

# SNOMED CT and Basic Formal Ontology – convergence or contradiction between standards? The case of “clinical finding”

Stefan Schulz<sup>a,\*</sup>, James T. Case<sup>b</sup>, Peter Hendler<sup>b</sup>, Daniel Karlsson<sup>c</sup>, Michael Lawley<sup>d</sup>, Ronald Cornet<sup>e,f</sup>, Robert Hausam<sup>g</sup>, Harold Solbrig<sup>h</sup>, Karim Nashar<sup>i</sup>, Catalina Martínez-Costa<sup>j</sup> and Yongsheng Gao<sup>k</sup>

<sup>a</sup> *Medical University of Graz, Austria*

<sup>b</sup> *SNOMED International, London, UK*

<sup>c</sup> *Socialstyrelsen, Stockholm, Sweden*

<sup>d</sup> *Australian e-Health Research Centre, Brisbane, Australia*

<sup>e</sup> *Amsterdam UMC – University of Amsterdam, The Netherlands*

<sup>f</sup> *Amsterdam Public Health, Methodology & Digital Health, Amsterdam, The Netherlands*

<sup>g</sup> *Hausam Consulting, Columbia, Missouri, USA*

<sup>h</sup> *Johns Hopkins University, Rochester, USA*

<sup>i</sup> *National Health Service, Stockport, UK*

<sup>j</sup> *University of Murcia, IMIB-Arrixaca, Spain*

<sup>k</sup> *SNOMED International, London, UK*

## Abstract.

**BACKGROUND:** SNOMED CT is a large terminology system designed to represent all aspects of healthcare. Its current form and content result from decades of bottom-up evolution. Due to SNOMED CT’s formal descriptions, it can be considered an ontology. The Basic Formal Ontology (BFO) is a foundational ontology that proposes a small set of disjoint, hierarchically ordered classes, supported by relations and axioms. In contrast, as a typical top-down endeavor, BFO was designed as a foundational framework for domain ontologies in the natural sciences and related disciplines. Whereas it is mostly assumed that domain ontologies should be created as extensions of foundational ontologies, a post-hoc harmonization of consolidated domain ontologies in use, such as SNOMED CT, is known to be challenging.

**METHODS:** We explored the feasibility of harmonizing SNOMED CT with BFO, with a focus on the SNOMED CT *Clinical Finding* hierarchy. With more than 100,000 classes, it accounts for about one third of SNOMED CT’s content. In particular, we represented typical SNOMED CT finding/disorder concepts using description logics under BFO. Three representational patterns were created and the logical entailments analyzed.

**RESULTS:** Under a first scrutiny, the clinical intuition that diseases, disorders, signs and symptoms form a homogeneous ontological upper-level class appeared incompatible with BFO’s upper-level distinction into continuants and occurrents. The *Clinical finding* class seemed to be an umbrella for all kinds of entities of clinical interest, such as material entities, processes, states, dispositions, and qualities. This suggests the conclusion that *Clinical finding* would not be a suitable upper-level class from an BFO perspective. On closer inspection of the taxonomic links within this hierarchy and the implicit meaning derived thereof, it became clear that *Clinical finding* classes do not characterize the entity (e.g. a fracture, allergy, tumor, pain, hemorrhage, seizure, fever) in a literal sense but rather the condition of a patient having that fracture, allergy, pain etc. This gives sense to the current characteristic of the *Clinical Finding* hierarchy, in which complex classes are modeled as subclasses of their constituents. Most of these taxonomic links are inferred, as the consequence of the ‘role group’ design pattern, which is ubiquitous in SNOMED CT and has often been subject of controversy regarding its semantics.

---

\*Corresponding author. E-mail: [stefan.schulz@medunigraz.at](mailto:stefan.schulz@medunigraz.at).

**CONCLUSION:** Our analyses resulted in the proposal of (i) equating SNOMED CT's 'role group' property with the reflexive and transitive BFO relation '**has occurrent part**'; and (ii) reinterpreting *Clinical Findings* as *Clinical Occurrences*, i.e. temporally extended entities in an organism, having one or more occurrences as temporal parts that occur in continuants. This re-interpretation was corroborated by a manual analysis of classes under *Clinical Finding*, as well as the identification of similar modeling patterns in other ontologies. As a result, SNOMED CT does not require any content redesign to establish compatibility with BFO, apart from this re-interpretation, and a suggested re-labeling. Regarding the feasibility of harmonizing terminologies with principled foundational ontologies post-hoc, our results provide support to the assumption that this does not necessarily require major redesign efforts, but rather a careful analysis of the implicit assumptions of terminology curators and users.

Keywords: Terminology, SNOMED CT, description logics, ontology harmonization

Accepted by: Janna Hastings

## 1. Introduction

### 1.1. Standards of meaning

Standards are agreements that facilitate the exchange of products and the joint participation in practices and operations (International Standards Organisation, 2022). Whereas industry standards, i.e. standards for manufactured entities, are well established, the extension of the standardization idea to natural kinds and to basic categories of being has not met the same level of acceptance yet (Schulz et al., 2018).

Awareness is growing that the creation of industry standards is something close to the practice of ontology and terminology engineering, and that both communities can learn from each other. Good practices as developed by the Applied Ontology community (Guarino and Musen, 2015) could then support discussions about ontology-based standards and support efforts to mutual discussions and collaboration towards interoperability as an ultimate goal.

It often occurs that standards contradict each other. Then, users have the difficult task of using one standard and rejecting the others. But standards can also complement each other. Here, the task is more rewarding. It is then centered on the creation of links between the components of the respective standards, as well as the identification of mappings in their overlapping areas. However, it is common that any two standards to be compared and aligned bear implicit assumptions that challenge interoperation, particularly in cases where they represent different communities with different views regarding the purpose of standardization and their effects on downstream use cases. Here, additional effort is needed to view, understand and re-interpret one standard in the light of the other one. Such an analysis ideally fills interpretation gaps in either standard, and the harmonization task can be described as creating convergence between the standards.

This paper will scrutinize two ontology-based standards in natural science; *viz.* Basic Formal Ontology (BFO) (Arp et al., 2015; Otte et al., 2022) and SNOMED CT (SNOMED CT, 2023). The impetus for this work were discussions within the SNOMED CT community regarding the meaning of certain design patterns that had been judged as underspecified, and reasoning entailments that were seen as questionable. Challenging SNOMED CT by the ontological rigor of a foundational ontology was seen as a useful strategy. An additional motivation followed from a series of meetings with proponents of the BFO community, driven by the fact that BFO was in the process of ISO standardization and the expectation of the standards community that proponents of standards should cooperate and ideally create interfaces between standards.

Interoperation between ontology standards is mostly driven by the need for interoperability of data that are annotated or coded by these standards. Lack of interoperability strongly affects the use of data

in the biomedical field. More and more ontologies used for research use BFO as their foundational level, whereas SNOMED CT gains more ground as a healthcare ontology. The interest in harmonization between SNOMED CT and BFO is therefore motivated by the interest in closing the data gap between healthcare and biomedical research.

Although there are a number of foundational ontologies, which might be equally suited as a counterpart to SNOMED CT (see Applied Ontology issue “Foundational Ontologies in Action” (Borgo et al., 2022), presenting seven foundational ontologies), BFO was chosen for the following reasons:

- Its high degree of consolidation, documented by the fact that it has recently become an ISO standard, committed to support the interchange of information among heterogeneous information systems (ISO/IEC, 2022);
- Its focus on the representation of entities relevant to natural science, particularly life science including healthcare;
- Its importance as an upper level of biomedical domain ontologies in the OBO Foundry (Smith et al., 2007);
- The fact that both BFO and SNOMED CT adopted OWL (OWL, 2023) as one representational language (besides others).

This does not mean that the review of SNOMED CT against other foundational ontologies would be less fruitful; we even postulate (albeit without providing evidence in this paper) that, generally, most foundational ontologies that like BFO subscribe to a three-dimensionalist view of the world would lead to solutions that are very similar to what we propose here.

## 1.2. SNOMED CT

SNOMED CT is a large clinical terminology standard, proposing interoperable codes linked to clinical terms in several languages. The mission of SNOMED International, the non-profit standards development organization that owns and maintains SNOMED CT is to support semantic interoperability between clinical care systems, as well as between clinical care systems and biomedical research environments, across institutions, jurisdictions and linguistic groups, by what they name a “global language for health”. The roots of this global language, SNOMED CT, lie in a nomenclature, i.e. a multiaxial and hierarchical, albeit informal, compilation of English medical terms, driven by the College of American Pathologists (CAP). The early versions SNOMED, SNOMED II and SNOMED 3 were followed by SNOMED RT (Spackman et al., 1997), underpinning the term collection with description logics axioms, aiming at the capacity of computing equivalence between the meaning of term compositions and pre-existing terms, thus addressing a desideratum formulated in 1994 (Campbell et al., 1994). After merging SNOMED RT with CTV3 (“Clinical Terms Version 3”), a hierarchical and systematized terminology used in the U.K. National Health Service, it became SNOMED CT (“Clinical Terms”) in 2002. In 2007, the intellectual property rights to all versions of SNOMED were acquired by IHTSDO, now SNOMED International. Thus, SNOMED CT is the result of a long bottom-up terminology engineering process based on what clinicians recorded and wished to retrieve about patients. Each code represents a standardized meaning, called a SNOMED CT concept. SNOMED CT concepts are ordered in multiple hierarchies that extend a domain-specific upper level with foundational classes such as *Clinical finding*, *Procedure*, *Organism* etc. In conclusion, SNOMED CT can be seen as an artifact that has increasingly incorporated notions of logic and ontology, but which is still tied to a strong legacy of more than 50 years (Cornet and de Keizer, 2008).

