the data using three different measures. Dos Reis et al. [49] look into the impact concerning mappings between two evolving ontologies. Cardoso et al. [50] identify the impact on annotation creation using an evolving ontology. Osborne et al. [51] present the pragmatic ontology evolution, in which they analyse the selection of concepts for a new version by evaluating the performance of four different tasks. However, [4, 48–51] focus on one ontology and its specific tasks, where we aim for a broader selection of ontologies and the materialisation — a task not only used by the bioinformatics community but also by other research fields as well as the materialisation is potentially used by further applications. We have previously investigated the impact of changes over embeddings [17] and reported changes in neighbourhoods as impact. Additionally, we used the types of changes and additional ontology information to learn a linear model to estimate the impact without calculating the embeddings [17].

Summarising, ontology evolution and evolution impact has been a topic of various research initiatives. To our knowledge, this is the first study to *define and investigate impact on the materialisation* and at the same time investigating the evolution at this scale. We focus on OBO, but our approach can be applied to any other ontology.

3. Approach

In this section, we define two different groups of measures: supportive and impact measures. Following the definitions, we explain our computation strategy.

3.1. Supportive measures

This section introduces the supportive measures (*popularity, ontology, and edit*) we use to investigate our research questions. *Popularity measures* address some simple methods of assessing the popularity of an ontology. The goal of the *ontology measures* is to capture the information about every single ontology O_i which is part of the evolving ontology $\mathcal{O}$. The *edit measures* capture how two consecutive ontologies O_i and O_{i+1} differ.

Generally, we focus on measures with low computational complexity. Measures with high computational complexity require a large amount of resources when the ontology size is large. We target scenarios where measures should be computed in an (interactive) online fashion, i.e. when the ontology engineer modifies the ontology [16]. Our focus is, therefore, on simple measures, which deliver information about the ontology structure without requiring an excessive amount of resources for calculation.

Popularity measures

We introduce three measures to determine if popularity influences the evolution of the ontologies. Possibly, if a larger number of engineers are working on an ontology or if it is being used more widely, it might influence the evolution of the ontology as additional knowledge needs to be consolidated over time.

First, we consider the number of authors who contribute to the ontology via GitHub, because OBO ontologies are often tracked and shared via GitHub. Using the GitHub API, we can retrieve all contributors and count them. Ekenayake et al. [43] have found that more contributors to a software repository lead to less stable bug prediction results. The authors used the number of contributors to decide when a new prediction model needed to be learned due to the old one becoming inaccurate. The same might be applicable in the domain of ontology engineering and therefore, have an influence on the evolution of ontologies.

Second, we can count the number of months between the start of the project and the latest version, by retrieving the first and last commit to the repository, which also includes the date. Also in this case, we utilise the GitHub API to retrieve such dates. The number of months gives a comparable measure across

ontologies, because ontologies are not being updated with the same periodicity.

Third, we measure the usage of the ontology. Since there is no reliable resource that tracks the ontology usage, we use Google Search to assess the number of mentions of the ontology's permanent link. For this, we use the permanent links of the ontology files, in both OWL and OBO formats, and report the number of entries found by the Google Search Engine. We consider links to both formats because the usage of either OWL or OBO heavily depends on the domain of the application. We do not use both formats in the analysis of the ontology evolution itself. We only consider both permanent links to determine their popularity more accurately.

Ontology measures

Through these measures and the edit measures which follow, we aim at studying the evolution of the ontology, as described with ***RQ1.1***. Table 2 shows the ontology measures, which were initially proposed in [31,52,54–60].

$|O^{\mathcal{T}}|_{\mathcal{EL}^{++}}$ and $|O^{\mathcal{A}}|_{\mathcal{EL}^{++}}$ count the number of $\mathcal{EL}^{++}$ axioms in the ontologies. As explained in Section 2.1, the TBox includes general concept inclusion, role inclusion, as well as domain and range axioms, while the ABox includes the concept and role assertions. To compute the number of axioms and assertions, we defined the SPARQL queries reported in Table 2, which extract the number of axioms for each type, and then we add them up. c and p count the atomic classes and properties defined in the TBox. The inheritance richness, calculated using c and the number of explicit subclass relationships, tells us much about the class hierarchy portion of the ontology.

Edit measures

The edit measures capture differences between consecutive ontology versions. We use these measures to investigate changes δ between two snapshots of an ontology, because the changes give a different perspective on the evolution beyond simple growth investigation of classes, relations, or annotations. Changes provide a more thorough view and allows for more detail, where the ontology measures are not sufficient.

We consider two different approaches for calculating edits between ontology: (i) a simple approach of counting additions, deletions, and moves [47] as well as (ii) a more complex approach which includes the classification of changes into so-called complex change actions [14]. Edit measures presented in Table 3 are solely based on counts of logical axioms and the notion of structural equivalence as defined by Motik et al. [61]. Structural additions refer to the axioms that are in an ontology but not in the previous version. Removals identify the axioms that were in the previous version but are no longer present in the ontology. Shared axioms include the axioms that are in both versions of the ontology. The definitions are shown in Table 3. We consider the axioms which are either additions or removals as *changes*. Just like for the calculation of ontology axioms ($|O|_{\mathcal{EL}^{++}}$), where axioms are counted according to $\mathcal{EL}^{++}$ logic as shown in Table 2, $|\cdot|$ refers to the counting of axioms also using $\mathcal{EL}^{++}$ logic. Therefore, we use the same queries for any $|\cdot|$ operation as we do for $|O|_{\mathcal{EL}^{++}}$, except where specified otherwise (e.g., number of subclasses h). This applies to the number of addition and removals. The query for subclass additions and removal is also already introduced in Table 2 (h in inheritance richness) and used for $h_{\delta_i^+}$, $\delta_{\delta_i^-}$ accordingly. The query is simply applied over the set of changes δ_i^+ or δ_i^- instead of the ontology O. All further usages of $|\cdot|$ are also implemented in this fashion, where the queries are applied over the appropriate set of changes δ_i, ontology O_i, materialisation M_i, or materialisation changes Δ_i.

Since we are dealing with $\mathcal{EL}^{++}$ OBO ontologies, we expect that most changes affect hierarchies of concepts. Therefore, we introduce an edit measure that relates the updates in the subclass relations. Such a measure, called hierarchical moves, counts how many removals have a corresponding addition, where the subject or object of the triple remained the same.

Additionally to the measures described in Table 3, we use COnto-Diff [14] for the classification of changes into complex change actions. The classification is rule-based and has nine atomic changes as basis: addition, deletion, and modification of concepts, attributes, and predicates. These atomic change actions are first identified and then condensed into complex change actions, e.g., the addition of a leaf node consists of the addition of a concept, of its attributes, and of a link to an already existing concept. The classification will enable a more thorough analysis of types of changes and the impact the generate on the materialisation.

3.2. Impact measures

The impact measures quantify how the ontology evolution affects the materialisation. The state of the art does not include ways of comparing materialisations between ontology versions. Known measures are either meant for ontology matching or for comparing two ontologies within the same domain but not related to each other. At the same time, we are looking for measures that are simple to compute, meaning not computationally intensive and computable in an online fashion. We propose the measures presented below to close this research gap. We have two groups of impact measures, the first focusing on impact relative to the size of the materialisation, which we refer to as *size-based metrics* further below (σ and $\sigma_{\sqsubseteq}$). The second set focuses on the number of changes applied to the ontology and, therefore, referred to as *change-based metrics* (γ, $\gamma_{\sqsubseteq}$, and $\gamma_{\not\sqsubseteq}$).

There are numerous ways to define impact, depending on the ontology, the task, and the analysis to be performed. We designed the measures to consider both the removed and added axioms and to make them symmetric (i.e. the impact between O_i and O_{i+1} is the same of the one between O_{i+1} and O_i). Moreover, we took into account that we are dealing with ontologies defining subclass hierarchies (hence: $\sigma_{\sqsubseteq}$, $\gamma_{\sqsubseteq}$, and $\gamma_{\not\sqsubseteq}$), with large TBoxes and small or absent ABoxes. Therefore, we do not pay spacial attention to the ABox in our impact definitions, however, we also do not exclude it.

Size-based metrics. In general, we have two families of impact measures as we previously mentioned. The first includes measures defined as ratios between how much the materialisation changes and how much has remained the same. Denominators signal the amount of axioms that are shared between the two materialisations, and numerators count the differences. We defined the values of impact measures ranging in $[0, \infty)$, where 0 indicates equality between two materialisations. When measures are greater than 1, the materialisation changes substantially, i.e. more than half of the axioms change. Since we are dealing with ontology that are evolving, we usually expect measures lower than 1.

The first impact measure is the percentage of the materialisation which changed due to ontology evolution. We define it as the ratio between the number of inferred axioms that change and of those that do not. To recap, axioms added to the materialisation are denoted with $\Delta_i^+ = M_i \backslash M_{i+1}$ and those removed from the materialisation are $\Delta_i^- = M_{i+1} \backslash M_i$, where the M_x only contains the inferred axioms (i.e., $M_x \cap O_x = \{\}$). We refer to the changes to the materialisation as Δ_i, which accounts for both the added

Table 2
Definition of the ontology measures and implementation (as SPARQL queries, command-line, or mathematical specification combining query results).

Name	Formula	Implementation	Reference
TBox axioms	$\|O^{\mathcal{T}}\|_{\mathcal{EL}^{++}}$	$= GCI + DomA + RanA + RI$	[52]
ABox axioms	$\|O^{\mathcal{A}}\|_{\mathcal{EL}^{++}}$	$= CA + RA$	[52]
general concept inclusion	GCI	`SELECT (COUNT(*) AS ?gci) WHERE {{ ?s rdfs:subClassOf ?o } UNION { ?s owl:disjointWith ?o }}`	[52]
domain axioms	$DomA$	`SELECT (COUNT(*) AS ?dom) WHERE { ?s rdfs:domain ?o }`	[52]
range axioms	$RanA$	`SELECT (COUNT(*) AS ?ran) WHERE { ?s rdfs:range ?o; a owl:ObjectProperty }`	[52]
role inclusion	RI	`SELECT (COUNT(*) AS ?ri) WHERE { {?s rdfs:subPropertyOf ?o } UNION ALL { ?s a owl:ReflexiveProperty } UNION ALL { ?s a owl:TransitiveProperty } UNION ALL { ?s owl:PropertyChainAxiom ?o } UNION ALL { ?s a owl:FunctionalDataProperty }}`	[52]
concept assertions	CA	`SELECT (COUNT(*) AS ?ca) WHERE { ?s a ?o. ?o a owl:Class }`	[52]
role assertions	RA	`SELECT (COUNT(*) AS ?ra) WHERE { ?s ?p ?o. ?p a owl:ObjectProperty }`	[52]
$\mathcal{EL}^{++}$ violations		`java -jar profilechecker.jar <file> OWL2ELProfile`	[53]
$\mathcal{DL}$ violations		`java -jar profilechecker.jar <file> OWL2DLProfile`	[53]
$\mathcal{EL}^{++}$ axioms	$\|O\|_{\mathcal{EL}^{++}}$	$= \|O^{\mathcal{T}}\|_{\mathcal{EL}^{++}} + \|O^{\mathcal{A}}\|_{\mathcal{EL}^{++}}$	[52]
classes	$c = \|\mathcal{C}^{\mathcal{T}}\|$	`SELECT (COUNT(DISTINCT ?s) AS ?c) WHERE { ?s rdf:type owl:Class. FILTER isURI(?s) }`	[54,55]
defined properties	$p = \|\mathcal{P}^{\mathcal{T}}\|$	`SELECT (COUNT(?p) AS ?pt) WHERE {{ ?p a owl:DatatypeProperty } UNION ALL { ?p a owl:ObjectProperty }}`	[54–57]
inheritance richness	h/c	$h =$ `SELECT (COUNT(*) AS ?h) WHERE { ?s rdfs:subClassOf ?o }`	[31,54,58–60]

Table 3
Definition of edit measures for two consecutive ontologies O_i and O_{i+1}, as well as their materialisations M_i and M_{i+1}.

Ontology (O_i and O_{i+1})	Materialisation (M_i and M_{i+1})	Name
$\|\delta_i^+\| = \|O_{i+1} \backslash O_i\|$	$\|\Delta_i^+\| = \|M_{i+1} \backslash M_i\|$	Structural additions
$\|\delta_i^-\| = \|O_i \backslash O_{i+1}\|$	$\|\Delta_i^-\| = \|M_i \backslash M_{i+1}\|$	Structural removals
$\|\delta_i\| = \|\delta_i^+\| + \|\delta_i^-\|$	$\|\Delta_i\| = \|\Delta_i^+\| + \|\Delta_i^-\|$	Structural changes
$\|\delta_i\| / \|O\|_{\mathcal{EL}^{++}}$	$\|\Delta_i\| / \|M\|_{\mathcal{EL}^{++}}$	Relative changes
$h_{\delta_i^+} = \|H_{\delta_i^+}\|$ with $H_{\delta_i^+} = \{c \mid (c, subclassOf, o) \in \delta_i^+\}$	$h_{\Delta_i^+} = \|H_{\Delta_i^+}\|$ with $H_{\Delta_i^+} = \{c \mid (c, subclassOf, o) \in \Delta_i^+\}$	Subclass additions
$h_{\delta_i^-} = \|H_{\delta_i^-}\|$ with $H_{\delta_i^-} = \{c \mid (c, subclassOf, o) \in \delta_i^-\}$	$h_{\Delta_i^-} = \|H_{\Delta_i^-}\|$ with $H_{\Delta_i^-} = \{c \mid (c, subclassOf, o) \in \Delta_i^-\}$	Subclass removals
$h_{\delta_i} = h_{\delta_i^+} + h_{\delta_i^-}$	$h_{\Delta_i} = h_{\Delta_i^+} + h_{\Delta_i^-}$	Subclass changes
$h_{\delta_i^+ \cap \delta_i^-} = \|H_{\delta_i^+} \cap H_{\delta_i^-}\|$	$h_{\Delta_i^+ \cap \Delta_i^-} = \|H_{\Delta_i^+} \cap H_{\Delta_i^-}\|$	Subclass moves
$h_{\delta_i^+ \cap \delta_i^-} / 2 \times \|\delta_i\|$	$h_{\Delta_i^+ \cap \Delta_i^-} / 2 \times \|\Delta_i\|$	Share of subclass moves

and the removed axioms. The unchanged axioms of the materialisation is captured by the intersection of the two materialisations ($M_i \cap M_{i+1}$). We can now define σ:

$$\sigma = \frac{|\Delta_i|}{|M_i \cap M_{i+1}|} \tag{1}$$

as the number of changes divided by the number commonalities between the materialisations. When ontologies are large, we expect the change impact to be close to 0, as the effect of the changes on the materialisation should be dominated by the number of inferred axioms that are not affected.

We have a second impact measure in this group, which focuses solely on the subclass hierarchy. The numerator and denominator only consider the subclass axioms of Δ_i and $M_i \cap M_{i+1}$:

$$\sigma_{\sqsubseteq} = \frac{h_{\Delta_i}}{h_{M_i \cap M_{i+1}}}, \tag{2}$$

where similar to h_{δ_i}, h_{Δ_i} is the number of `SubClassOf` axioms in Δ_i. $h_{M_i \cap M_{i+1}}$ can be understood accordingly, where the `SubClassOf` axioms are counted for the shared axioms between the two materialisations $M_i \cap M_{i+1}$. Hence, $\sigma_{\sqsubseteq}$ corresponds to the number of added and removed subclass axioms in the materialisation divided by the joint subclass axioms of the two materialisations. This impact focuses on the hierarchy, and we expect it to be effective with biomedical ontologies, which often define taxonomies.

Change-based metrics. The second family of impact measures do not consider the size of the ontology. Rather, they focus on the changes in the materialisation (Δ_i) compared to the changes in the ontology (δ_i). Therefore, the general impact based on changes is:

$$\gamma = \frac{|\Delta_i|}{|\delta_i|} \tag{3}$$

In contrast to the previous impact measures, γ only considers the amount of changes applied to the ontology and the following changes to the materialisation. This means that when there are more changes in the materialisation than in the underlying ontology, impact is high (above 1.0). Consequently, when impact is below 1.0 the ontology changes did not lead to as many changes in the materialisation.

We also introduce a specific impact measure for taxonomies, which only regards the subclass changes to the materialisation (h_{Δ_i}) in the numerator and the subclass changes to the ontology (h_{δ_i}), as previously defined in Table 3. Instead of only counting the subclass changes to the materialisation we also subtract the subclass moves ($h_{\delta_i^+ \cap \delta_i^-}$). This allows us to divide the subclass changes to the materialisation by the changes to the ontology, which do not have matches between additions and deletions.

$$\gamma_{\sqsubseteq} = \frac{h_{\Delta_i}}{h_{\delta_i} - 2 \times h_{\delta_i^+ \cap \delta_i^-}} \tag{4}$$

Also for this impact measure, we expect the impact value to be around 1.

Lastly, since we anticipate most impact to come from the class hierarchy, we are also curious about the remaining changes and if they impact the materialisation as well. To get the number of changes without the subclass changes, we simply subtract the subclass changes (h_{Δ_i} for the nominator and h_{δ_i} for the denominator) from the changes overall (Δ_i for the nominator and δ_i for the denominator).

$$\gamma_{\not\sqsubseteq} = \frac{|\Delta_i| - h_{\Delta_i}}{|\delta_i| - h_{\delta_i}} \tag{5}$$

With this last impact measure, we can investigate the impact on the materialisation without the hierarchy.

We further analyse the usefulness of the impact measures in Section 5. We use ontologies and the supportive measures for the analysis. We then discuss and answer research ***RQ2*** in more detail with real world data.

3.3. Metrics computation

All the measures presented above are implemented in the framework depicted in Fig. 2. The workflow is supplied with an evolving ontology and produces sequences of ontology, evolution, and edit measures.

The *Materialiser* takes an ontology O_i as input and computes the materialised axioms M_i. We implemented this component using the OWL2 API [62], which allows exploiting state of the art reasoners, among which we chose HermiT [19]. Since some ontologies are in OBO format, the component uses the Robot module [63] to convert them in OWL before materialisation.

The *Diff-calculator* processes pairs of consecutive materialisations M_i and M_{i+1} to generate the sets of added axioms Δ_i^+, deleted axioms Δ_i^-, and shared axioms $M_i \cap M_{i+1}$. In parallel, the Diff-calculator also processes pairs of consecutive ontologies O_i and O_{i+1} to generate the sets of added axioms δ_i^+, deleted axioms δ_i^-, and shared axioms $O_i \cap O_{i+1}$. We use the Ecco framework [64] as implementation of this component.

COnto-Diff[1] is the module which also calculates a difference between ontology [14]. However, unlike Ecco, this tool categorises the diff into multiple complex changes. Simple changes (additions, deletions of concepts, relations, or attributes) are grouped together to form more complex changes. We expanded the implementation to return the number of the different complex change actions instead of the entire list.

The measures are calculated using three separate python modules. The *edit measures* module computes the edit measures by processing the added/deleted/shared axiom sets $\langle \delta_i^+, \delta_i^-, O_i \cap O_{i+1} \rangle$. The *impact measures* module takes as input the data about the materialisations, i.e. M_i and M_{i+1}, as well as their differences and intersection $\langle \Delta_i^+, \Delta_i^-, M_t \cap M_{i+1} \rangle$, to compute the impact measures. The *ontology measures* module takes an ontology O_i as input and calculates the ontology measures. The three components were developed in Python 3.6, using the rdflib's SPARQL interface.

Finally, the calculated measures are analysed by the *Measure Analyser* — an R script that computes the monthly growth, averages, and standard deviations. We also use it to test hypotheses, where applicable, and to generate tables and plots displayed in the next sections. Based on the results from the Measure Analyser, we assess and answer our research questions.

The code of the overall running script as well as the R code used in the analysis are open source and available on our project repository[2] and published under the GNU GPLv3 License. It also includes the modified version of Ecco and COnto-Diff as well as the Materialiser. For all three components code is available, as well as a build jar-file which was used during our calculations.

4. Datasets: Analysed ontologies

Even though there are many ontologies freely available, only a few record the edit history and make it available in a reusable format. GO and NCIT are two well known ontologies, which publish not only the most current version, but also the previous ones.[3] Additionally, we consider seven ontologies from the OBOFoundry [65]: DOID, FYPO, UBER, PWO, RSO, ASV, PTO.[4] In this section, we describe our strategy for choosing the ontologies. Next, we introduce each of them ordered by their size.

To select the OBO Foundry ontologies, we analysed its 260 ontologies. 49 of them are marked as obsolete, ten as inactive, and six as orphaned. Out of the 179 remaining ontologies, 16 have either no available files or broken links to their repositories. We found that 37 repositories include two files, one marked as releases (called simply `<ontology>.owl/.obo`) and the other as edit file (`<ontology>-edit.owl/.obo`). The edit file is used for changing the ontology until it is deemed good enough for release. The commit messages of the release files follow the format "Release ⟨date⟩" and signal official releases with no intermediate commits to these files. We consider the release files, as the edit ones lead to extremely small changes that are not relevant for the reasoning task [66]. Moreover, the released ontologies are ready to be used, and should not contain errors due to the ongoing editing. Finally, we selected the ontologies with more than 100 releases, to have a sufficient amount of versions to observe their evolution and run statistical testing. Some remaining ontologies fit these criteria but were not selected because of parsing or materialisation issues with many of their available versions, resulting in seven ontologies selected from OBO Foundry.

Table 4 shows the ontology measures and the profile violations for the considered ontologies. We report and discuss the ontologies starting with the largest ontology and ending with the smallest one. The first row reports the number of versions for each ontology, followed by the number of months they cover. The second block of rows reports the compliance of the ontologies to the OWL2 EL profile, as well as OWL2 DL by listing the number of violations. We expect the ontologies to be in the $\mathcal{EL}^{++}$ profile.

[1] https://github.com/dbs-leipzig/conto_diff.

[2] gitlab.ifi.uzh.ch/ddis-public/chimp-mat.

[3] Raw data can be downloaded from http://www.geneontology.org and https://evs.nci.nih.gov/ftp1/NCI_Thesaurus/archive/.

[4] http://www.obofoundry.org/.

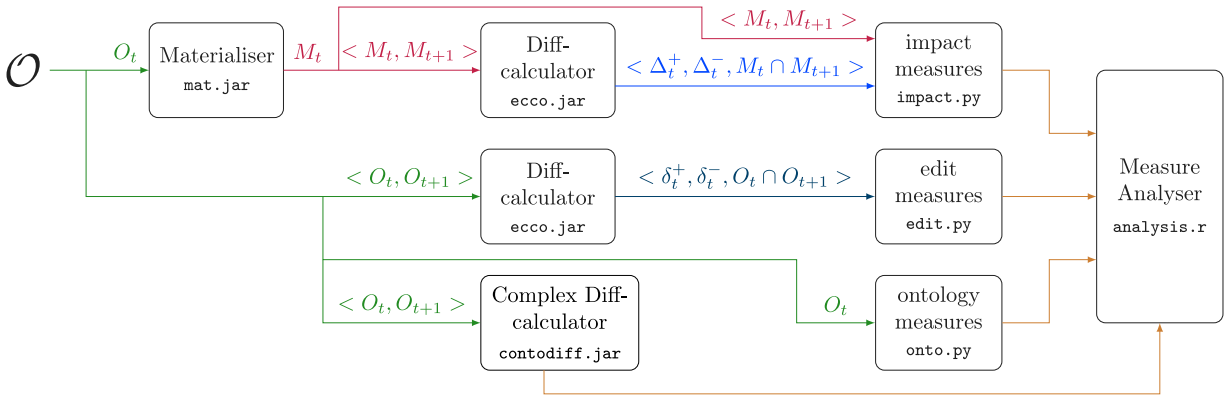


Fig. 2. Calculation Pipeline.

Table 4
The number of versions and months, which are being considered, and violations for OWL2 profiles as well as min, max and monthly change (mc) of the number of $\mathcal{EL}^{++}$ axioms ($|O|_{\mathcal{EL}^{++}}$) and classes (c), properties (p), and inheritance richness (h/c). Reported are also the number of versions and months which are being considered and violations for OWL2 profiles.