### 1.3. Basic Formal Ontology (BFO)

In contrast, Basic Formal Ontology (BFO) (Arp et al., 2015; Basic Formal Ontology, 2022) is the result of a top-down process, carried out by a cross-disciplinary academic team with a strong anchoring in analytic philosophy. It resulted in a foundational ontology that proposes a small set of disjoint, hierarchically ordered types (universals), accompanied by formal binary and ternary relations, textual definitions and elucidations, as well as formal axioms. Foundational ontologies like BFO typically introduce upper-level distinctions by disjoint classes, with the aim to provide clear-cut dissections of reality – in terms of types, properties (relations) and constraints – in benefit of domain ontologies that import them as their top layer. It is expected that domain ontologies using the same upper layer better interoperate.

BFO uses first-order logic; an approximate rendering in OWL-DL description logics (Baader et al., 2008) is mostly finished. BFO has recently become an ISO/IEC standard (ISO/IEC, 2022). As a three-dimensional ontology, BFO is well known for its top-level bipartition into *continuants* and *occurrents*, terms coined by the logician William Ernest Johnson in the 1920s (Simons, 2000). As the *continuant-occurrent* distinction is crucial for the remainder of this paper it will be elucidated in detail:

- Continuants are those things that exist in time and have no temporal parts. Material entities, spaces or qualities are typical continuants, e.g. an aspirin tablet, an operation theater, the cavity of a stomach, or a broken bone.
- Occurrents are entities in time like processes and events with temporal parts, i.e. phases or temporal slices, such as July in a year, the opening of the chest in a heart transplant procedure, adolescence in a human's life, or the event of a bone fracture or its healing process.

The difference is that you can never take away temporal parts from occurrents, e.g. July 2022 from the year 2022 or adolescence from one's life (then it would no longer be the same), but in contrast you could take away parts of continuants, e.g. a tooth from one's body without affecting its identity. Material continuants typically have a volume and/or a mass (which may vary every instant), as opposed to occurrents, which have a duration. Occurrents “happen”, whereas continuants “are there” and maintain identity across time. Typically, continuants participate in occurrents, e.g. a heart participates in a heart transplant, a person participates in an exam, but also a biological organism is a participant in this organism's life. In BFO, an important descendant class of *bfo:Occurrent* is *bfo:Process*, for *bfo:Continuant* the classes *bfo:Material entity*, *bfo:Immaterial entity*, *bfo:Quality*, *bfo:Disposition*, as well as *bfo:Generically Dependent Continuant* as a container of data, information and alike.

### 1.4. BFO – SNOMED CT synopsis

Table 1 displays key features of SNOMED CT and BFO 2. Given that one is a domain ontology and the other is a top-level ontology, it is not intended for direct comparison. What can be sensibly compared between SNOMED CT and BFO is (i) SNOMED CT's top hierarchy (the classes directly underneath “SNOMED CT concept”) with the BFO class hierarchy, (ii) the BFO relations with SNOMED CT's linkage concepts (binary relations), and (iii) SNOMED CT concept model (a set of domain and range constraints on SNOMED CT object properties (SNOMED International, 2023)) with BFO axioms.

In terms of scope, SNOMED CT would ideally fit underneath BFO, as there is no or minimal overlap (Fig. 1). There are, however, controversies regarding the possibility of aligning/harmonizing the two ontologies, especially due to BFO's strict desiderata concerning domain ontologies linked to it (see OBO Foundry criteria (Smith et al., 2007)). However, interoperation between the two artifacts does not

Table 1  
Synopsis SNOMED CT vs. BFO, cf. Smith et al. (2005), Lowe (2006), Arp et al. (2015)

	SNOMED CT (July 2023)	BFO Version 2020
Scope	Clinically relevant entities in a broad range of abstractions	Most general “categories of being”, as deemed relevant for representing natural science (Arp et al., 2015)
Focus	Clinical medicine, health care, social care, biomedical research	Physical reality, scientific research
Intended use	Making electronic health records (EHRs) standardized, computable and interoperable. Ontology in which the meaning of clinical terms in many languages is grounded.	Providing a foundational system of types (and their extension to classes) to support the creation and maintenance of interoperable and computable domain ontologies for science and engineering
Intended audience	Developer of clinical systems. End users (clinicians) will see the terms (and not the ontology behind) and should trust that the latter does the right job (inferencing, disambiguation)	Ontology and terminology developers
Size	Very huge (362,738 active classes and properties)	Very tiny (122 classes and properties)
Top level divisions	Flat, mostly disjoint top-level concepts under “SNOMED CT concept”. However, things like material entities and processes can be found in several upper level hierarchies	Uppermost node “entity” split into “continuant” and “occurrent”. Everything in the world is either a continuant or an occurrent
Nodes represent	“Clinical ideas”, i.e. intensional meanings, which extend to classes of potentially clinically relevant entities	Universals (which only exist in their instances), but which also extend to non-empty and very general classes
Relations	Binary relations (“linkage concepts”), corresponding to OWL object properties and datatype properties	Binary and ternary relations. The latter ones raise problems when creating an OWL version (require reification). Relations made their way into BFO only recently (Version 2); they are largely based on the OBO Relation Ontology (Smith et al., 2005)
Formal representation	Description logics OWL EL	First order logic (Description logics representation not straightforward due to time-indexed relations), in its several OWL approximations use of OWL DL
Naming	Numeric concept IDs, artificial, self-explaining labels, in English and Spanish, called Fully Specified Names, real-world terms (quasi-synonyms)	Artificial labels, no synonyms
Textual scope notes	Low coverage of textual definitions, underspecification of many primitive concepts due to lack of textual scope notes	Highly elaborated definitions and elucidations, refined in numerous iterations
References to external sources/standards	Standards, clinical literature for the curation of terminology content, e.g. Gray’s anatomy, TA, FMA, and others for the body structure. Other examples, such as classifications for many clinical conditions, e.g. fracture, ulcers, etc.	Scientific realism (Chakravartty, 2017), with references to Aristotle and Quine; continuant/occurrent distinction borrowed from Johnson (Simons, 2000); four-category ontology defended by Lowe (Lowe, 2006)
Hierarchies	Multiple (the taxonomic relation is-a, interpreted as OWL SubClassOf or SubPropertyOf). Top-level classes directly under SNOMED CT root node, are considered as disjoint classes (except ‘physical object’ and ‘pharmaceutical/biologic product’). The rest are multiple but still following the disjointness from the top-level classes.	Single hierarchies. All divisions and subdivisions are strictly disjoint but not necessarily exhaustive
Equivalence axioms	32 % (varying from ~100 % to 0 % dependent of the hierarchy)	No equivalence axioms

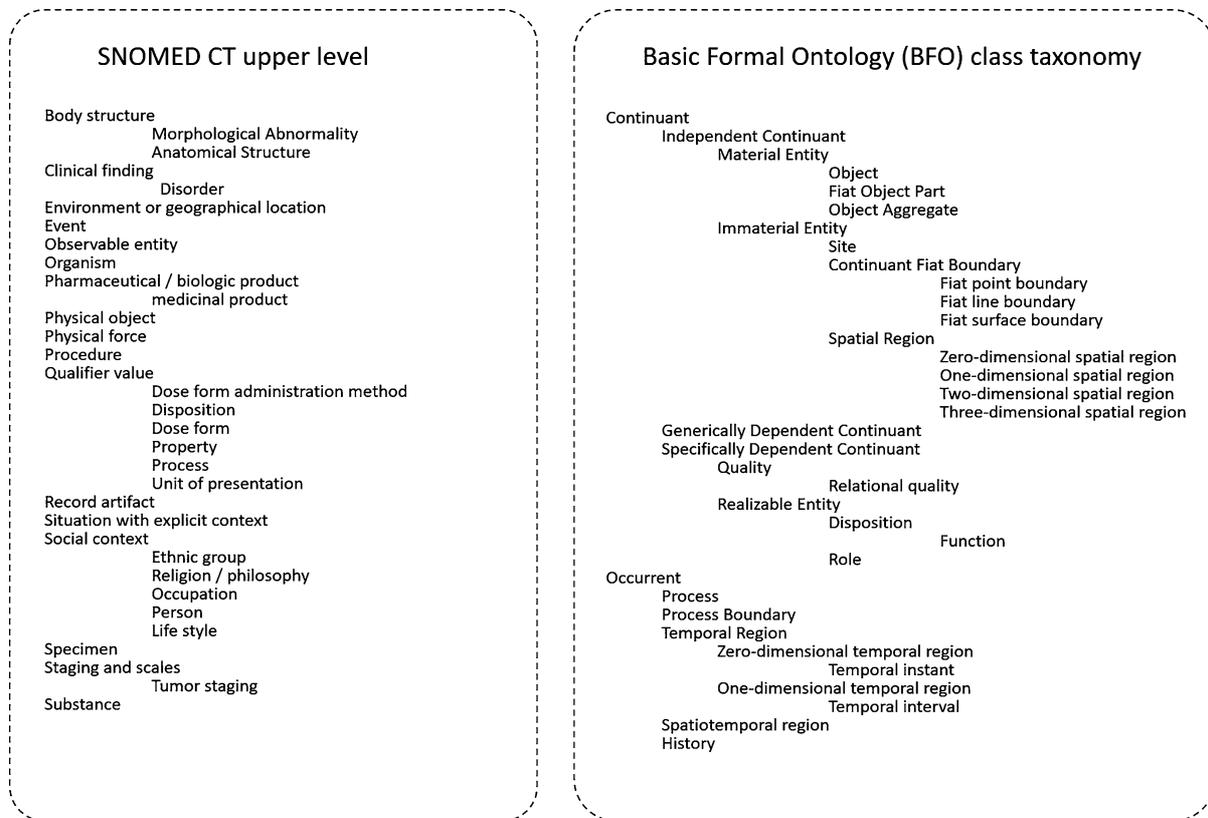


Fig. 1. Class hierarchies in SNOMED CT and BFO (complete). Indentations indicate subclasses.

stop at technicalities and alignment tasks, which are common in knowledge representation circles, where the fitness for specific use cases are the criterion for achievement. BFO claims that it represents reality independent of any purpose because tailoring ontologies to address specific purposes would undermine their ability to serve interoperability (Smith, 2018).