Ontology		NCIT	GO	DOID	FYPO	UBER	PWO	RSO	ASV	PTO
Versions		185	122	107	356	254	104	140	211	144
Months		108	195	51	56	116	98	101	58	108
viol.	EL	232	2	30	3	22	OK	OK	269	1
	DL	97	1	17	3	OK	OK	OK	44	OK
$\|O\|_{\mathcal{EL}^{++}}$	min	50'937	52'522	31'892	19'011	8'014	3'262	1'515	1'065	925
	max	322'434	120'259	48'930	22'270	19'295	6'532	1'722	3'258	1'783
	mc	2'281.487	340.820	261.036	57.175	116.300	65.400	3.569	21.930	7.461
	mean	190787	83995	12178	40208	20841	2157	5639	1255	1606
	sd	75643.8	20444.9	4067.3	4282.2	1098.0	731.1	747.5	255.6	95.8
c	min	28'757	31'989	8'446	8'495	9'192	2'544	1'347	857	642
	max	124'025	50'119	13'984	9'396	12'451	4'728	1'549	2'642	1'412
	mc	800.571	93.454	65.764	15.807	33.598	43.680	3.483	17.850	6.696
	mean	87338	43209	10683	10597	9004	1733	4048	965	1438
	sd	27278.8	5618.8	1136.6	1255.1	299.6	597.7	535.1	254.9	92.7
p	min	66	4	44	64	12	1	2	1	68
	max	203	9	55	69	20	1	16	1	89
	mc	0.908	0.021	0.109	−0.035	−0.010	0.000	0.241	0.000	0.148
	mean	169.82	5.52	14.88	46.63	66.18	1.00	1.00	78.67	7.60
	sd	32.82	1.90	2.26	1.45	1.62	0.00	0.00	7.14	7.04
h/c	min	1.247	1.641	2.567	1.749	0.871	1.273	1.103	1.220	1.006
	max	2.313	1.958	2.896	1.870	1.440	1.418	1.117	1.270	1.037
	mc	−0.004	0.001	0.001	0.002	0.006	0.002	−0.0002	−0.0001	−0.0003
	mean	1.37	1.83	1.05	2.78	1.82	1.24	1.39	1.02	1.11
	sd	0.246	0.096	0.209	0.053	0.039	0.017	0.024	0.012	0.004

However, if not, we also check for the DL profile, to ensure that HermiT can be used as reasoner for all ontologies. The remaining part of the table contains the ontology measures. For each measure, the table provides the minimum, maximum, monthly change, mean, and standard deviation. The minimum and maximum do not have to be from the first and last versions of each ontology, they are to signal the range of the ontology measure. The monthly change for the number of classes c is $c_n - c_1 / months - 1$, where c_1 and c_n are the numbers of classes in the first and last version, respectively. The raw data we collected and calculated (without aggregations) is available for further analyses.[5]

NCIT is a widely recognised standard for biomedical coding and Refs. [67]. It provides a vocabulary for diverse medical fields: clinical care, transitional and basic research, public information, and administrative activities. We use 190 versions of NCIT from October 2003 to December 2019. NCIT shows the largest number of violations: 232 axioms violate OWL2 EL and 97 axioms violate OWL2 DL. However, those numbers are negligible when considering the total number of axioms. Among the ontologies we consider, NCIT is the biggest and the one with the largest growth.

GO is a well-known ontology in the biomedical domain and has been maintained by the Gene Ontology Consortium since 2000. GO provides a precise and common vocabulary to describe the role of genes and gene products in any organism [2]. We use 123 versions from January 2010 to December 2019. The older versions of GO that we are using are not directly available in OWL. Therefore, we decided to use all the versions in the OBO format, and convert them with Robot [63]. GO shows two violations of the OWL2 EL profile, related to the metadata of the ontology: one is an axiom declaring an inverse property in the context of the `versionIRI` and `OntologyID`, and the other is an undeclared annotation (`license`). Since both violations do not pose a problem, GO can be considered compliant to $\mathcal{EL}^{++}$. GO grows faster than the ontologies presented below when considering the number of classes. But the number of defined properties increases by five throughout the evolution, which is small compared to the growth of number of classes.

DOID is the human disease ontology [9]. The ontology has undergone significant expansions in the past three years. Though an expansion from single asserted classification to multiple-inferred mechanistic classification, it provides a new perspective on related diseases. DOID had 107 commits on its GitHub repository, which cover 51 months starting from November 2015. The

5 https://gitlab.ifi.uzh.ch/ddis-public/chimp-mat.

profile-checker reports that there are 30 violations for OWL2 EL and 17 for OWL2 DL. This ontology grows significantly in the considered time period. The slope for the number of properties is negative, indicating that the last version of DOID has fewer defined properties than the first one.

FYPO is an ontology of phenotype observed in fission yeast [7]. FYPO versions range from July 2015 to February 2020, and it is the ontology with the highest number of versions (355) among the ones we consider. FYPO shows three violations to the OWL2 EL Profile, all of which are the use of undeclared annotation properties (`title`, `description`, and `license`). Therefore, it is safe to assume that FYPO is an $\mathcal{EL}^{++}$ ontology. Both the number of axioms and classes increase over time, while the number of properties does not grow significantly.

UBER is an integrative cross-species anatomy ontology [8]. It is organised according to traditional anatomical classification criteria, and models concepts in a species-neutral way. UBER reuses several concepts and properties of other anatomical ontologies. UBER has 305 versions, but we consider 254 because the 46 versions cannot be parsed. The versions we use cover 116 months starting in September 2010. This ontology does not show any OWL2 DL violations, while it has 22 OWL2 EL violations. UBER is the third largest ontologies among the ones we consider. It is also the third fastest growing ontology, with a growth rate slope values slightly higher than DOID. The number of properties has the highest growth among the ontologies we consider, but it is small compared to classes and axioms.

PWO is the Pathway Ontology and it contains all known types of biological pathways, including disease paths [10]. It also incorporates relations among pathways, creating an acyclic directed graph structure. PWO has 104 versions, which cover 98 months starting from March 2011. PWO passes the profile-checker without violations for OWL2 EL, and consequently OWL2 DL. The number of properties does not change over time, and we observe that the number of axioms grows faster than classes.

RSO is another ontology developed by the maintainers of PWO. It is a structured vocabulary for facilitating access to rat strain data [11]. It models the breeding history, parental background, and genetic manipulations. RSO cover 101 months starting from February 2011, and it has 146 versions, but we considered 140 of them (6 raised parsing errors). Just like PWO, RSO passes the profile-checker without violations for OWL2 EL and DL. The size of RSO is also comparable to PWO. The number of properties is constant, and the growth of axioms is larger than the growth of the classes.

ASV provides definitions that are necessary for inter-operation between epidemic simulators and public health application software [12]. Versions of ASV range from May 2015 to February 2020, with 216 commits on the owl file. Despite its small size, this is the ontology with the highest number of violations of the OWL2 EL profile. The smallest version has fewer than 1'000 axioms, and grows to a maximum of 1'781 axioms. However, as the slope shows, the first and last version have the same size. Therefore, ASV grew during the covered time, but also shrank to its original size with the last version. This is visible for classes and properties. It is also interesting to note that this ontology has a large number of defined properties, considering its size.

PTO is the Plant Trait Ontology; it is part of the Plant Ontology, maintained by the Plant Ontology Consortium [13]. As the name suggests, PTO encodes traits of plants. It is used in many other ontologies focusing on specific trait of plants and on genetics of plants. PTO has 144 versions, and it covers 59 months beginning in July 2015. We were unable to run the profile-checker on most versions of this ontology. This ontology is a OWL2 DL ontology. The one violation for OWL2 EL is an inverse object property definition. This profile was derived from version 24. All later versions either gave a `FileNotFoundException` for importing ontologies or a `NullPointerException` after the loading of the ontology with the profile-checker. The growth rate between axioms and classes is close to each other for PTO.

In conclusion, our analysis of the selected datasets confirms our expectations: the ontologies have small or empty ABoxes and most comply to the OWL2 EL profile. At the same time, the ontologies show to have different sizes, growth rates, and number of versions.

As previously stated in this section, we analysed ontologies in OBO Foundry, which are about 30% of those referenced by BioPortal (877)[6] [68], and selected the ones fit for our analysis. Therefore, of the 30% analysed ontologies, we are confident that we nonetheless selected a fairly representative sample of ontologies which fit our selection criteria and therefore, are appropriate to study the research questions we introduced in Section 1.

5. Changes and their impact

First, we analyse the evolution of the chosen nine ontologies. We take a closer look at the changes and some trends within them. We will then address the different impact measures and what aspects they capture together with an analysis of the effect of evolution on the materialisation.

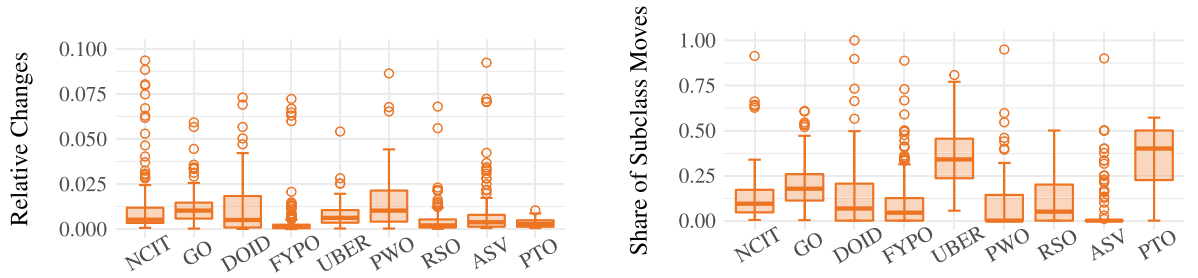
5.1. Ontology evolution

In Fig. 3, we report different distributions of the relative changes. The substantial ontology size difference also leads to a large difference in the number of changes, therefore, we visualise relative numbers rather than absolute ones. Additionally, visualising ratios allows us to see outliers more clearly. Firstly, Fig. 3(a) shows the distributions of structural changes relative to the respective ontology's size ($|\delta_i|/|o|_{\mathcal{EL}^{++}}$). Secondly, Fig. 3(b) reports the distribution of the share of hierarchical moves relative to the number of structural changes ($h_{\delta_i^+ \cap \delta_i^-}/|\delta_i|$). In Table 5, we also report the means of the relative and structural changes. Additionally, this table shows the number of contributors on GitHub and the number of Google search results of the ontologies official file link for both the OWL and OBO file, as explained in Section 3.1, Popularity Measures.

The number of absolute structural changes is directly related to the size of the ontology. This can be read in Table 5. However, this relation does not apply to relative changes. We confirmed both of these observations using a pairwise t-test to evaluate if samples are taken from the same distribution. A significant result reports that the means of the two samples are indeed different from each other and, therefore, from different distributions. We cannot report only significant results, neither for relative nor for absolute changes, which means that not all ontology comparisons yields a significant difference between means. So, even though we observe a relation between the size of the ontology and number of changes, we do not confirm it statistically.

Since size does not explain the differences, we investigate the number of contributors and the popularity (mentions) of the ontology reported in Table 5. By observation, we cannot see a connection between size of ontology and these two other aspects. We did not test this statistically because the correlation result with only eight points is unreliable. However, there seems to be a connection between the popularity of an ontology and the number of contributors, with two exceptions: PWO and RSO. These two ontologies are developed by the same team and are not properly tracked using GitHub. Therefore the number of contributors is not reliable for these ontologies. Additionally, we do

[6] https://bioportal.bioontology.org, last access: 29/05/2021.



(a) Distributions of number of relative changes (to size of ontology, $\mathcal{EL}^{++}$ axioms). 52 outliers were cut off from the visualisation.

(b) Distributions of the share of subclass moves within structural changes. 34 special cases (value > 1.0) have been removed from the visualisation.

Fig. 3. Distributions of shares of different type of changes within the structural changes.

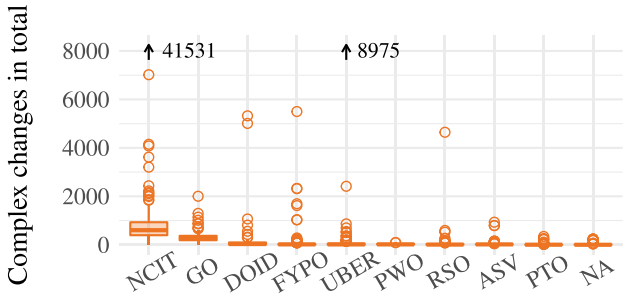


Fig. 4. Distribution of total number of complex changes computed with COnto-Diff [14] (with 0s). Two outliers were omitted from the graph and are indicated by including there actual number at the top of the plot.

Table 5
Mean of relative ($|\delta_i|/|o|_{\mathcal{EL}^{++}}$) and structural ($|\delta_i|$) changes, ontologies ordered by absolute number of changes (same as size). Number of Google search results on 15.01.2021.

	Contributors	Mean changes		Mentions	
		relative	structural	obo	owl
NCIT	–	0.205	11'774.125	–	16'900
GO	26	0.018	1'429.635	25'700	10'300
DOID	9	0.049	540.210	238	263
FYPO	4	0.006	261.195	88	83
UBER	12	0.009	185.667	5'630	4'090
PWO	(3)	0.007	14.660	1'120	949
RSO	(3)	0.003	14.729	81'200	31'300
ASV	5	0.014	22.207	–	8
PTO	8	0.003	4.333	419	285

not have a number of contributors for NCIT, because this ontology is not updated or shared via GitHub. We also note that for GO and RSO we see a large difference in mentions for OBO versus OWL files. We believe that these two ontologies are much more used within the biomedical domain compared to elsewhere, which explains the higher number of mentions of the OBO file format.

We can now answer the research ***RQ1.1***: *What is the relation between the number of changes applied to the ontology and its size for open biomedical ontologies, the evolution of which is available online?* We observe the intuition within the presented data that larger ontologies experience more changes. However, we are unable to confirm this statistically at this point.

At the same time, the number of contributors does not explain the absolute or relative number of changes. We observe a relation between the number of contributors and the popularity of an ontology. Even though we notice growth tendencies and how ontologies evolve based on size, there is still some uniqueness to each of the ontologies that should not be left unnoticed, because size does not totally explain all differences between the ontologies.

Further, Fig. 3(b) shows the distribution of hierarchical moves in relation to structural changes ($h_{\delta_i^+ \cap \delta_i^-}/2\times|\delta_t|$). A move consists of one addition and one deletion of subclass axioms, hence, we calculate the ratio using the number of changes divided by 2. In 34 versions (removed from the visualisation), certain moves were counted multiple times because there was more than one addition or deletion identified as part of the move. Hence, in these cases the ratio is larger than 1.0 and they are omitted from Fig. 3(b). In most cases, the main part of the distribution (the box) is under 0.5 (ASV, DOID, FYPO, NCIT, PWO, RSO). This signals that for most versions the moves only make up about 25% of the changes. This is different for PTO and UBER. The means are around 0.3, which is higher than for the other seven ontologies, and there are almost no outliers. Therefore, these two ontologies experience this type of change more often. For DOID, FYPO, PWO, RSO, and ASV, the subclass moves do not occur as often, as for GO and NCIT. GO and NCIT's distributions of the share of subclass moves is around 0.125, with theirs whiskers touching 0. Whereas for the remaining five ontologies, the boxes end at 0 and no whiskers are visible.

Fig. 4 shows the distribution of all complex changes detected with COnto-Diff [14]. Two outliers have been cut off and are labelled at the top of the graph. Fig. 5 shows the distributions of the relative number of the specific complex change action types in comparison to the total number of change actions. For completeness, we report the min, max, mean, median, and standard deviation values in the appendix (Table A.8). We can see that the addition of leaves (*AddLeaf*), moves (*Move*), and the changes of attribute values (*ChgAttValue*) build the most common complex change actions as categorised by Hartung et al. [14]. Changes to attribute values are present often, but do not necessarily make up the largest part of the changes except for ASV and PTO or NCIT, where there are none. In contrast, NCIT experiences the highest number of additions of leaves and a smaller number of moves when compared with GO, DOID, FYPO, UBER, PWO, and RSO. ASV and PTO show a slightly different pattern with a lot of changes to attribute values but a small amount of leaf additions and even fewer moves. The remaining complex changes are visibly less present. Therefore, we observe different types of behaviours which we categorise into three groups:

(1) Ontologies in the first group experience many additions (leaves), small amount of moves and close to no change to attributes, visible in the evolution of NCIT. We see the goal of the evolution to be the addition of new information but only small amount of modifications to already present information with moves. Ontologies in this group show a very large monthly growth.
(2) The second group shows changes mostly in the form of leaf additions, but these make up less than 50% of the changes on average. Moves and changes to attributes together make up the rest of the changes. This group includes six ontologies: GO, DOID, FYPO, UBER, PWO, and RSO. These ontologies still

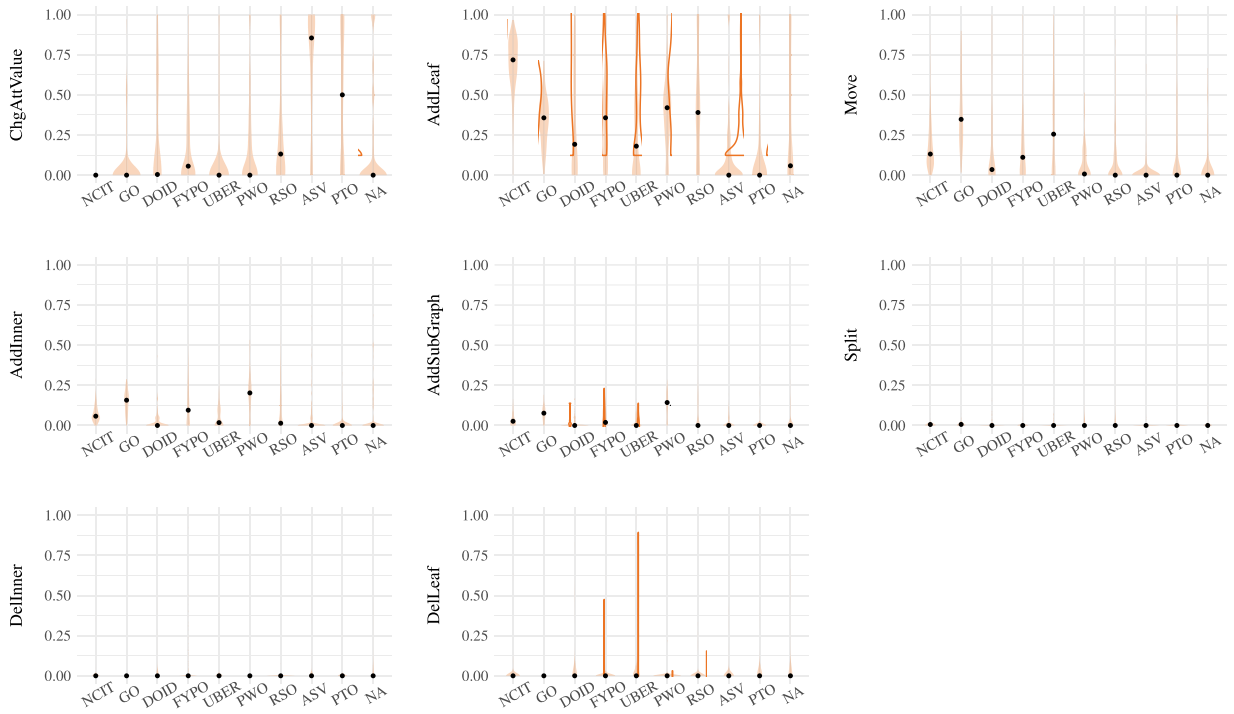


Fig. 5. Violin plots which show the distribution of complex changes relative to the total of complex changes. Outliers are connected to the violin and the point denotes the median. 0s are not visualised, as seen for example by *ChgAttValue* of NCIT.

grow substantially, but far less than those in the first group (NCIT).

(3) Lastly, the third group experiences mainly maintenance in the form of changes in attribute values and also moves. These ontologies (ASV and PTO) do not experience that many additions, and therefore, also show the lowest growth rate compared to the first two groups.

We could also see this as a spectrum, where on the one end we find NCIT with the highest growth (group 1) and on the other and we have ASV and PTO that are mostly maintained and do not grow much over time (group 3). The second or middle group experiences both, addition of new information (growth) and maintenance at the same time.

At this point we can also answer the research ***RQ1.2***: *What are the most common complex changes which are applied to open biomedical ontologies, the evolution of which is available online?* The most common changes are additions of leaves, moves, and changes to attribute values. Based on these three types of changes, we also observe three slightly different change behaviours among the ontologies. In short, they can be classified as either growing (1), being mostly maintained (3), or something in between, growing and also experiencing maintenance (2).

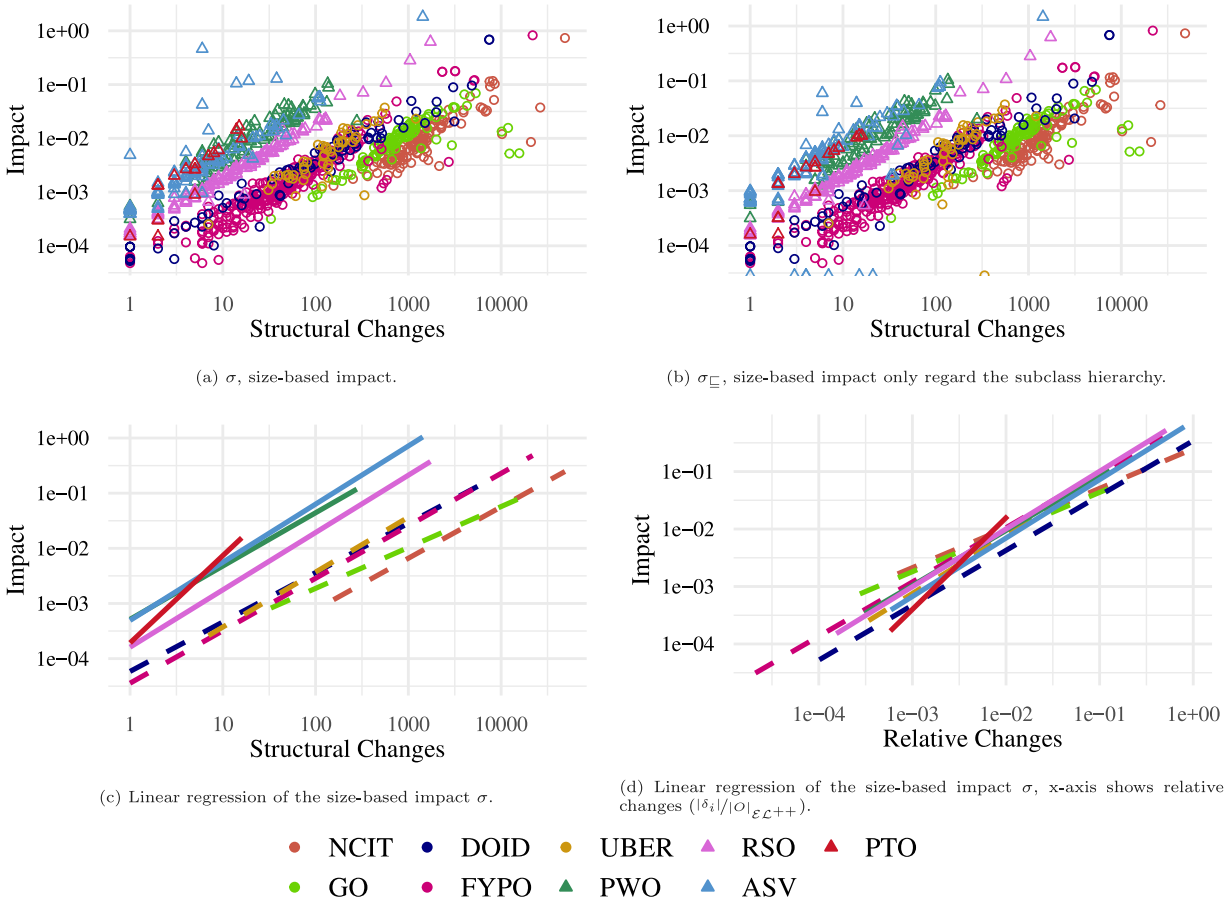
5.2. Influence on materialisation

To further discuss the second research question and to answer the last one, we investigate the evolution's impact on the materialisation. Fig. 6 shows the impact measures based on size, in comparison to the number of structural changes. Figs. 6(a) and 6(b) show the individual points in the evolution of ontologies for σ and $\sigma_{\sqsubseteq}$, and Figs. 6(c) and 6(d) show the linear regressions for each ontology for σ. Fig. 6(c) relates the impact to structural changes, whereas Fig. 6(d) relates it to relative changes. In Figs. 6(a)–6(c) we observe two clusters of points or lines. The ontologies use different shapes or line types based on size to improve readability. In the scatter plots, triangles are associated with small ontologies, which have fewer than 6'587 classes; the remaining ontologies are large and are illustrated with circles. In the line plots, full lines denote the small ontologies and dashed lines denote the large ontologies. When the same amount of change is involved, the large ontologies have a lower impact on the materialisation compared to the small ontologies. This behaviour confirms our initial idea: the smaller the ontology, the higher impact of one single change.

σ and $\sigma_{\sqsubseteq}$ penalises large ontologies because of their size. We see that the absolute number of changes for large ontologies is larger than in small ones (Table 5). As the figures indicate that the size of the ontology provides a classification into the two visible clusters. Labelling each point according to the cluster visible in Fig. 6(a), we trained a decision tree model to identify the size of the ontology which explains this classification visualised with the two shapes of points (circles and triangles). The classification into circles and triangles is used as the independent variable for the decision tree model. Using relative changes, as visible in Fig. 6(d), we do not observe these two clusters of impact based on the size of the ontology. Therefore, we infer that ontologies experience the same amount of relative changes, which coincides with the same amount of impact. This observation is supported by the parallel linear regression lines in Fig. 6(c) as well as in Fig. 6(d). The slope and intercept numbers for both figures are also reported in Table 6 on the left side. The right side of Table 6 lists the Spearman correlations between impact values and number of changes (either structural or relative) and their p-values. We can see that the correlation is almost one for all ontologies and all are highly significant.