SNOMED CT has never raised that universal claim, and from its history, it has always been committed to clinical documentation tasks and therefore driven by the need of providing standardized meaning to human language expressions used in clinical care contexts and materialized in electronic health records (EHRs). Although nowhere explicitly stated in the SNOMED CT documentation, we make the assumption in the further course of our deliberations that SNOMED CT categorizes things in reality, whenever used for the purpose of clinical documentation, which range from physical entities and processes to qualities and information entities under several distinct upper level concepts.

### 1.5. Purpose of this study

A comprehensive agenda for BFO-SNOMED CT harmonization would require identifying appropriate BFO categories corresponding to each SNOMED CT hierarchy and, if needed, subdivisions thereof, by scrutinizing the current state of SNOMED CT in light of the precepts of formal-ontological analysis in general, and the foundational divisions proposed by the BFO ontology in particular.

Preliminary work showed that important parts of current SNOMED CT can easily be aligned with the upper-level classes of BioTopLite, an experimental domain upper level ontology, partially aligned with BFO (Schulz and Martínez-Costa, 2015), particularly organisms, devices, procedures and substances. For those other hierarchies where there are still open issues regarding BFO compatibility, *Clinical findings* (CF) stands out not only regarding its content, which roughly accounts for one third of SNOMED CT and is particularly rich in formal axioms, but also because it has been the focus of controversial discussions among members of the Applied Ontology, Medical Informatics and SNOMED CT communities for more than one decade.

The goal of this paper is therefore to investigate the harmonization of the CF hierarchy with BFO. We expect, as a result, a formal framework that is consistent with the current content and useful for the modeling of new CF content, formulated as a set of recommendations. This requires first of all to clarify the ontological commitment of SNOMED CT, by adding more precision to what is currently provided by the SNOMED CT Concept Model (SNOMED International, 2023).

Criteria of success would be (i) to reach a consensus among SNOMED CT users and maintainers on the resulting clarifications and recommendations, (ii) to facilitate the use of SNOMED CT together with other domain ontologies based on BFO, (iii) to facilitate its use with information models for clinical data interoperability, (iv) to better support biomedical data representation and management, and (v) finally to get closer to an answer to the question whether two ontologies with such different histories and criteria can be reconciled.

We highlight that (i) this study is preliminary in the sense that it does not propose any experimental approach to assess the interoperability claims, and (ii) limited in scope as it focuses on the CF hierarchy only. We also refrain (iii) from discussing possible ontological criteria for distinguishing between findings and disorders, as well as (iv) between findings that are necessarily pathological and those that are pathological only in certain contexts, because this does not affect the ontological nature of CFs.

## 2. Resources and methods

### 2.1. Methodological considerations

Most of this section is devoted to an in-depth description and clarification of the resource SNOMED CT and its basic tenets, in the light of BFO. The notion of “Concept” in SNOMED CT and its boundaries regarding a BFO-compatible interpretation is analyzed and illustrated by means. SNOMED CT Clinical Findings (CF) are then discussed in light of the BFO *Continuant/Occurrent* dichotomy. In particular, the syntactic/semantic phenomenon of so-called role groups, a fundamental design principle in the CF hierarchy (but not limited to it), is presented. Content with dispositional meaning within CF is another aspect to be elucidated. All this lays the ground for the qualitative methodology, which is the core of this work, and whose results are presented in the following section, to be summarized as follows:

- Modeling of the SNOMED CT class *Clinical finding*, together with the prototypical examples as OWL models, implementing several design patterns that represent competing views of SNOMED CFs. Protégé is used as an ontology editor;
- Reduction of the impact on current SNOMED CT editorial principles to a minimum;
- Testing the logical entailments of these patterns, using HermiT as a description logic reasoning engine;
- Visualization of the modeling and reasoning results;

- Discussion of the results, selection of a preferred pattern and analyzing its impact on the current state of the CF hierarchy;
- Testing the preferred model for plausibility against a selection of active CF classes of the July 2022 release, i.e. those with the hierarchy tags “(disorder)” or “(finding)”. Out of a total of 115,998 classes, 62,875 had multiple stated and inferred parents. A random sample ( $n = 100$ ) of these classes was generated and manually reviewed by a domain expert.
- Formulation of recommendations for the future content development of the CF hierarchy.

Throughout the paper, class names are shown in *italics*, object property names in **bold**. For SNOMED CT and BFO the namespace prefixes ‘sct’ and ‘bfo’ are used, respectively. Symbols without namespace prefixes belong to formalisms as proposed by the authors. For description logics expressions we use the OWL Manchester Syntax (Horridge and Patel-Schneider, 2012). Outside the OWL context we also use SNOMED CT’s standard notation, which combines the ID with the fully specified name with the semantic tag in parentheses. An example is “279039007|*Low back pain (finding)*”. Regarding capitalization, we follow the conventions of the source ontologies. See also (Hitzler et al., 2012) for an introduction to OWL.

## 2.2. “Concepts” in a BFO-compatible interpretation of SNOMED CT

The term “concept” has repeatedly been subject of heated discussions between ontologists and terminology builders. “Concept” had been introduced as a cornerstone of terminology theory, with its ISO definition as “unit of knowledge created by a unique combination of characteristics” (ISO). SNOMED CT defines “concept” as “clinical idea”, which comes close to the ISO definition. In contrast, the creators of BFO repeatedly rejected the notion of concept (Smith, 2004), as being incompatible with the precepts of Scientific Realism they defend, in which an ontology’s representational units correspond to universals, in the sense as discussed in Western philosophy since Plato and Aristotle.

The attempt to harmonize BFO and SNOMED CT could already end at this point, but from a pragmatic point of view there is a common denominator, namely the set-theoretical semantics of OWL. Description logics specifies only classes and properties and is therefore agnostic regarding the question whether the members of a class are included in the corresponding extension of a concept or a universal.

However, at the level of SNOMED CT itself, the word “concept” requires clarification, because – unorthodoxically, for SNOMED any entity that has a SNOMED ID is a concept, which deliberately includes relations and metadata elements.

This is why the authors refrain from the use of the word “concept” in the remainder of this paper and introduce the following definitions, in accordance with SNOMED International (2023).

1. SNOMED CT concepts that correspond to OWL classes will be referred to by “SNOMED CT classes”.
2. Those SNOMED CT concepts that are descendants of “Concept model attribute” (and are, in fact, binary relations) correspond to OWL object and datatype properties. We will name these “SNOMED CT properties”.

Many SNOMED CT classes have formal definitions in OWL axioms. E.g. ‘*sct:Low back pain*’ is a pain located in the lumbar region of back. The axioms in SNOMED CT class definitions correspond to an intensional meaning (Fitting, 2020), i.e. the conjunction of properties that defines class membership. Whereas fully defined classes in SNOMED CT have at least one OWL equivalent class axiom, primitive ones are only specified by one or more OWL subclass axioms.

It should not be overlooked that there is a small number of SNOMED CT classes that are used for individual things, e.g. *China*, *French language*, *Zen Buddhism*. According to the recommendation of SNOMED International's Modeling Advisory Group, individuals are modeled as classes in order to avoid logical complexity and due to no particular benefit for reasoning. However, this is irrelevant to the focus of this paper.

Nevertheless, individuals – although they are rarely ever named in the biomedical domain – are fundamental, because OWL axioms are always quantifications over all individuals that belong to a class. Whereas 279039007|*Low back pain (finding)*|, 57190000|*Myopia (disorder)*|, and |1290040004|*Entire eye proper (body structure)*| are classes, the first author's backache, his shortsightedness, or his left eye are individuals, i.e. members of these corresponding classes.

BFO's view of an ontology as a system of universals leads to another limitation. The assumption that universals, by definition, exist in their instances (as they represent what individual things have in common) obviously precludes uninstantiated universals, but also universals defined by negation, such as representational units that correspond to the terms “non-smoker” or “absence of rib”. For BFO such terms would not have any relevance to an ontology. This has sparked controversies in the past, mostly with the argument that biomedical ontologies have to account for representing the entirety of scientific discourse. Here, terms that do not yet or do possibly not denote anything in reality, or that denote phenomena that might exist in the future or whose existence is disputed cannot be avoided. Examples are whole-body transplants or Qi deficiency (Schulz et al., 2011). Again, the ontologically neutral ground of OWL shows a way out of this dilemma. Given an ontology representing universals in the BFO sense and implemented in OWL would not preclude it being enhanced by additional classes that are defined on the basis of existing ones by using the logical constructors provided by the language, e.g. introducing a class *Non-Smoker* defining it via the classes *Person* and *Smoking* by using the negation operator. Such an OWL enhancement would not be an ontology in the BFO sense but nevertheless acceptable. From a SNOMED point of view, the distinction of the ontology proper and OWL models logically derived thereof seems academic, but it smoothens the concept-universal controversy and it does not contradict the principles of BFO.