Fig. 7 shows the set of impact measures related to the number of changes. The three scatter plots relate the number of structural changes with γ, $\gamma_{\sqsubseteq}$, and $\gamma_{\not\sqsubseteq}$, respectively. Assuming that every change would effect at least one materialised axiom, we expect the plots to hover near or above 1.0. Numbers far above 1.0 would indicate structural changes with a large impact, numbers below 1.0 would indicate the presence of changes without any impact on the materialisation.

We see differences when comparing Figs. 7(a) and 7(b). Though overall the cluster prevails, $\gamma_{\sqsubseteq}$ is situated slightly higher than



(a) σ, size-based impact.

(b) $\sigma_{\sqsubseteq}$, size-based impact only regard the subclass hierarchy.

(c) Linear regression of the size-based impact σ.

(d) Linear regression of the size-based impact σ, x-axis shows relative changes ($|\delta_i|/|O|_{\mathcal{EL}++}$).

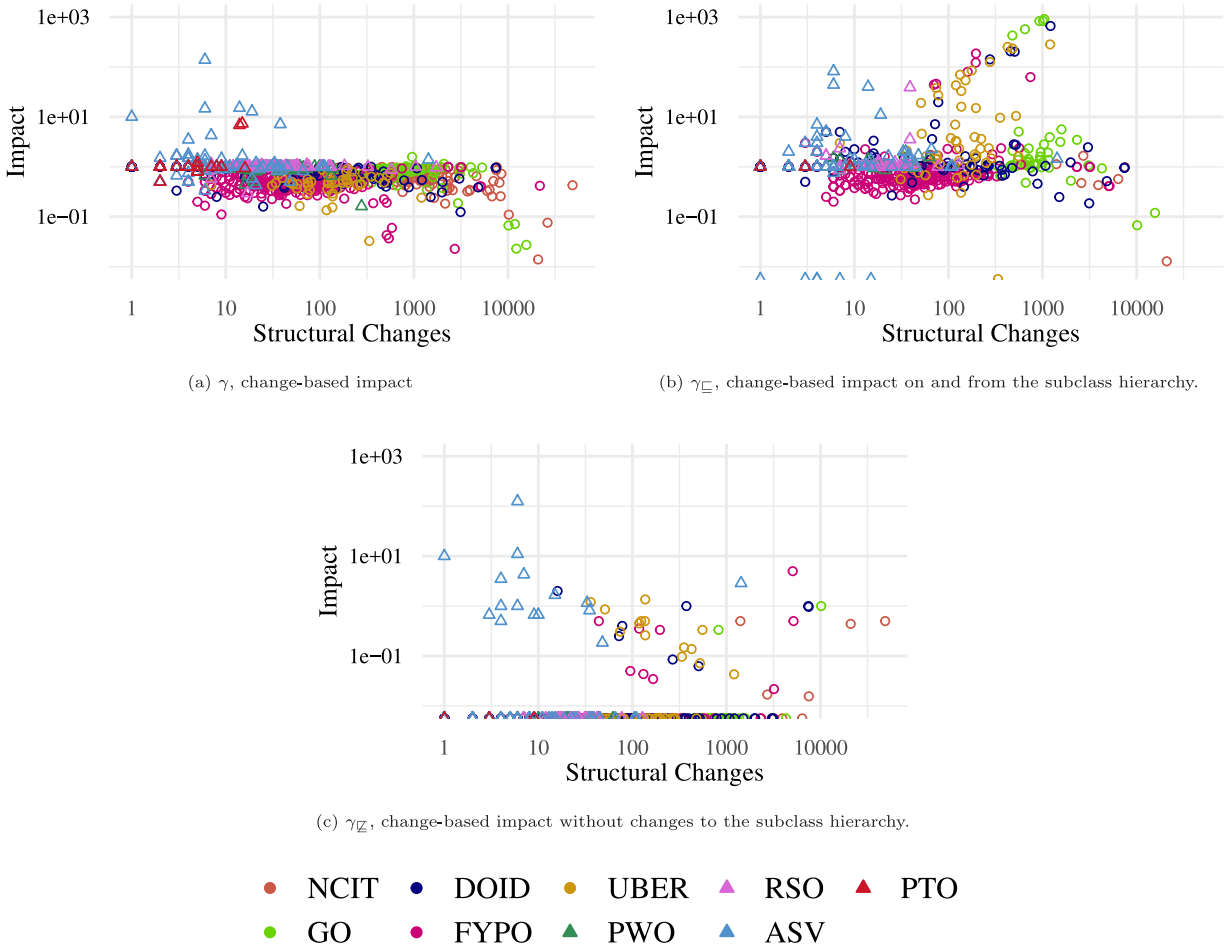
Fig. 6. Scatter plots showing the size-based impact (y-axis), where the nominator denotes change to the materialisation (Δ_i) and the denominator is the size of the unchanged part of the materialisation ($M_i \cap M_{i+1}$). The x-axis shows structural ($|\delta_i|$) or relative changes. Both axes are in logarithmic scale. Legend applies to all plots.

Table 6
Intercept and Slope of LM displayed in Figs. 6(c) and 6(d) and the correlation between σ and number of structural as well as relative changes.

	Linear regression				Spearman correlation	
	Structural changes		Relative changes		Structural changes	Relative changes
Ontology	Intercept	Slope	Intercept	Slope	$\sigma \sim \lvert\delta_i\rvert$	$\sigma \sim {}^{\lvert\delta_i\rvert}/_{\lvert O\rvert}$
	0.0069 (0.002)*	0.00001 (0.000)*	−0.0006 (0.690)	0.95437 (0.000)*	0.39683 (0.000)*	0.80844 (0.000)*
NCIT	−0.0071 (0.060)	0.00001 (0.000)*	0.0024 (0.460)	0.62029 (0.000)*	0.86687 (0.000)*	0.91932 (0.000)*
GO	0.0103 (0.000)*	0.00000 (0.002)*	0.0107 (0.000)*	0.08543 (0.006)*	0.85432 (0.000)*	0.87819 (0.000)*
DOID	−0.0122 (0.044)*	0.00007 (0.000)*	−0.0072 (0.041)*	0.67460 (0.000)*	0.98059 (0.000)*	0.98629 (0.000)*
FYPO	−0.0003 (0.690)	0.00004 (0.000)*	−0.0009 (0.130)	1.71652 (0.000)*	0.95404 (0.000)*	0.95739 (0.000)*
UBER	0.0019 (0.063)	0.00003 (0.000)*	0.0015 (0.120)	0.70530 (0.000)*	0.83493 (0.000)*	0.83864 (0.000)*
PWO	0.0053 (0.002)*	0.00032 (0.000)*	0.0037 (0.004)*	0.69446 (0.000)*	0.93294 (0.000)*	0.98742 (0.000)*
RSO	−0.0038 (0.000)*	0.00032 (0.000)*	−0.0010 (0.018)*	1.16621 (0.000)*	0.98873 (0.000)*	0.99293 (0.000)*
ASV	−0.0021 (0.620)	0.00125 (0.000)*	−0.0074 (0.110)	2.20945 (0.000)*	0.92498 (0.000)*	0.89882 (0.000)*
PTO	−0.0015 (0.029)*	0.00095 (0.000)*	−0.0014 (0.075)	1.50841 (0.000)*	0.91468 (0.000)*	0.95744 (0.000)*

Table 7
Regression and Correlation between γ, $\gamma_{\sqsubseteq}$ and $\gamma_{\not\sqsubseteq}$ as indicated in the header of the table. There are not enough points for PTO to calculate and PWO and RSO only have 0 for $\gamma_{\not\sqsubseteq}$.

	Regression			Correlation	
Ontology	Intercept	$\gamma_{\sqsubseteq}$	$\gamma_{\not\sqsubseteq}$	$\gamma \sim \gamma_{\sqsubseteq}$	$\gamma \sim \gamma_{\not\sqsubseteq}$
	0.718956 (0.000)*	−0.000036 (0.283)	1.099227 (0.000)*	0.460669 (0.000)*	−0.103911 (0.008)*
NCIT	0.494509 (0.009)*	0.000008 (0.631)	−0.470080 (0.367)	0.761905 (0.037)*	0.240981 (0.565)
GO	0.829993 (0.000)*	0.000053 (0.379)	−0.685616 (0.001)*	0.603774 (0.000)*	−0.304631 (0.028)*
DOID	0.754517 (0.000)*	−0.000364 (0.080)	0.080905 (0.403)	0.183767 (0.101)	−0.096839 (0.390)
FYPO	0.533731 (0.000)*	−0.000011 (0.614)	−0.022965 (0.564)	0.896273 (0.000)*	−0.052464 (0.438)
UBER	0.515435 (0.000)*	−0.000373 (0.384)	−0.090611 (0.327)	0.076109 (0.594)	−0.088682 (0.536)
PWO	0.857426 (0.000)*	0.117145 (0.023)*	NA	0.471399 (0.001)*	NA
RSO	0.998021 (0.000)*	0.000077 (0.873)	NA	0.177227 (0.142)	NA
ASV	0.663118 (0.000)*	0.212025 (0.000)*	0.959593 (0.000)*	0.016188 (0.866)	0.107977 (0.259)
PTO	1.000000 (0.000)*	NA	NA	NA	NA



(a) γ, change-based impact

(b) $\gamma_{\sqsubseteq}$, change-based impact on and from the subclass hierarchy.

(c) $\gamma_{\not\sqsubseteq}$, change-based impact without changes to the subclass hierarchy.

Fig. 7. Scatter plots showing the change-based impact (y-axis) where the nominator denotes changes to the materialisation (Δ_i) and the denominator are changes on the ontology (δ_i). The number of structural changes ($|\delta_i|$) are on the x-axis. Both axes are shown in logarithmic scale. Legend applies to all plots.

γ, but still around 1.0, and it also displays more outliers above 1.0. Therefore, the changes on the class hierarchy have a high contribution to the overall impact. We can confirm this with the scatter plot in Fig. 7(c), which plots $\gamma_{\not\sqsubseteq}$ and shows impact without the class hierarchy. All points situated at the bottom of this plot have no impact besides that of the class hierarchy, because their $\gamma_{\not\sqsubseteq}$ is equal to 0.0. The remaining points show ontology version pairs which experience impact not connected to the hierarchy. This impact is mostly below 1.0 as well, showing that other changes have less influence on the materialisation.

Further investigation into the characterisation of the points between $10e^{-2}$ and $10e^{2}$ in Fig. 7(c) was not fruitful. No feature emerged that would identify such ontology version pairs, which experience high impact outside of the hierarchy. We investigated other common ontology measures known from works like [55, 56,58], such as property richness, annotation richness, or class-property ratio. We also analysed the axioms in δ_i, ignoring general concept inclusion axioms with `type` and `SubClassOf` relations, but we found no common changes among these points. There are only 58 ontology pairs in this range (out of 1'090), and most of such points relate to ASV and UBER. This behaviour is therefore seldom, and in the cases of PWO, RSO, and PTO it never happens. We also observe that in all the ontologies, the most common changes are related to general concept inclusions, mostly `type` triples, followed by `subClassOf` ones. Finally, the ontologies we consider are biomedical $\mathcal{EL}^{++}$ ontologies, which typically have a strong focus on the class hierarchy relations and have no or a small ABox.

Further, We calculated regressions and correlation to investigate the observation of most impact being related to the subclass hierarchy. Results are shown in Table 7 with the regression on the left side and correlation on the right side. The first line shows the results over all ontologies combined. The regression target is γ with $\gamma_{\sqsubseteq}$ and $\gamma_{\not\sqsubseteq}$ as input. Unfortunately, the regressions are not significant in most cases. This looks different when looking at the correlation, where some ontologies show a significant result between γ and $\gamma_{\sqsubseteq}$. Even though the results are slightly ambiguous for individual ontologies, the results over all ontologies still suggests that class hierarchy is the deciding factor for impact on the materialisation. Hence, we conclude that the changes to the hierarchy are the ones that most prevailingly influence the materialisation. This confirms our initial intuition of creating impact measures tailored to class inheritance.

Additionally, in Fig. 7(a), and therefore for γ, there is no clear distinction between the different sizes of ontologies. This is in contrast to the split we observe for σ in Fig. 6(a).