What might be less acceptable to BFO is the notion of mental constructs as defining principles for entities within a scientific reality. Medical practice and biomedical sciences have always been characterized by such constructs, rooted in natural language expressions and used in contexts in which they were related to some biological entity. Examples are disease entities like sepsis or rheumatoid arthritis, which are ill-defined or repeatedly re-defined. Their intensional aspects represent mental constructs such as scientific hypotheses or disease models, rather than things that can be pointed to such as a broken bone or a red eye. Whether such constructs are eventually confirmed by a material understanding of an underlying phenomenon or abandoned is a result of progress of science. Nevertheless, such terms with changing meanings, merely phenomenological descriptions and supposed phenomena of future obsolescence are and will be important elements of reasoning, communicating and decision-making in the clinical realm, as well as subject to scientific investigations. This includes that traces of such “concepts” persist in clinical records and that we have to acknowledge that, at the time they were used, they pointed to some instance in reality – most generically to be understood as some state, event or series thereof, characterizing presence during some part of a patient's life.

Whereas a term like “Qi” (in the Tradition Medicine extension of SNOMED CT) in an ontology under BFO would conflict with BFO's commitment to the exact sciences, this should be less of a problem with a class '*sct:Qi deficiency*' if the referent of this class is not an entity of dubious existence but a state within the life of a patient about whom '*sct:Qi deficiency*' in traditional Chinese medicine is predicated.

As a general and not negotiable principle, SNOMED CT cannot blind out clinical terms with unclear, ill-defined or debatable meanings, because its mission is to represent the entirety of clinical discourse. Harmonization with BFO therefore means either to identify areas of SNOMED for which no attempt at harmonization should be made or, to find a mutual agreement on the referents (the entities in reality) SNOMED CT classes denote.

### 2.3. SNOMED CT Clinical Findings (CF) in the light of BFO Continuant/Occurrent dichotomy

The CF hierarchy includes 119,833 classes (July 2023 release). Part of them have the hierarchy tag “finding”, the rest “disorder”. There is no clear-cut criterion that distinguishes findings from disorders, though some distinctions have been proposed, apart from the fact that due to taxonomic inheritance all disorders are taxonomic descendants of some finding. The distinction often depends on circumstances and individual judgment. SNOMED defines CFs as follows: “normal/abnormal observations, judgments, or assessments of patients”, whereas disorders are always and necessarily abnormal clinical conditions. This definition leaves several questions open, particularly regarding the ontological high-level classes to which CF content belongs. This is not only a matter for interoperability with BFO, but for any foundational ontology with non-overlapping upper-level classes.

In the case of diseases/disorders, or more generally, clinically relevant body conditions, clinicians use, often interchangeably, terms like “disease”, “disorder”, “clinical course”, “clinical evolution”, “clinical picture”, or in other languages “sjukdom”, “Krankheit”, “maladie”, “enfermedad”, “disturbio”. In Chinese, they are the same “疾病” (jī-bing). Those terms and their hyponyms do not clearly denote types of entities that can be unequivocally put into the “continuant” or “occurrent” basket.

A tentative characterization of CF classes is that all of them are (mostly dynamic) bodily, mental, and social features that are subject to health-related investigation and scrutiny. To define them, the SNOMED Concept model provides a large part of following object properties: **‘sct:Finding site’**, **‘sct:Associated morphology’**, **‘sct:Temporally related to’**, **sct:Before**, **sct:During**, **sct:After**, **‘sct:Due to’**, **‘sct:Causative agent’**, **sct:Severity**, **‘sct:Clinical course’**, **‘sct:Episodicity’**, **‘sct:Pathological process’**, **‘sct:Has realization’**, **sct:Interprets**, **‘sct:Has interpretation’**, **‘sct:Finding method’**, **‘sct:Finding informer’**, **‘sct:Associated with’**, **sct:Occurrence**.

Most of these properties intuitively suggest that CF classes are occurrents. For instance, a strep throat can be seen as a process with an internal dynamic. A counterexample would be Trisomy 21, a material entity consisting of morphologically different chromosomes, which are, finally, bearers of dispositions that determine the known trisomy 21 phenotype. A supernumerary toe is, in contrast to the chromosomal condition, a macroscopic continuant, as well as ulcers or hematomas, which undergo morphological change, so that the occurrents in which they participate are more in the foreground. Finally, there are conditions, e.g. a specific gait or speech pattern, in which the participating continuants (limbs, joints, bones, tongue, pharynx) do not exhibit, in a snapshot view, any particularity, but their time-dependent configurations and movements catch the eye of the observer. Here, the focus is on the occurrent entity only. It is noteworthy that for each and every continuant class *X* an occurrent class ‘*Life of X*’ can be trivially constructed, whereas the contrary is more complex. There is no simple continuant correlative to an epileptic seizure, stuttering, or a bouncy gait. Instead, numerous continuants and particular configurations thereof correlating to the occurrent may be described, e.g. the CNS, neurotransmitters, etc., although the complete set of continuants may not be known.

In the light of BFO, this flexibility regarding considering CF classes as continuants or occurrents not only clashes with the disjointness between the classes *bfo:continuant* and *bfo:occurrent*, but also with the

fact that nearly all domains and ranges of BFO object properties are constrained by either *bfo:continuant* and *bfo:occurent*.

This is also an issue in the following paragraphs where we will introduce two additional characteristic features of the CF hierarchy, viz. Role groups and Dispositions. The importance of both features will become apparent in our later modeling approaches, so that their introduction is justified at this place.

#### 2.4. Role groups in the Clinical Finding (CF) hierarchy

Representing compound procedure and findings such as 64550003 [*Removal of foreign body from stomach by incision (procedure)*] or 75857000 [*Fracture of radius AND ulna (disorder)*] was a problem for SNOMED for some time. There are numerous such complexes where this issue arises.

However, to meet clinicians' requirements, compound findings must appear classified and be retrieved under each of their component parts so that, for example, queries concerning the class *Fracture of radius* include *Fracture of radius and ulna* – i.e. a fracture of the two bones of the forearm resulting from a single trauma at the same place on the arm. To this end, role groups were introduced in SNOMED CT's predecessor SNOMED RT (Spackman et al., 2002). Without role groups, *Fracture dislocation of elbow joint*, it would be unclear and produces incorrect query results, e.g. a fracture of joint. In the SNOMED CT compositional grammar, role groups appear as a specific operator {curly braces}, and they were translated in the SNOMED CT OWL version to the '**sct:Role group**' object property (Cornet and Schulz, 2009). That role groups have never been given a clear semantics and no clear-cut algebraic properties has been repeatedly subject to criticism (Schulz et al., 2009) and interpretation proposals (Schulz et al., 2006) by several authors of this paper. It turned out that if role groups were used only for complexes, the above conclusions could not be drawn by logical reasoning. This explains, why in the transformations to OWL, role groups were introduced uniformly in all formal definitions of CF, even where there is only one "group" like in:

$$\begin{aligned}
 & \text{'sct:Fracture of radius' EquivalentTo} \\
 & \text{sct:Disease and 'sct:Role group' some} \\
 & \quad \left( \left( \text{'sct:Finding site' some 'sct:Bone structure of radius'} \right) \text{ and} \right. \\
 & \quad \left. \left( \text{'sct:Associated morphology' some sct:Fracture} \right) \right)
 \end{aligned} \tag{1}$$

This modeling pattern formally corresponds to a defined class, but does not answer the question about the difference to its potential non-grouped model variant. Non-grouped variants do not occur in pre-coordinated definitions, but may result from post-coordination, i.e. by the creation of a compositional expression using SNOMED CT content and constructors. In OWL, a post-coordinated expression could result in the following:

$$\begin{aligned}
 & \text{sct:Disease and} \\
 & \quad \left( \text{'sct:Finding site' some 'sct:Bone structure of radius'} \right) \text{ and} \\
 & \quad \left( \text{'sct:Associated morphology' some 'sct:Fracture'} \right)
 \end{aligned} \tag{2}$$

A closer inspection of the children of *Fracture of radius* reveals the concept *Fracture of radius AND ulna*, which suggest that the meaning of the former is not exactly the intuitive one but a more inclusive

one with the sense “body condition including a broken radius” or “patient having a broken radius” (Schulz et al., 2011).

### 2.5. Dispositions in the Clinical Finding hierarchy

Dispositions, in BFO *bfo:disposition*, a subclass of ‘*bfo:specifically dependent continuant*’ denote all those properties that silently inhere in material entities and only become manifest under certain circumstances, like the propensity of a glass to break or an animal to mate (Choi and Fara, 2021). Ontologically, dispositions are defined via their potential manifestations or realizations, expressed by the relation ‘*bfo:has-realization*’. The following tripartition, *viz.* (i) material entity, (ii) a disposition that inheres in it, and (iii) the disease process that realizes this disposition, has been described by the Ontology of General Medical Sciences (OGMS) (Scheuermann et al., 2009), which specializes BFO. When talking about allergic conditions, clinicians distinguish between (silent) allergic dispositions, the anatomic correlates of these dispositions (e.g. the nasal mucosa), and (active) allergic manifestations (processes such as hay fever). In a similar vein, clinicians distinguish between cancer as an underlying disposition (which inheres in some combination of physical components), and its manifestation as malignant growth. Whereas the manifestation always realizes the disposition, the contrary does not always hold: the detection of cancer dispositions that are not yet manifest is currently one of the main drivers of biomarker research (Goossens et al., 2015).

In SNOMED CT, both dispositions and their manifestations are in the CF hierarchy, e.g. *300910009* |Allergy to pollen (finding)| for the disposition and *21719001* |Allergic rhinitis caused by pollen (disorder)| for the manifestation.

## 3. Results

### 3.1. Basic BFO framework

We present several OWL models. Their basic elements required for the representation of central ontological aspects of CFs in BFO are depicted in Fig. 2. We use *Cancer* as a prototypical example, due to its importance in medicine on the one hand and to the variety of interpretations on the other hand, which occur whenever clinicians write “cancer (of . . .)” in clinical notes. Here we found the following distinctions:

- *Cancer disease process* (#1), A *bfo:process*, which is an entity that perdures through time: an *occurrent*; it has temporal parts like early stage or late stage cancer.
- *Cancer* (#2), A material object, which is a ‘*bfo:independent continuant*’, representing the mass of tissue participating in #1.
- *Cancer tissue quality* (#3) A *bfo:quality* that inheres in some #2. It describes the shape of the tissue that characterize #2.

But also:

- *Cancer genomic structure* (#4) Again a ‘*bfo:material object*’, like #2, but here it means the germline structure of a human organism that favors a cancer disease process #1
- *Cancer disposition* (#5) A *bfo:disposition*, which inheres in #4, regardless of whether it is realized, i.e. whether there is a manifestation of #1 or not.

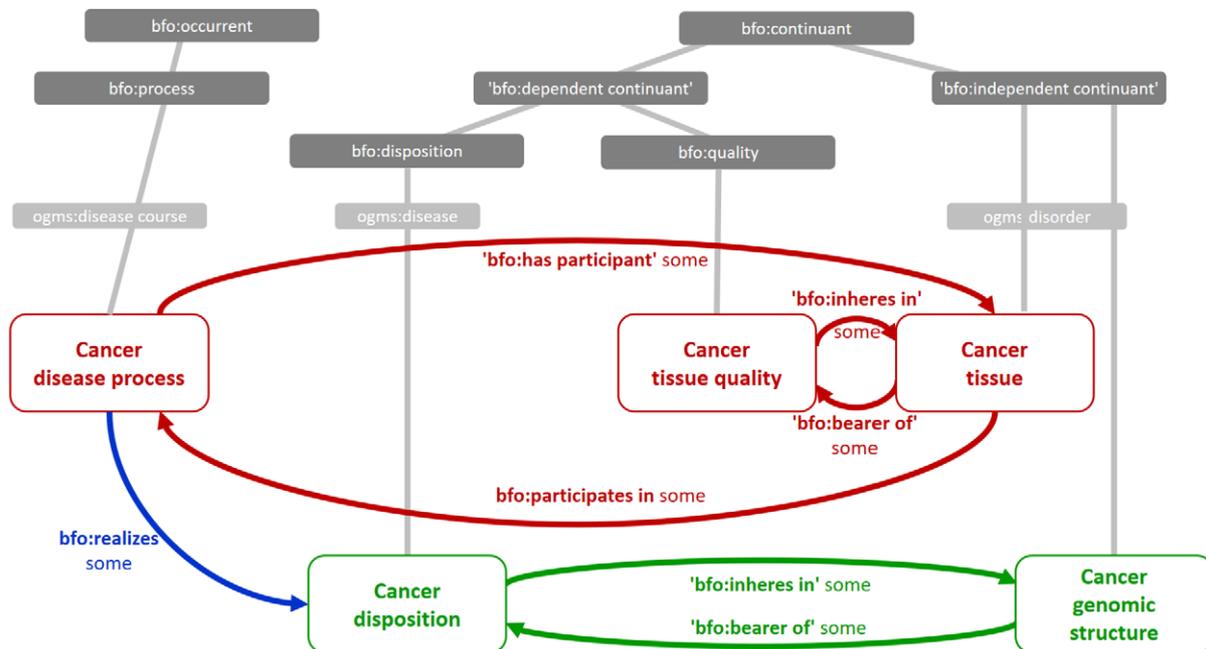


Fig. 2. The conceptual space of Clinical Findings (CFs) in the context of BFO. The example shows different entity types involved, typed by their BFO upper-level categories and OGMS (Ontology of General Medical Science) classes (Scheuermann et al., 2009). “Cancer genomic structure” relates to the germline genomic structure underlying a disposition for malignant growth.

The fact that clinical terminologies (in the sense of collection of human language terms) often do not distinguish between entities of the type #1, #2, and #3 on the one hand, and between entities of the type #4 and #5 on the other hand, does not mean that the referent of a term is a combination of them, e.g. that the referent of “cancer” is the intersection of a class under *bfo:process* and a *bfo:material object* at the same time, which would contradict the disjointness axiom between *bfo:occurent* and *bfo:continuant*. Instead, such an expression actually refers to several, distinct but mutually related entities. This is depicted in Fig. 1 by the “red block” (#1, #2, #3) as well as by the “green block” (#4 and #5):

- Whenever there is a cancer occurrent, there is also a (material) cancer. Whenever there is a (material) cancer, a cancer occurrent exists (or existed – in case we also classify dead tissue as cancer) (#1 ↔ #2).
- Whenever there is a material cancer, it exhibits a cancer quality and vice versa (#2 ↔ #3)

But also

- Whenever there is a genomic structure for cancer, there is a disposition for cancer and vice versa (#4 ↔ #5)

If clinicians speak or write “the patient has cancer”, whether they refer to the quality, the occurrent, or the material correlative thereof does not matter, because all of them exist whenever one of them exists. Only a more precise utterance like “the tumor has a size of 3 cm” disambiguates the term “cancer”, because only a material entity can have a size.

Note that there is no such mutual dependence between any of #1, #2, #3 and any #4, #5. Therefore, clinicians will always make a terminological difference between “risk of cancer” (#4, #5) and “cancer

manifestation” (#1, #2, #3). This is depicted by the relation **bfo:realizes**, which shows that a manifestation implies a disposition but not vice-versa.

Such mutually dependent entities are closely related in an ontological sense, but fall into different categories. This phenomenon had been termed “dot objects” and “logical polysemy” (Pustejovsky and Bouillon, 1995). Common examples of this are “University” (building vs. institution) or “Book” (printed copy vs. intellectual product), or in biology “Enzyme” (protein vs. function). Medical language abounds of this kind of polysemy, e.g., “inflammation” as denoting some morphologically altered tissue, which is the result of an inflammatory process, or “biopsy” for a procedure, but also as the outcome of this procedure, the biopsy sample (often called biopsy, too). Our built-in contextual awareness and extensive background knowledge explains why across all realms of discourse ontologically distinct but mutually dependent entities are often not distinguished by different words, even in technical language. Domain terms have different aspects of meaning, which in some context need to be distinguished, in some not. Each aspect, however, denotes a distinct entity in the domain, often belonging to disjoint upper-level type such as *bfo:continuant* or *bfo:occurrent*.

The SNOMED CT concept model allows representation of these aspects in the CF hierarchy; by defining CFs in their relation to processes by the SNOMED CT relation ‘**sct:Pathological process**’; as well as conditions that are related to continuants via the relation ‘**sct:Associated morphology**’.

Our modeling efforts resulted in the following three patterns. Pattern 1 defined the disjunctive class *Condition*, pattern 2 places CFs under a new class, *Clinical life phase*, and pattern 3 redefines CF as *Clinical occurrent*.

### 3.2. Pattern 1: CFs as descendants of condition, a disjunctive class

Intensive discussions around the ontological interpretation of CF content in SNOMED CT have taken place more than ten years ago by SNOMED CT terminologists and members of the SNOMED International special interest group “Event-Condition-Episode”. They argued for a *Condition* class in a clinical ontology as the disjunction between *Disposition*, *Material entity* and *Process* (Schulz et al., 2011). However, compatibility with BFO was not aimed at that time. In the meantime, the name “condition” has been established as an umbrella term for all clinically relevant phenomena (excluding medical procedures), particularly due to the resource *Condition* in FHIR (Ayaz et al., 2021). Additionally, the Human Phenotype Ontology (Robinson and Mundlos, 2010) has emphasized the view of clinical conditions as qualities. Formula (3) shows an updated disjunctive definition of *Clinical condition*:

$$\begin{aligned} &\text{‘Clinical condition’ SubClassOf} \\ &\quad \text{‘bfo:material entity’ or ‘bfo:quality’ or ‘bfo:disposition’ or ‘bfo:process’} \end{aligned} \quad (3)$$

Referring to or defining some condition *C* in a non-disjunctive sense  $x \in \{\text{‘material entity’}, \text{quality}, \text{disposition}, \text{process}\}$  can be expressed by:

$$C_x \text{ SubClassOf ‘clinical condition’ and bfo:x} \quad (4)$$

All axioms on these disjunctive SNOMED CT classes should work for all interpretations, e.g.

$$C_x \text{ SubClassOf (‘located in’ some A)} \quad (5)$$

understanding by ‘**located in**’ a universal spatiotemporal inclusion relation. In SNOMED CT this would correspond to

$$C_x \text{ SubClassOf } (‘\text{sct:Finding site}’ \text{ some } A) \quad (6)$$

This is plausible because the  $C_{process}$  as well as the  $C_{disposition}$ ,  $C_{quality}$  and  $C_{material-object}$  can be seen as located in the anatomical site  $A$ .

Formula (7) introduces our running example, modelled along Pattern 1:

$$\begin{aligned} & \text{sct:Osteosarcoma} \text{ EquivalentTo } ‘\text{Condition}’ \text{ and} \\ & (‘\text{sct:Associated morphology}’ \text{ some } ‘\text{sct:Sarcoma morphology}’) \text{ and} \\ & (‘\text{sct:Finding site}’ \text{ some } ‘\text{sct:Bone structure}’) \end{aligned} \quad (7)$$

Relevant reasoning patterns (e.g. that a condition of a part is a condition of a whole) could be shown to work also without committing to the precise ontological nature of that condition (every osteosarcoma is located in a bone, regardless of whether we mean by “osteosarcoma” the malignant growth process, its material correlate or its morphological quality).