We also use the change classification from COnto-Diff [14] for a further analysis. We calculate a linear model to learn the influence of the different types of changes on the impact. Unfortunately, no change type emerged across ontologies, for which we can say it has a significant influence on the materialisation impact. For some ontologies one or two types were significant, but most of the time they differed from each other. Results are presented in the Appendix in Tables A.9–A.12.

Through the analyses we presented in this section, we can answer research ***RQ2***: *What are measures that are simple to compute and allow to assess the impact of the ontology change on the ontology's materialisation?* Our analyses support the fact that the impact measures capture different aspects of the consequences of

changes: where one set focuses on the size of the ontology (σ and $\sigma_{\sqsubseteq}$) and the other on the number of edits (γ, $\gamma_{\sqsubseteq}$, and $\gamma_{\not\sqsubseteq}$). σ and $\sigma_{\sqsubseteq}$ show similar behaviour and no influence when only regarding the hierarchy of the ontology. On the contrary, γ and $\gamma_{\sqsubseteq}$ display differences in the scatter plots. σ and $\sigma_{\sqsubseteq}$ are heavily influenced by the size, penalising the impact for large ontologies when small part of axioms changes. Due to their design, γ and $\gamma_{\sqsubseteq}$ do not have this problem. The measures are also simple to compute. We used Ecco to calculate the difference between two materialisations. We have also looked into the implementation within Protégé [15] and will include the impact measures in the ChImp plugin [16].

To conclude, we analyse research ***RQ3***: *What aspects of changes and which complex change types have the largest share in the impact on the materialisation?* This question concerns the ontology evolution's real consequences on the nine chosen OBO ontologies. Once an ontology exceeds roughly 6'500 classes, we consider it to be a large ontology. This clear distinction is visible in Figs. 6(a) and 6(b). When using σ and $\sigma_{\sqsubseteq}$ as impact measures, large ontologies require more impactful changes to reach comparable impact. However, relating the changes to the size, we found a high correlation between the relative changes in the impact for all ontologies. At the same time, the relationship reported is similar for all ontologies, which allows us to generalise to other ontologies with a similar profile. When considering γ and $\gamma_{\sqsubseteq}$, we observe that most impact comes from hierarchy changes and also consequences mostly on the hierarchy within the materialisation. When hierarchy is not taken into consideration, we observed that less than 2% of the ontology versions experience impact. Therefore, we identify the evolution of the subclass hierarchy as having the largest share in the impact on the materialisation. This finding is in line with related work, where the features capturing hierarchy relations were performing the best in the change prediction task [37,50]. Surprisingly, no specific change actions emerged to have a significant influence on the materialisation impact. However, these change actions do not distinguish between changes on the subclass hierarchy and other changes. The type of change, where the addition or deletion occurs does not reliably signal impact on the materialisation. The changes on the subclass hierarchy are much more dominant in this regard.

6. Limitations and future work

The largest threat to validity is the analysis of *only* nine ontologies, when there are roughly 877 ontologies[7] available in BioPortal [68]. However, because of the selection criteria and computational intensity of this research, we could not select an excessive amount of ontologies. Many ontologies available online do not openly share their evolution, but rather just the recent release. Additionally, a more specific investigation has to be conducted on the ABox, its changes and their impact on the materialisation. Our chosen ontologies did not include ABoxes and they are also rare in the biomedical domain. Hence, we cannot draw any conclusions about ontologies which make use of ABoxes. In our future work, we will extend the analysis towards other types of ontologies, investigating both other description logics and domains outside of biomedicine, but our constraint on the availability of the evolution of an ontology remains.

Our analyses show that every ontology has its unique characteristics, even though they are from the same domain and mostly the same description logic. This supports recent claims that ontology engineering and ontology evolution is still an open research area that requires novel and more supporting instruments [69].

In this study, we do not focus on the analysis of the outliers, as we are interested in the overall picture, as in previous studies such as Gonçalves et al. [70] in the case of NCIT in 2011, and Gross et al. on GO [4] in 2012. Such detailed analyses are still important and necessary, as they may let properties and findings emerge that are complementary to the ones we found in our study.

Lastly, even if the impact measures we introduced are effective to carry out our analyses, they need to be validated and evaluated by the ontology engineering community. As we have added the measures in our ChImp Protégé plugin [16] our next step includes a detailed user-study on the usefulness and informativeness of the introduced measures. Additionally, this will give us the opportunity to study the awareness of change consequences in general. We are interested in studying if such measures are useful to help users to better understand the impact of their changes, and in changing their behaviour.

7. Conclusions

It is a known fact that knowledge evolves constantly. As we capture knowledge in ontologies, we inherently have to deal with its evolution. As ontologies are often built to power subsequent applications, and the ontology evolution also impacts the applications built on top of such ontologies. In this work, we studied how ontology evolution affects their materialisation. Our goal was two-fold: we defined materialisation impact measures and analysed the relationship between ontology evolution and the impact using the proposed measures. We carried out this analysis for nine OBO ontologies, two of which are very large and well established. At this point, we want to advocate for more transparency in the editing process and ask ontology maintainers and engineers to share the editing history of their ontologies.

Our primary contribution and the answer to our research ***RQ2*** is the definition of five impact measures, divided into two sets. The first set concerns the consequences of changes on the materialisation concerning the materialisation's size. The second set focuses on both the number of changes to the materialisation and to the ontology.

To start the analysis, we took a closer look at the evolution of ontologies themselves, which acted as a stepping stone to understand the impact on the materialisation. We found that even though we observed more changes being applied to already large ontologies, we could not confirm this statistically (***RQ1.1***). Additionally, using the relative number of changes to the size of the ontology, no such distinction could be observed at all. We investigated the types of changes in detail as well and found that the most common complex change actions are addition of leaves, moves, and changes of attributes (***RQ1.2***) classified using the rules defined by Hartung et al. [14].

We then continued to use the proposed materialisation impact measures ***RQ2*** to investigate the influence of the evolution on the materialisation. The first set of impact measures displays a clear division based on the size of the ontologies, which allows us to conclude that smaller ontologies experience a more considerable impact on their materialisation regarding their size. Once we correct for the size of the ontology and use relative instead of absolute changes, no distinction between ontologies is visible. However, there is a very strong correlation between the number of changes and the impact, meaning that more changes also mean more impact on the ontology. With the help of the second set of impact measures, we found that most impact on the materialisation is related to the hierarchy of the ontology, except for some noise. This is also supported by the first set of impact measures, as no difference could be found between the general impact and the impact on the class hierarchy, showing the same distributions. Also taking into account the complex changes calculated using COnto-Diff [14], we were unable to identify commonalities among the ontologies, meaning that no

[7] https://bioportal.bioontology.org/ — Accessed on 29/05/2021.

single or multiple change types have a significant share in the impact on the materialisation. Therefore, we conclude that OBO ontologies experience changes and their impact mostly on the class hierarchy within the ontology or materialisation but not specific to complex change actions. These findings answer ***RQ3***.

Knowing that some changes or the accumulation thereof are causing a tremendous impact, could signal the ontology user that they need to update their version and recalculate tasks based on the new version of the ontology. There is significant potential in future work following this research for investigating the awareness of change effect with ontology engineers. Additionally, because of the simplicity of the measures, they can easily be incorporated into ontology editors as well as release notes. It is open to see if knowledge about impact can improve processes and communication of ontology evolution between engineers and users. Hopefully, with time, we will be able to observe such improvements, which could result not only in better ontologies but also in a better usage of ontologies in other applications.

CRediT authorship contribution statement

Romana Pernisch: Conceptualization, Methodology, Software, Investigation, Writing – original draft, Visualization, Validation, Data curation. **Daniele Dell'Aglio:** Conceptualization, Methodology, Writing — review & editing, Supervision. **Abraham Bernstein:** Conceptualization, Methodology, Writing – review & editing, Supervision, Funding acquisition.

Declaration of competing interest

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

Acknowledgements

We thank the Swiss National Science Foundation for their partial support under contract number 407550_167177. We also thank the anonymous reviewers for their insightful feedback, and Luca Rossetto and Florian Ruosch for their comments and discussions.

Appendix A. COnto-Diff tables

The following tables are addressed in Section 5 but not discussed in detail.

Table A.8
COnto-Diff complex changes relative to the total amount of complex changes in median, mean, and standard deviation for each change type and ontology with 0s, when a change action was not present in compared versions (instead of NA).

		Total	ChgAttvalue	AddLeaf	Move	AddInner	AddSubGraph	Split	DelInner	DelLeaf
	min	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	median	604.000	0.000	0.718	0.133	0.058	0.026	0.006	0.000	0.000
NCIT	max	41'531.000	0.000	0.966	0.878	0.292	0.165	0.073	0.363	0.637
	mean	1'039.556	0.000	0.694	0.190	0.072	0.037	0.008	0.002	0.005
	sd	3'136.318	0.000	0.167	0.159	0.049	0.036	0.008	0.027	0.049
	min	12.000	0.000	0.006	0.016	0.000	0.000	0.000	0.000	0.000
	median	279.000	0.000	0.357	0.348	0.157	0.077	0.007	0.000	0.000
GO	max	2'004.000	0.971	0.712	0.899	0.289	0.191	0.074	0.000	0.000
	mean	321.345	0.048	0.342	0.385	0.148	0.076	0.009	0.000	0.000
	sd	264.110	0.153	0.149	0.214	0.069	0.044	0.011	0.000	0.000
	min	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	median	31.000	0.121	0.298	0.083	0.001	0.000	0.000	0.000	0.000
DOID	max	5'322.000	1.000	1.000	1.000	0.437	0.167	0.167	0.500	0.833
	mean	172.606	0.304	0.381	0.207	0.031	0.019	0.014	0.019	0.039
	sd	718.885	0.352	0.325	0.274	0.061	0.036	0.031	0.076	0.134
	min	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	median	17.000	0.063	0.373	0.121	0.102	0.026	0.000	0.000	0.000
FYPO	max	5'499.000	1.000	1.000	1.000	1.000	0.250	0.200	1.000	0.623
	mean	70.952	0.185	0.391	0.195	0.127	0.048	0.004	0.050	0.004
	sd	367.108	0.244	0.285	0.228	0.135	0.059	0.019	0.140	0.041
	min	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	median	16.000	0.000	0.356	0.404	0.067	0.018	0.000	0.000	0.000
UBER	max	8'975.000	1.000	1.000	1.000	0.500	0.167	1.000	0.490	1.000
	mean	89.816	0.123	0.344	0.408	0.070	0.029	0.019	0.003	0.024
	sd	591.625	0.276	0.242	0.282	0.070	0.035	0.081	0.036	0.119
	min	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	median	16.500	0.007	0.427	0.020	0.211	0.143	0.000	0.000	0.000
PWO	max	95.000	1.000	1.000	0.514	0.533	0.286	1.000	0.000	0.222
	mean	25.676	0.173	0.401	0.084	0.208	0.131	0.023	0.000	0.002
	sd	23.210	0.266	0.217	0.122	0.121	0.075	0.110	0.000	0.022
	min	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	median	9.500	0.143	0.400	0.000	0.023	0.000	0.000	0.000	0.000
RSO	max	4'644.000	1.000	1.000	0.845	1.000	0.400	Inf	0.079	0.333
	mean	64.580	0.266	0.446	0.113	0.110	0.058	Inf	0.001	0.007
	sd	399.612	0.324	0.334	0.185	0.152	0.087		0.007	0.043
	min	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	median	17.000	0.906	0.036	0.000	0.000	0.000	0.000	0.000	0.000
ASV	max	933.000	1.000	1.000	0.692	0.686	0.200	0.067	0.478	0.653
	mean	31.884	0.754	0.135	0.024	0.048	0.017	0.001	0.008	0.014
	sd	99.167	0.323	0.241	0.089	0.111	0.034	0.008	0.054	0.072
	min	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	median	4.000	0.700	0.000	0.000	0.000	0.000	0.000	0.000	0.000
PTO	max	344.000	1.000	1.000	1.000	0.333	0.303	0.101	0.250	1.000
	mean	15.624	0.613	0.127	0.213	0.016	0.007	0.001	0.002	0.024
	sd	45.706	0.393	0.277	0.312	0.051	0.036	0.010	0.024	0.137

Table A.9
Results of the linear model to estimate the **size impact** σ using the complex actions classified by COnto-Diff.