The proposed approach leaves open how to interpret the SNOMED CT relation ‘**sct:Associated morphology**’. In an BFO alignment context this would require agreement on the BFO class to which SNOMED CT morphologies belong, *viz.* either as qualities (morphological shapes, which would be subsumed by ‘*bfo:specifically dependent continuant*’) or the material objects (subsumed by ‘*bfo:independent continuant*’) that exhibit these shapes. In the former case, which is currently discussed to become the preferred one in SNOMED CT circles, the distinction material bearer vs. quality would align with the current SNOMED CT concept model (SNOMED International, 2023). In the latter case, both the material morphology and the interpretation of *Clinical condition* as ‘*bfo:material entity*’ would become indistinguishable.

Whereas both cases would still preserve the ambiguity between *bfo:process* and ‘*bfo:material object*’ as above, it would also imply that processes have (material) qualities, which contradicts BFO and its domain restriction of ‘**bfo:bearer of**’ to ‘*bfo:independent continuant*’. As a result, if the pattern were “sct:condition ‘**bfo:bearer of**’ some x” then all conditions would be classified as under ‘*bfo:independent continuant*’ by a description logics reasoner, as demonstrated in Fig. 3 by the class *sct:Osteosarcoma*.<sup>1</sup> This contradicts the premise of the disjunctive reading. The explanation is exactly the restricted scope of *bfo:quality*, which rules out the reading as *bfo:process*, *bfo:disposition* or *bfo:quality*.

As stated above, we could ignore qualities and represent morphologies as subclasses of ‘*sct:Body part*’, which demands the relation ‘**bfo:has continuant part**’.<sup>2</sup> This, again, entails ‘*bfo:independent continuant*’, because in BFO processes, dispositions or qualities cannot have independent continuants as parts.

These problems did not appear when the disjunctive approach was originally proposed, because it was modeled under BioTop (Schulz et al., 2017) as ontological upper-level, which has a broader notion of quality, and provides a generic spatiotemporal inclusion relation. So it did not result in unintended models.

<sup>1</sup>The corresponding OWL model is available as supplementary material (BFO\_SCT\_pattern\_1a.owl).

<sup>2</sup>See OWL model BFO\_SCT\_pattern\_1b.owl

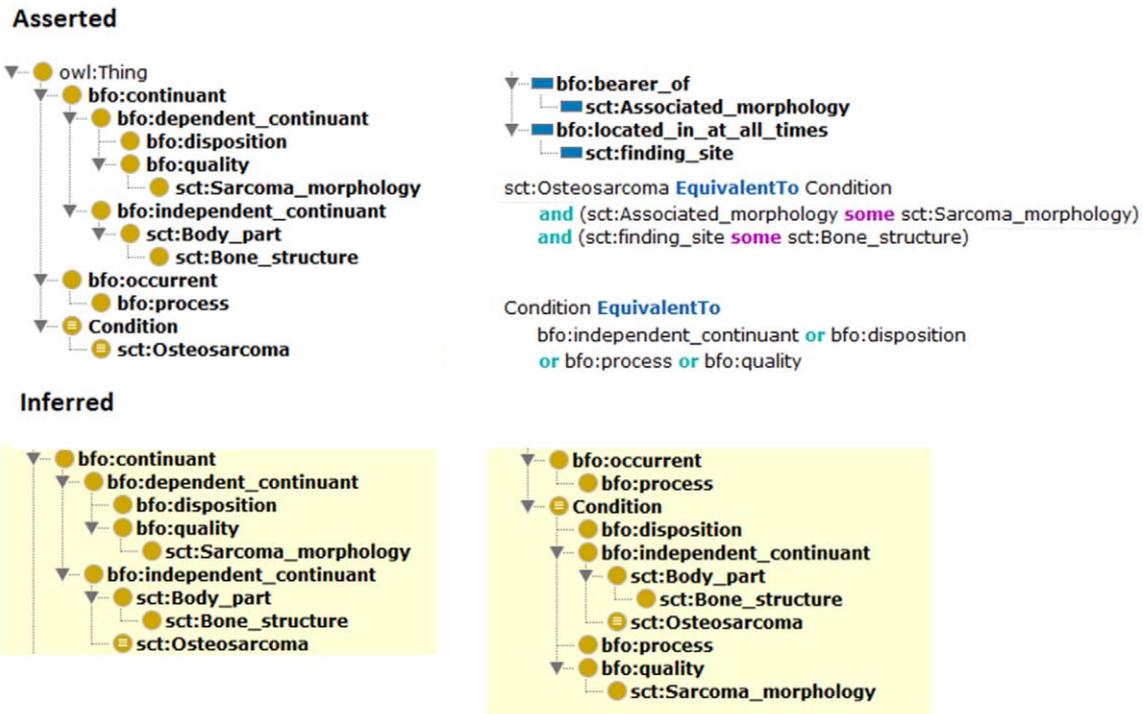


Fig. 3. Pattern 1: expressing SNOMED CT findings/disorders as clinical conditions yields an unintended model: *sct:Osteosarcoma* is classified under '*bfo:independent continuant*', which excludes its intended interpretation as *bfo:occurent*, which was the intention of the introduction of *Condition* as disjunctive class. The yellow-shaded hierarchy is the inferred one.

Back to BFO, the only compatible solution would be to introduce a new relation, e.g., '**located in**' as a common parent of the relations '**bfo:located in at all times**' and '**bfo:occurs in**'.

There are several reasons for not favoring this modeling approach. Neither a new top-level object property '**located in**' nor a disjunctive class *Condition*, would be compatible with BFO. CFs as subclasses of an inherently ambiguous *Condition* class are, additionally, prone to cause confusion when these classes need to be mapped to other ontologies.

Nevertheless, this model takes into account the undeniable fact that from a clinical-terminological point of view, many CF terms are polysemous under formal-ontological scrutiny, with their actual meanings (e.g. whether "tumor" means a lump of tissue of a growth process) only becoming clear from the context in which they are used.

### 3.3. Pattern 2: CFs as descendants of clinical life phase

The problem of how to represent complex diseases and syndromes had already occupied the creators of the GALEN ontology (Rector et al., 1997) in the 1990s, who termed it *Clinical situation*. Although *Situation* is used in a different sense in SNOMED CT, the representation of complex descriptions within the CF hierarchy has repeatedly been discussed in SNOMED circles, where the expression *Clinical life phases* was proposed, but a consensus on this matter was never reached (Cheetham et al., 2015). Clinical pragmatics is the question here, frequently debated in standards committees, of whether searches for cases of "Pulmonary stenosis" should include cases of "Tetralogy of Fallot", or searches for cases

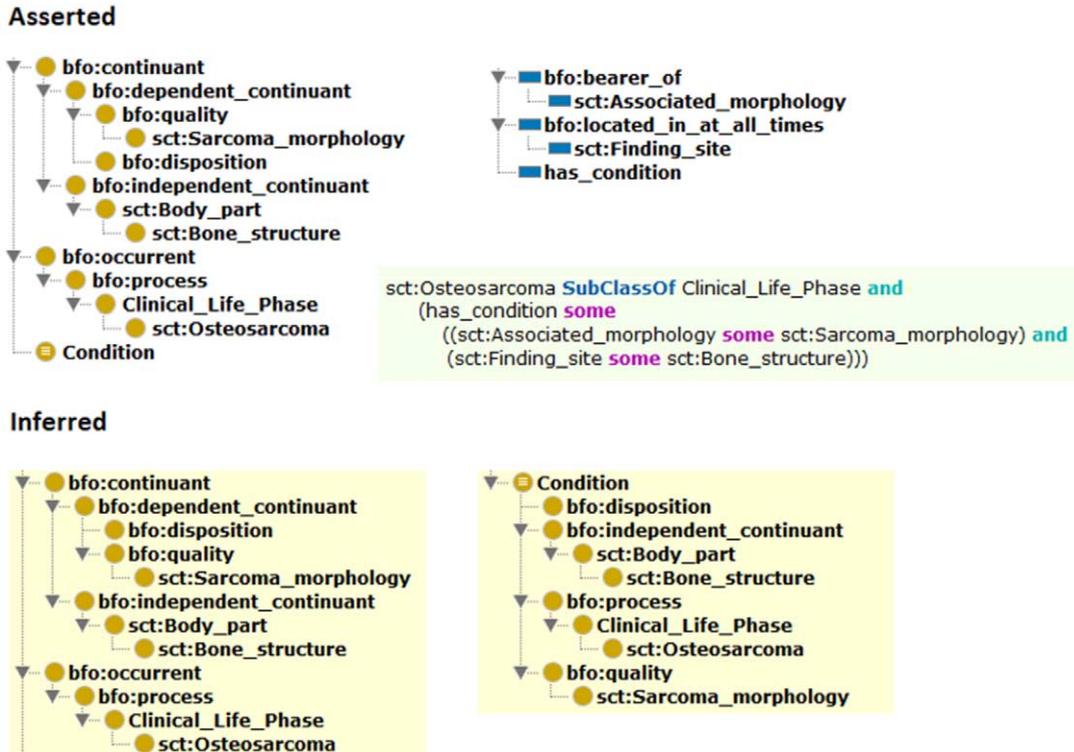


Fig. 4. Pattern 2: introducing clinical life phases as the root of the SNOMED CT findings/disorders. The object property ‘**has condition**’ represents the role group operator in findings/disorder. The yellow shade highlights the inferred OWL model.

of “Combined Fracture of the forearm” should be retrieved by the query “Fracture of the ulna”. The consensus has almost always come out that they should. However, if this reasoning is made possible by the architecture of the CF hierarchy itself, then the actual interpretation of  $X$  would be “case with  $X$ ” of “clinical life phase with  $X$ ”, rather than “condition  $X$ ”.

Figure 4 shows a modeling attempt under the class ‘*Clinical life phase*’.<sup>3</sup> Whereas *Condition* is still a disjunction as in Fig. 3, *Clinical life phase* is a process because lives are processes, so necessarily life phases (i.e. their temporal parts) are processes, too. Compared to the model depicted in Fig. 3 and discussed in Section 2.3.2, this additionally provides a plausible semantics to the role group object property (which is here renamed as ‘**has condition**’) and avoids the unintended model in Fig. 3.

Osteosarcoma would be modeled as following:

$$\begin{aligned}
 & \text{sct:Osteosarcoma EquivalentTo 'Clinical life phase' and} \\
 & \text{'has condition' some} \\
 & \left( \left( \text{'sct:Associated morphology' some 'sct:Sarcoma Morphology'} \right) \text{ and} \right. \\
 & \left. \left( \text{'sct:Finding site' some 'sct:Bone structure'} \right) \right)
 \end{aligned} \tag{8}$$

<sup>3</sup>See OWL model BFO\_SCT\_pattern\_2.owl

The problem here is that because *Condition* is still a disjunctive class, therefore, ‘**has condition**’ – with its range constrained by *Condition* cannot be aligned with any BFO object property. In addition, this solution would not allow the classification of *sct:Osteosarcoma* under a class-like post-coordinated expression such as used in a query (which would be of the type *Condition*):

$$\begin{aligned} & \text{('sct:Associated morphology' some 'sct:Sarcoma Morphology')} \text{ and} \\ & \text{('sct:Finding site' some 'sct:Bone structure')} \end{aligned} \quad (9)$$

It would require to complement the query with the ‘**has condition**’ object property:

$$\begin{aligned} & \text{'has condition' some} \\ & \text{((('sct:Associated morphology' some 'sct:Sarcoma Morphology')} \text{ and} \\ & \text{('sct:Finding site' some 'sct:Bone structure'))} \end{aligned} \quad (10)$$

Compared to Pattern 1, this pattern still has problems around the class *Condition*, which can only be represented as a disjunction of BFO classes and is therefore extremely weak. That the object property ‘**sct:Role group**’ is here re-interpreted as ‘**has condition**’ clarifies its meaning as linking clinical life phases with conditions. However, its harmonization with BFO fails due to the range restrictions of BFO object properties. In addition, the parallelism of *Clinical life phase* and *Clinical condition* could negatively affect the acceptance of this approach.

### 3.4. Pattern 3: CFs redefined as Clinical Occurrent under *bfo:occurrent*, with morphological abnormalities as descendants of ‘*bfo:independent continuant*’

The third pattern (cf. Figure 5) attempts to reconcile (i) parsimony, (ii) compatibility with BFO, (iii) expressibility in OWL-EL, and (iv) the redefinition of role groups as object properties that link occurrents to parts of occurrents. All of what is currently named ‘*Clinical finding*’ would then be ‘*Clinical occurrent*’, defined as BFO occurrents that describe reportable phenomena in and around patients, excluding health care procedures performed by healthcare professionals.

In contrast to the former two models, the classes *Condition* and *Clinical life phase* are abandoned.

In this modeling approach, **all SNOMED CFs – re-interpreted as “clinical occurrents” – are BFO occurrents**, i.e. temporally extended entities having one or many occurrents as temporal parts (including themselves), and having continuants as locations and participants. Again, the interpretation of role groups as OWL object properties is crucial. E.g. ‘*sct:Osteosarcoma* (grouped)’ means “any clinical occurrent that includes osteosarcoma”, whereas ‘*sct:Osteosarcoma* (ungrouped)’ means osteosarcoma in a strict sense (A query using the ungrouped concept would therefore not retrieve patient data annotated with the SNOMED CT concept ‘733064004 |*Osteosarcoma, limb anomalies, erythroid macrocytosis syndrome (disorder)*’).

$$\begin{aligned} & \text{'Osteosarcoma(grouped)' EquivalentTo 'Clinical occurrent'} \\ & \text{and ('bfo:has occurrent part' some} \\ & \text{((('sct:Associated morphology' some 'sct:Sarcoma morphology')} \\ & \text{and ('sct:Finding site' some 'sct:Bone structure'))} \end{aligned} \quad (11)$$

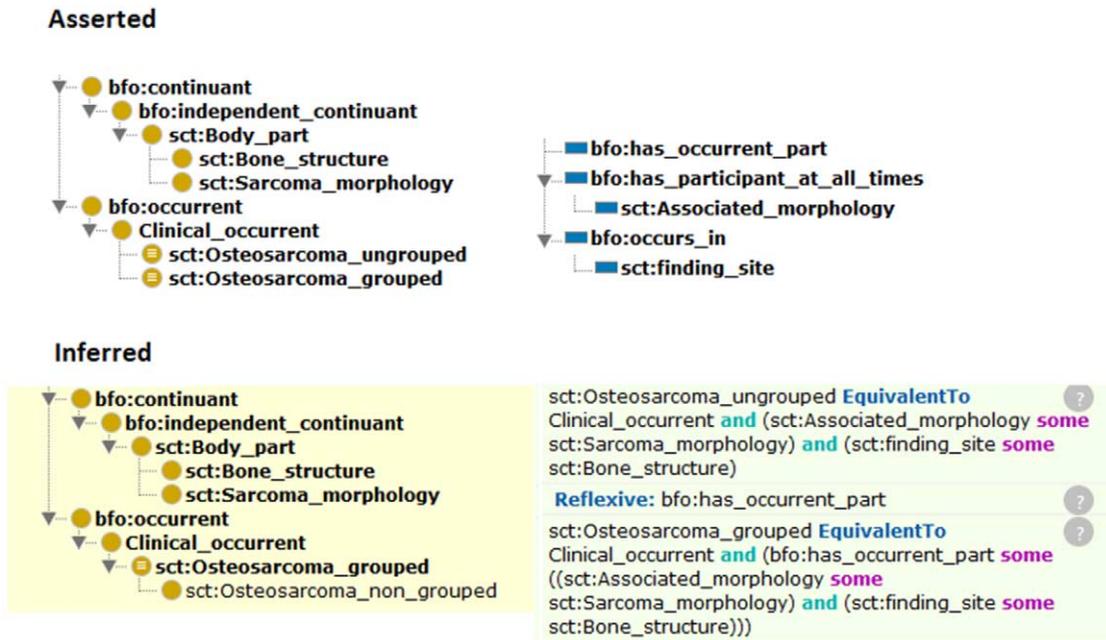


Fig. 5. Pattern 3: SNOMED CT findings/disorders as Clinical occurrences. Here, the reflexive and transitive object property ‘**bfo:has occurrent part**’ represents the role group property in CF. The yellow shaded graph shows the inferred OWL model.

Opposed to ‘*sct:Osteosarcoma* (ungrouped)’, with just means the processual aspect of osteosarcoma, and typically corresponds to an expression that results from SNOMED CT post-coordination:

$$\begin{aligned}
 & \text{‘Osteosarcoma(ungrouped)’ EquivalentTo ‘Clinical occurrent’} \\
 & \text{and (‘sct:Associated morphology’ some ‘sct:Sarcoma morphology’)} \\
 & \text{and (‘sct:Finding site’ some ‘sct:Bone structure’)} \tag{12}
 \end{aligned}$$

Since nearly all definitional axioms of CFs use role groups, the implicit meaning of a CF class *C* is always “any clinical occurrence that includes *C*” or “having *C*”.

The interpretation of CFs in the sense of “having *C*” (rather than “*C*” itself) also resolves the problem of identifying the exact instances of ill-defined and controversial concepts like rheumatoid arthritis or sepsis. “Having sepsis” as pointing to an umbrella process in a patient, which is assumed to include ill-defined subprocesses and process participants and qualities is just easier to conceive than an instance of sepsis, itself. Assuming that sepsis had once been defined as a pathophysiological process that has proven nonexistent afterwards, “having sepsis” as an ailment ascribed to a patient is less problematic to reconcile with BFO’s underlying philosophical theory, because the referent is always something ongoing in the patient, i.e. some *bfo:occurent*, which exists – even if the existence of the thing ascribed to it is controversial or speculative.

It is to highlight that this interpretation of all findings and disorders includes *states*, i.e. things that happen without any relevant dynamics or evolution from a given perspective (but which can have temporal parts and are therefore occurrences). Examples are findings of temperature, color, size, shape, or weight, but also CFs defined by the absence of a canonic body part, like 205306000 |*Congenital com-*

plete absence of upper limb (disorder)]. A correct interpretation of this SNOMED CT class as states or processes would be to refer to the totality of processes and states of an organism that lacks an upper limb. Where states are placed in an ontology and whether they are considered at all, varies between ontologies. So is BFO agnostic regarding any difference between processes, events and states (Galton, 2016). This makes our choice of ‘*Clinical occurrent*’, to be placed directly under *bfo:occurrent* the least controversial option. In addition, we propose the introduction of the subclass ‘*Clinical occurrent state*’ for those cases which are clearly considered static.

Another decision in this modeling strategy is to interpret ‘**sct:Role group**’ as the BFO relation ‘**bfo:has occurrent part**’, which is both transitive and reflexive. Reflexivity is a crucial point, because it allows that a class ‘*Clinical occurrent C*’ is instantiated both by *C* alone and by complex occurrents that co-occur with *C*. The only drawback here is caused by the limitation of reflexive relations in OWL, where they do not consider domain and range restrictions. Whereas ‘**bfo:has occurrent part**’ has its domain and range restricted to *bfo:continuant* in the FOL version of BFO, this is not the case in the OWL approximation. The solution is either to drop these restrictions in the specific implementation or to introduce a generic spatiotemporal inclusion relation, such as ‘**btl2:is included in**’ in the BTL2 ontology (Schulz et al., 2017). On the one hand, this would reduce compatibility with BFO, on the other hand, this could offer a solution for SNOMED CT role groups in all hierarchies, which still require further investigation.