		(Intercept)	Total	ChgAttvalue	AddLeaf	Move	AddInner	AddSubGraph	Split	DelInner	DelLeaf
NCIT	estimate	−0.052	**0.000**		0.052	0.079	0.129	−0.072	**0.432**	9.472	
	p.value	0.671	**0.038**		0.667	0.519	0.316	0.569	**0.016**	0.681	
GO	estimate	−0.002	**0.000**	−0.026	−0.003	0.014		0.021	−0.051		
	p.value	0.914	**0.000**	0.125	0.855	0.418		0.622	0.403		
DOID	estimate	0.342	**0.000**	**−0.330**	**−0.327**	**−0.337**	**1.241**	**−1.690**	−0.032	−0.328	**−0.334**
	p.value	0.000	**0.008**	**0.000**	**0.000**	**0.000**	**0.015**	**0.008**	0.929	0.091	**0.002**
FYPO	estimate	0.003	0.000	0.017	−0.001	0.009	0.014	−0.014	−0.058	−0.003	0.016
	p.value	0.812	0.826	0.276	0.941	0.565	0.614	0.808	0.689	0.907	0.870
UBER	estimate	0.002	**0.000**	−0.003	−0.005	0.001	−0.004	−0.003	0.000		
	p.value	0.525	**0.000**	0.407	0.236	0.862	0.562	0.814	0.979		
PWO	estimate	**−0.024**	**0.001**	**0.011**	0.031	**0.058**	0.039	−0.025	0.017		
	p.value	**0.585**	**0.000**	**0.811**	0.486	**0.218**	0.370	0.621	0.083		
RSO	estimate	0.000	**0.001**	**−0.015**	−0.004	**−0.028**	0.076	**−0.145**	−0.031	0.458	−0.079
	p.value	0.996	**0.000**	**0.639**	0.891	**0.407**	0.058	**0.005**	0.205	0.310	0.367
ASV	estimate	**0.046**	0.000	−0.018	−0.047	−0.080	−0.051	−0.054	−0.110	0.041	−0.006
	p.value	**0.246**	0.878	0.683	0.560	0.743	0.781	0.917	0.947	0.924	0.984
PTO	estimate	−0.110	**−0.010**	0.324	0.154	0.098	−0.057	0.667			
	p.value	0.908	**0.046**	0.735	0.873	0.918	0.957	0.657			

Table A.10
Results of the linear model to estimate the **subclass size impact** $\sigma_{\sqsubseteq}$ using the complex actions classified by COnto-Diff.

		(Intercept)	Total	ChgAttvalue	AddLeaf	Move	AddInner	AddSubGraph	Split	DelInner	DelLeaf
NCIT	estimate	−0.052	**0.000**		0.053	0.079	0.129	−0.071	**0.431**	9.466	
	p.value	0.668	**0.038**		0.664	0.517	0.314	0.573	**0.016**	0.681	
GO	estimate	−0.002	**0.000**	−0.026	−0.003	0.014		0.021	−0.051		
	p.value	0.911	**0.000**	0.126	0.858	0.415		0.620	0.404		
DOID	estimate	**0.341**	**0.000**	**−0.328**	**−0.326**	**−0.336**	**1.235**	**−1.683**	−0.032	−0.327	**−0.333**
	p.value	**0.000**	**0.008**	**0.000**	**0.000**	**0.000**	**0.015**	**0.008**	0.930	0.091	**0.002**
FYPO	estimate	0.003	0.000	0.017	−0.001	0.009	0.013	−0.014	−0.058	−0.003	0.016
	p.value	0.812	0.827	0.276	0.941	0.564	0.615	0.810	0.688	0.905	0.870
UBER	estimate	0.000	**0.000**	−0.002	−0.003	0.002	−0.002	−0.002	0.000		
	p.value	0.895	**0.000**	0.599	0.467	0.535	0.808	0.846	0.994		
PWO	estimate	**−0.024**	**0.001**	**0.011**	0.031	**0.059**	0.039	−0.025	0.017		
	p.value	**0.585**	**0.000**	**0.810**	0.486	**0.217**	0.371	0.622	0.084		
RSO	estimate	0.000	**0.001**	**−0.015**	−0.004	**−0.028**	0.076	**−0.145**	−0.031	0.458	−0.079
	p.value	0.996	**0.000**	**0.639**	0.891	**0.407**	0.058	**0.005**	0.206	0.310	0.367
ASV	estimate	**0.007**	0.000	0.010	−0.004	−0.024	−0.007	−0.030	0.042	0.079	0.044
	p.value	**0.734**	0.719	0.677	0.928	0.855	0.943	0.915	0.962	0.734	0.801
PTO	estimate	−0.101	**−0.010**	0.309	0.144	0.089	−0.053	0.652			
	p.value	0.914	**0.046**	0.741	0.878	0.924	0.959	0.657			

Table A.11
Results of the linear model to estimate the **change impact** γ using the complex actions classified by COnto-Diff.

		(Intercept)	Total	ChgAttvalue	AddLeaf	Move	AddInner	AddSubGraph	Split	DelInner	DelLeaf
NCIT	estimate	0.075	0.000		0.616	0.569	0.182	0.881	3.208	−217.487	
	p.value	0.963	0.273		0.706	0.729	0.916	0.605	0.181	0.483	
GO	estimate	0.060	0.000	0.586	0.827	0.894		1.230	−0.749		
	p.value	0.895	0.381	0.213	0.101	0.067		0.312	0.665		
DOID	estimate	**0.914**	**0.000**	0.019	−0.155	**−0.374**	0.975	−0.606	0.097	0.025	−0.450
	p.value	**0.000**	**0.042**	0.913	0.383	**0.041**	0.429	0.693	0.913	0.957	0.079
FYPO	estimate	**0.582**	0.000	−0.049	−0.094	−0.048	−0.027	0.006	−0.884	−0.127	−0.299
	p.value	**0.000**	0.228	0.445	0.110	0.465	0.799	0.978	0.130	0.185	0.440
UBER	estimate	−0.259	0.001	0.419	0.658	**0.846**	1.031	0.681	−1.388		
	p.value	0.384	0.110	0.188	0.064	**0.009**	0.059	0.468	0.095		
PWO	estimate	0.583	0.000	0.175	0.448	0.412	0.125	0.670	0.127		
	p.value	0.330	0.567	0.769	0.456	0.519	0.834	0.330	0.342		
RSO	estimate	**1.000**	0.000	−0.039	0.010	0.023	−0.031	0.007	0.036	−0.278	0.065
	p.value	**0.000**	0.860	0.721	0.924	0.837	0.816	0.967	0.669	0.855	0.826
ASV	estimate	**13.149**	−0.011	**−11.695**	−10.569	−13.058	−10.942	−10.335	−14.045	−7.723	−14.004
	p.value	**0.000**	0.892	**0.005**	0.112	0.641	0.499	0.805	0.965	0.848	0.597
PTO	estimate	−5.229	0.027	7.542	5.674	6.695	−3.077	0.249			
	p.value	0.807	0.831	0.727	0.794	0.754	0.897	0.994			

Table A.12
Results of the linear model to estimate the **subclass change impact** $\gamma_{\sqsubseteq}$ using the complex actions classified by COnto-Diff.

		(Intercept)	Total	ChgAttvalue	AddLeaf	Move	AddInner	AddSubGraph	Split	DelInner	DelLeaf
NCIT	estimate	−3926.145	0.200		4007.178	4012.148		−1685.175			
	p.value										
GO	estimate	−291.050	**1.602**	664.451	250.391	206.909		−2770.407	1411.984		
	p.value	0.852	**0.000**	0.708	0.884	0.900		0.482	0.785		
DOID	estimate	0.979	**0.120**	**−29.701**	−35.928	**228.445**	−1267.714	1271.949	229.436	−373.248	−6.294
	p.value	0.991	**0.040**	**0.748**	0.705	**0.020**	0.057	0.123	0.629	0.144	0.963
FYPO	estimate	23.697	0.035	290.789	−45.488	−56.870	−37.542	163.949	−978.889	−69.989	−346.372
	p.value	0.904	0.878	0.238	0.839	0.824	0.931	0.851	0.728	0.862	0.938
UBER	estimate	−5.535	**0.675**	3.279	1.786	45.251	−95.799	−137.676	−106.980		
	p.value	0.961	**0.011**	0.978	0.989	0.706	0.641	0.702	0.734		
PWO	estimate	**0.941**	0.000	**−0.114**	0.060	**2.227**		0.174	0.120		
	p.value	**0.000**	0.845	**0.543**	0.743	**0.000**		0.704	0.261		
RSO	estimate	1.000	0.004	**−0.613**	−0.212	**29.348**	−5.373	2.086	3.042		0.870
	p.value	0.723	0.852	**0.847**	0.941	**0.000**	0.181	0.698	0.232		0.918
ASV	estimate	**10.685**	−0.005	−8.471	−9.248	−15.052	−9.253	−12.846	−16.511	−14.302	−1.422
	p.value	**0.000**	0.939	0.008	0.067	0.490	0.449	0.691	0.946	0.662	0.961
PTO	estimate	1.000	0.000		0.000						
	p.value										

References

[1] N.F. Noy, M. Klein, Ontology evolution: not the same as schema evolution, Know. Inf. Sys. 6 (4) (2004) 428–440.
[2] M. Bada, R. Stevens, C.A. Goble, Y. Gil, M. Ashburner, J.A. Blake, J.M. Cherry, M.A. Harris, S. Lewis, A short study on the success of the gene ontology, Web Semantics 1 (2) (2004) 235–240.
[3] G. Fragoso, S. de Coronado, M. Haber, F. Hartel, L. Wright, Overview and utilization of the NCI thesaurus, Int. J. Genomics 5 (8) (2004) 648–654.
[4] A. Gross, M. Hartung, K. Prüfer, J. Kelso, E. Rahm, Impact of ontology evolution on functional analyses, Bioinformatics 28 (20) (2012) 2671–2677.
[5] E.Y. Chen, C.M. Tan, Y. Kou, Q. Duan, Z. Wang, G.V. Meirelles, N.R. Clark, A. Ma'ayan, Enrichr: Interactive and collaborative HTML5 gene list enrichment analysis tool, BMC Bioinformatics 14 (1) (2013) 128.
[6] P. Gaudet, N. Škunca, J.C. Hu, C. Dessimoz, Primer on the gene ontology, in: The Gene Ontology Handbook, Humana Press, New York, NY, 2017, pp. 25–37.
[7] M.A. Harris, A. Lock, J. Bähler, S.G. Oliver, V. Wood, FYPO: The fission yeast phenotype ontology, Bioinformatics 29 (13) (2013) 1671–1678.
[8] C.J. Mungall, C. Torniai, G.V. Gkoutos, S.E. Lewis, M.A. Haendel, Uberon, an integrative multi-species anatomy ontology, Genome Biology 13 (1) (2012) R5.
[9] L.M. Schriml, E. Mitraka, J. Munro, B. Tauber, M. Schor, L. Nickle, V. Felix, L. Jeng, C. Bearer, R. Lichenstein, K. Bisordi, N. Campion, B. Hyman, D. Kurland, C.P. Oates, S. Kibbey, P. Sreekumar, C. Le, M. Giglio, C. Greene, Human disease ontology 2018 update: Classification, content and workflow expansion, Nucleic Acids Research 47 (D1) (2019) D955–D962.
[10] V. Petri, P. Jayaraman, M. Tutaj, G.T. Hayman, J.R. Smith, J. De Pons, S.J. Laulederkind, T.F. Lowry, R. Nigam, S.-J. Wang, M. Shimoyama, M.R. Dwinell, D.H. Munzenmaier, E.A. Worthey, H.J. Jacob, The pathway ontology – updates and applications, Biomedical Semantics 5 (1) (2014) 7.
[11] R. Nigam, D.H. Munzenmaier, E.A. Worthey, M.R. Dwinell, M. Shimoyama, H.J. Jacob, Rat strain ontology: Structured controlled vocabulary designed to facilitate access to strain data at RGD, Biomedical Semantics 4 (1) (2013) 36.
[12] W.R. Hogan, M.M. Wagner, M. Brochhausen, J. Levander, S.T. Brown, N. Millett, J. DePasse, J. Hanna, The apollo structured vocabulary: An OWL2 ontology of phenomena in infectious disease epidemiology and population biology for use in epidemic simulation, Biomedical Semantics 7 (1) (2016) 50.
[13] T.P.O. Consortium, The plant ontology™ consortium and plant ontologies, Comp. Funct. Genomics 3 (2) (2002) 137–142.
[14] M. Hartung, A. Gross, E. Rahm, Conto-diff: Generation of complex evolution mappings for life science ontologies, J. Biomed. Inform. 46 (1) (2013) 15–32.
[15] M.A. Musen, The Protégé project: A look back and a look forward, AI Matters 1 (4) (2015) 4–12.
[16] R. Pernisch, M. Serbak, D. Dell' Aglio, A. Bernstein, Chimp: visualizing ontology changes and their impact in protégé, in: Visualization and Interaction for Ontologies and Linked Data, Co-Located with ISWC2020, CEUR-WS.org, 2020.
[17] R. Pernischova, D. Dell'Aglio, M. Horridge, M. Baumgartner, A. Bernstein, Towards predicting impact of changes in evolving knowledge graphs, in: Posters & Demonstrations, Springer, Aukland, NZ, 2019.
[18] F. Baader, S. Brandt, C. Lutz, Pushing the EL envelope, in: IJCAI, Professional Book Center, 2005, pp. 364–369.
[19] B. Glimm, I. Horrocks, B. Motik, G. Stoilos, Z. Wang, Hermit: an OWL 2 reasoner, J. Autom. Reason. 53 (3) (2014) 245–269.
[20] D. Tsarkov, I. Horrocks, Fact++ description logic reasoner: system description, in: Proceedings of the Third International Joint Conference on Automated Reasoning, IJCAR'06, Springer, Berlin, Heidelberg, 2006, pp. 292–297.
[21] Y. Kazakov, M. Krötzsch, F. Simančík, The incredible ELK: From polynomial procedures to efficient reasoning with ontologies, J. Autom. Reason. 53 (1) (2013) 1–61.
[22] F. Baader, C. Lutz, B. Suntisrivaraporn, CEL—a polynomial-time reasoner for life science ontologies, in: U. Furbach, N. Shankar (Eds.), Proceedings of the 3rd International Joint Conference on Automated Reasoning, IJCAR'06, in: Lecture Notes in Artificial Intelligence, vol. 4130, Springer-Verlag, 2006, pp. 287–291.
[23] E. Thomas, J.Z. Pan, Y. Ren, Trowl: tractable OWL 2 reasoning infrastructure, in: ESWC (2), in: Lecture Notes in Computer Science, vol. 6089, Springer, 2010, pp. 431–435.
[24] B. Motik, Y. Nenov, R.E.F. Piro, I. Horrocks, Incremental update of datalog materialisation: The backward/forward algorithm, in: AAAI, AAAI Press, 2015, pp. 1560–1568.
[25] E. Sirin, B. Parsia, B.C. Grau, A. Kalyanpur, Y. Katz, Pellet: A practical OWL-DL reasoner, Web Semantics 5 (2) (2007) 51–53.
[26] Y. Ren, J.Z. Pan, Optimising ontology stream reasoning with truth maintenance system, in: CIKM, ACM, 2011, pp. 831–836.
[27] F. Zablith, G. Antoniou, M. d'Aquin, G. Flouris, H. Kondylakis, E. Motta, D. Plexousakis, M. Sabou, Ontology evolution: A process-centric survey, Knowl. Eng. Rev. 30 (1) (2015) 45–75.
[28] M. Hartung, J.F. Terwilliger, E. Rahm, Recent advances in schema and ontology evolution, in: Schema Matching and Mapping, in: Data-Centric Systems and Applications, Springer, 2011, pp. 149–190.
[29] M. Rashid, M. Torchiano, G. Rizzo, N. Mihindukulasooriya, O. Corcho, A quality assessment approach for evolving knowledge bases, Semantic Web 10 (2) (2019) 349–383.
[30] M. Quesada-Martínez, A. Duque-Ramos, M. Iniesta-Moreno, J.T. Fernández-Breis, Preliminary analysis of the OBO foundry ontologies and their evolution using oquare, Stud. Health Technol. Inform. 235 (1) (2017) 426–430.
[31] A. Duque-Ramos, M. Boeker, L. Jansen, S. Schulz, M. Iniesta, J.T. Fernández-Breis, Evaluating the good ontology design guideline (GoodOD) with the ontology quality requirements and evaluation method and metrics (OQuaRE), PLoS One 9 (8) (2014).
[32] M. Tury, M. Bielikova, An approach to detection ontology changes, in: ICWE Workshops, in: ACM International Conference Proceeding Series, vol. 155, ACM, 2006, p. 14.
[33] M. Klein, N.F. Noy, A component-based framework for ontology evolution, in: Workshop on Ontologies and Distributed Systems at IJCAI, vol. 3, 2003, p. 4.
[34] V. Papavasileiou, G. Flouris, I. Fundulaki, D. Kotzinos, V. Christophides, High-level change detection in rdf(s) kbs, ACM Transactions on Database Systems 38 (1) (2013) http://dx.doi.org/10.1145/2445583.2445584.
[35] M. Hartung, A. Gross, E. Rahm, CODEX: Exploration of semantic changes between ontology versions, Bioinformatics 28 (6) (2012) 895–896.

[36] G. Flouris, D. Manakanatas, H. Kondylakis, D. Plexousakis, G. Antoniou, Ontology change: Classification and survey, Knowl. Eng Rev. 23 (2) (2008) 117–152.
[37] C. Pesquita, F.M. Couto, Predicting the extension of biomedical ontologies, PLoS Comput. Biol. 8 (9) (2012) 1–16, http://dx.doi.org/10.1371/journal.pcbi.1002630.
[38] S.D. Cardoso, C. Pruski, M. Da Silveira, Supporting biomedical ontology evolution by identifying outdated concepts and the required type of change, J. Biomed. Inform. 87 (2018) 1–11.
[39] A. Meroño"-"Peñuela, E. Kontopoulos, S. Darányi, Y. Kompatsiaris, A study of intensional concept drift in trending DBpedia concepts, in: SEMANTICS Workshops, in: CEUR Workshop Proceedings, vol. 2063, CEUR-WS.org, 2017.
[40] A. Wegmann, F. Lemmerich, M. Strohmaier, Detecting different forms of semantic shift in word embeddings via paradigmatic and syntagmatic association changes, in: International Semantic Web Conference, Springer, 2020, pp. 619–635.
[41] T.G. Stavropoulos, S. Andreadis, E. Kontopoulos, I. Kompatsiaris, SemaDrift: A hybrid method and visual tools to measure semantic drift in ontologies, J. Web Semant. (2018).
[42] G. Capobianco, D. Cavaliere, S. Senatore, OntoDrift: a semantic drift gauge for ontology evolution monitoring, in: MEPDaW@ISWC, 2020.
[43] J. Ekanayake, J. Tappolet, H.C. Gall, A. Bernstein, Tracking concept drift of software projects using defect prediction quality, in: MSR, IEEE Computer Society, 2009, pp. 51–60.
[44] J. Chen, F. Lecue, J.Z. Pan, H. Chen, Learning from ontology streams with semantic concept drift, in: IJCAI, ijcai.org, 2017, pp. 957–963.
[45] J. Gao, W. Fan, J. Han, On appropriate assumptions to mine data streams: Analysis and practice, in: Proceedings of the 7th IEEE International Conference on Data Mining (ICDM 2007), October 28-31, 2007, Omaha, Nebraska, USA, IEEE Computer Society, 2007, pp. 143–152, http://dx.doi.org/10.1109/ICDM.2007.96.
[46] R. Trivedi, H. Dai, Y. Wang, L. Song, Know-evolve: deep temporal reasoning for dynamic knowledge graphs, in: ICML, in: Proceedings of Machine Learning Research, vol. 70, PMLR, 2017, pp. 3462–3471.
[47] R.S. Goncalves, B. Parsia, U. Sattler, Categorising logical differences between OWL ontologies, in: CIKM, ACM, 2011, pp. 1541–1546.
[48] T. Gottron, C. Gottron, Perplexity of index models over evolving linked data, in: ESWC, in: Lecture Notes in Computer Science, vol. 8465, Springer, 2014, pp. 161–175.
[49] J.C. dos Reis, C. Pruski, M. Da Silveira, C. Reynaud-Delaître, Understanding semantic mapping evolution by observing changes in biomedical ontologies, J. Biomed. Inform. 47 (2014) 71–82.
[50] S.D. Cardoso, C. Pruski, M.D. Silveira, Y.-C. Lin, A. Gross, E. Rahm, C. Reynaud-Delaitre, Leveraging the impact of ontology evolution on semantic annotations, in: EKAW, in: Lecture Notes in Computer Science, vol. 10024, 2016, pp. 68–82.
[51] F. Osborne, E. Motta, Pragmatic ontology evolution: reconciling user requirements and application performance, in: International Semantic Web Conference (1), in: Lecture Notes in Computer Science, vol. 11136, Springer, 2018, pp. 495–512.
[52] F. Baader, C. Lutz, S. Brandt, Pushing the EL envelope further, in: OWLED (Spring), in: CEUR Workshop Proceedings, vol. 496, CEUR-WS.org, 2008.
[53] B. Motik, B.C. Grau, I. Horrocks, Z. Wu, A. Fokoue, C. Lutz, OWL 2 Web Ontology Language: Profiles, Technical report, second ed., W3C, 2012.
[54] S. Tartir, I.B. Arpinar, A.P. Sheth, Ontological evaluation and validation, in: Theory and Applications of Ontology: Computer Applications, Springer, 2010, pp. 115–130.
[55] S.L. Tomassen, D. Strasunskas, An ontology-driven approach to web search: Analysis of its sensitivity to ontology quality and search tasks, in: Proceedings of the 11th International Conference on Information Integration and Web-Based Applications & Services, 2009, pp. 130–138.
[56] A.M. Orme, H. Yao, L.H. Etzkorn, Indicating ontology data quality, stability, and completeness throughout ontology evolution, J. Softw. Maint. Evol. Res. Pract. 19 (1) (2007) 49–75.
[57] C. Tempich, R. Volz, Towards a benchmark for semantic web reasoners-an analysis of the DAML ontology library., in: EON, vol. 87, 2003.
[58] R. Djedidi, M.-A. Aufaure, ONTO-EVO A L an ontology evolution approach guided by pattern modeling and quality evaluation, in: International Symposium on Foundations of Information and Knowledge Systems, 2010, pp. 286–305.
[59] B. Lantow, K. Sandkuhl, An analysis of applicability using quality metrics for ontologies on ontology design patterns, Intell. Syst. Account. Finance Manag. 22 (1) (2015) 81–99.
[60] S.S. Sahoo, C. Thomas, A. Sheth, W.S. York, S. Tartir, Knowledge modeling and its application in life sciences: A tale of two ontologies, in: Proceedings of the 15th International Conference on World Wide Web, 2006, pp. 317–326.
[61] B. Motik, P.F. Patel-Schneider, B. Parsia, OWL 2 Web Ontology Language: Structural Specification and Functional-Style Syntax, Technical report, W3C, 2012.
[62] M. Horridge, S. Bechhofer, The OWL api: A java API for OWL ontologies, Semantic Web 2 (1) (2011) 11–21.
[63] R.C. Jackson, J.P. Balhoff, E. Douglass, N.L. Harris, C.J. Mungall, J.A. Overton, ROBOT: A tool for automating ontology workflows, BMC Bioinform. 20 (1) (2019) 407:1–407:10.
[64] R.S. Gonçalves, B. Parsia, U. Sattler, Ecco: A hybrid diff tool for OWL 2 ontologies, in: OWLED, in: CEUR Workshop Proceedings, vol. 849, CEUR-WS.org, 2012.
[65] B. Smith, M. Ashburner, C. Rosse, J. Bard, W. Bug, W. Ceusters, L.J. Goldberg, K. Eilbeck, A. Ireland, C.J. Mungall, et al., The OBO foundry: Coordinated evolution of ontologies to support biomedical data integration, Nat. Biotechnol. 25 (11) (2007) 1251–1255.
[66] R. Pernischova, The butterfly effect in knowledge graphs: Predicting the impact of changes in the evolving web of data, in: Doctoral Consortium, CEUR-WS.org, Aukland, NZ, 2019.
[67] N. Sioutos, S. de Coronado, M.W. Haber, F.W. Hartel, W.-L. Shaiu, L.W. Wright, NCI Thesaurus: A semantic model integrating cancer-related clinical and molecular information, J. Biomed. Inform. 40 (1) (2007) 30–43.
[68] N.F. Noy, N.H. Shah, P.L. Whetzel, B. Dai, M. Dorf, N. Griffith, C. Jonquet, D.L. Rubin, M.-A. Storey, C.G. Chute, M.A. Musen, BioPortal: ontologies and integrated data resources at the click of a mouse, Nucleic Acids Res. 37 (suppl_2) (2009) W170–W173, http://dx.doi.org/10.1093/nar/gkp440.
[69] T. Tudorache, Ontology engineering: Current state, challenges, and future directions, Semantic Web 11 (1) (2020) 125–138.
[70] R.S. Goncalves, B. Parsia, U. Sattler, Analysing the evolution of the NCI Thesaurus, in: 2011 24th International Symposium on Computer-Based Medical Systems, CBMS, 2011, pp. 1–6.