The corresponding OWL model<sup>4</sup> shows that there is no difference regarding the current CF structure in SNOMED CT, apart from the reinterpreting of CF as ‘*Clinical occurrent*’ to which a subclass link to ‘*bfo:occurrent*’ is added, as well as the substitution of the object property ‘**sct:Role group**’ by ‘**bfo:has occurrent part**’.

Finally, bodily dispositions like allergy to pollen need to be fitted under ‘*Clinical occurrent*’, too. However, according to this modeling proposal they would be more precisely identified as a ‘*Clinical Occurrent*’ with allergic disposition state. This connection is indirect, because in BFO (as well as in the current literature to dispositions, cf. Choi and Fara (2021)), occurrents cannot be bearers of dispositions. Notwithstanding, such dispositional states could always be interpreted as clinical states that occur in material entities that are the bearers of the disposition. E.g., *Pollen allergy* is the state of an organism that has cells that are the bearers of allergic dispositions.

Currently, SNOMED CT links allergic dispositions to allergic processes via the relation ‘**has realization**’ as shown in Formula (13).

$$\begin{aligned}
 & \text{‘sct:Allergy to pollen (clinical finding)’ EquivalentTo} \\
 & \quad \text{‘sct:Propensity to adverse reaction’ and} \\
 & \quad \text{‘sct:Role group’ some} \\
 & \quad \quad \text{((‘sct:Has realization’ some ‘sct:Allergic process(qualifier value)’)) and} \\
 & \quad \quad \text{(‘sct:Causative agent’ some ‘sct:Pollen (substance)’))} \tag{13}
 \end{aligned}$$

Assuming that the latter is a BFO process (the modeling of processes as qualifier values in SNOMED CT may be debatable, but is out of scope in this paper), the left hand side (a BFO process, too) would

---

<sup>4</sup>See BFO\_SCT\_pattern\_3.owl

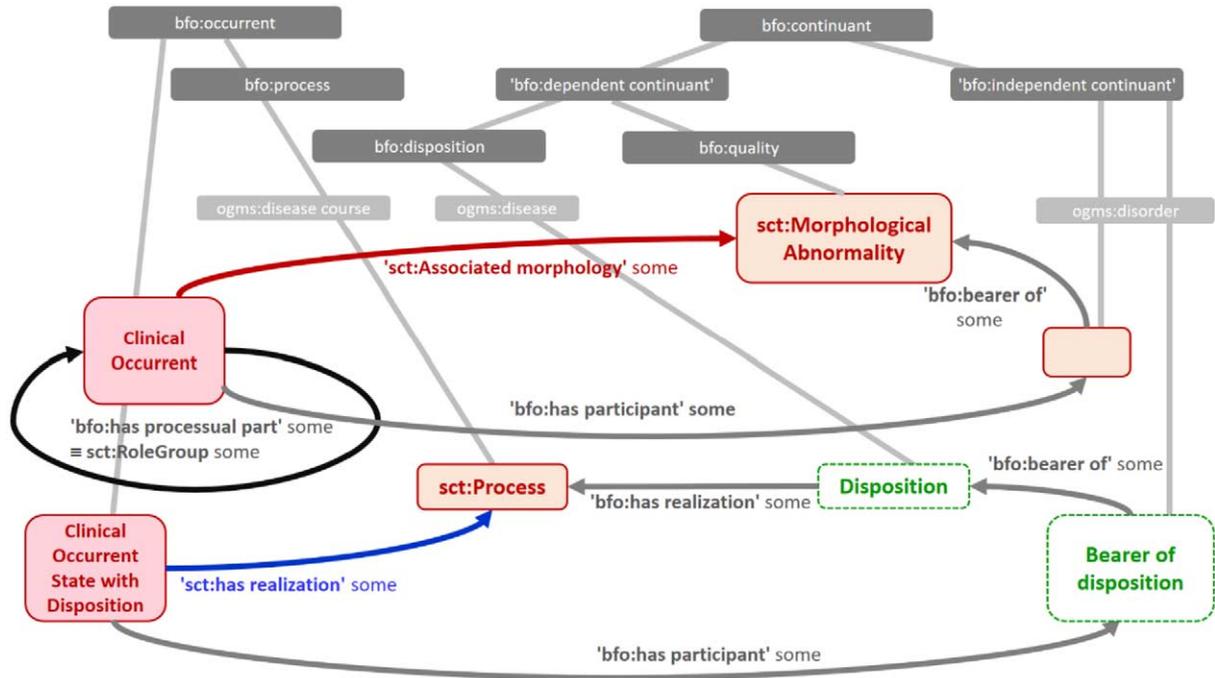


Fig. 6. SNOMED CT CF classes as clinical occurrents and clinical occurrent states. Compliance with BFO classes and properties. It highlights the need to refer to dispositional findings via clinical occurrents (i.e. “states with a disposition”), which requires to express the SNOMED CT object property ‘sct:has realization’ by a chain of BFO object properties (cf. Formula (15)).

not qualify as the domain of a BFO compatible ‘bfo:has realization’ relation. Correctly modeled (cf. Figure 6) it would be phrased as

$$\begin{aligned}
 & \text{‘sct:Allergy to pollen (clinical occurrent state)’ EquivalentTo} \\
 & \text{‘sct:Propensity to adverse reaction’ and} \\
 & \text{‘bfo:has occurrent part’ some} \\
 & \text{((‘bfo:occurs in’ some (‘bfo:bearer of’ some (‘bfo:disposition’ and} \\
 & \text{‘bfo:has realization’ some ‘sct:Allergic process (qualifier value)’)) and} \\
 & \text{(‘sct:Causative agent’ some ‘sct:Pollen(substance)’))} \\
 & \hspace{15em} (14)
 \end{aligned}$$

with ‘sct:Propensity to adverse reaction’ being a subclass of ‘clinical occurrent state’.

By using an OWL property path axiom, the disturbance caused by this difference can be minimized (with “o” being used as concatenation operator):

$$\begin{aligned}
 & \text{‘bfo:occurs in’ o ‘bfo:bearer of’ o ‘bfo:has realization’ SubPropertyOf} \\
 & \text{‘sct:Has realization’} \\
 & \hspace{15em} (15)
 \end{aligned}$$

This is the price to be paid here, otherwise the CF hierarchy would never become “clean” under a BFO perspective. Due to the lack of equivalent property statements in OWL, the equivalence between (13) and (14) can only be approximated: the classifier identifies *sct:Allergy to pollen (clinical occurrent state)* as a subclass of ‘*sct:Allergy to pollen (finding)*’, not as an equivalent class.<sup>5</sup> Again, this proposal is only a background re-interpretation of the current SNOMED CT content and axioms, because it does not require re-modelling.

Finally, the ‘*Clinical occurrent*’ approach could offer a clearer way to deal with negations, such as presence of a condition while another condition is absent. E.g., ‘146291000119108|Vomiting without nausea (disorder)’ is currently modelled as a primitive subclass of ‘422400008|Vomiting (disorder)’, while ‘162056003|No nausea (situation)’ does not qualify as an additional parent because it is in the ‘*Situation with explicit context*’ hierarchy. Content revision in this hierarchy under the aspect of conformance with the ‘*Clinical occurrent*’ pattern could be a step towards more clarity in dealing with absence of clinical conditions without extending the representation language.

### 3.5. Case study

The manual review of 100 CF classes randomly selected from the July 2022 release revealed that twenty of them, ([13–29%] in a 95% confidence interval) would lead to inconsistencies when literally interpreted.

Table 2 shows these 20 examples, classified by their characteristics. Only if analyzed under the assumption that each class has the meaning of *Clinical Occurrent* as suggested in 3.4, viz. “any clinical occurrent that includes *C*” or “having *C*”, these subclass links seem plausible.

## 4. Discussion

Our scrutiny of clinical finding and disorder classes in SNOMED CT, i.e. the descendants of ‘*sct:Clinical finding*’ (CF) has followed up earlier discussions regarding the ontological commitment of this pivotal type of entity (Schulz et al., 2011) and how it can be best interpreted in terms of BFO. It also builds on past deliberations on an appropriate semantic interpretation of the ‘**sct:Role group**’ object property (Spackman et al., 2002; Cornet and Schulz, 2009). We elaborate an interpretation of the current state of that hierarchy, which not only proposes a clear semantics of SNOMED CT findings and disorders as descendants of the proposed class ‘*sct:Clinical occurrent*’, but also to ‘**sct:Role group**’, identified as the reflexive relation ‘**bfo:has occurrent part**’, which relates occurrents with their (proper and improper) parts. Thus, our original claim that SNOMED CT could be based on BFO 2020, addressing the requirement of mutual exclusiveness between continuants and occurrents can be fulfilled in the scope under scrutiny, viz. descendants of ‘*sct:Clinical finding*’. By interpreting them as *Clinical occurrents*, as suggested here, these SNOMED classes can be smoothly integrated with BFO without major changes either to the structure of SNOMED CT or to BFO. The recommended changes, according to our preferred representational pattern (formulae (11)–(15)) would then be limited to:

- Interpreting ‘*sct:Clinical finding*’ as ‘*Clinical Occurrent*’, a direct subclass of *bfo:occurent*.
- With ‘*sct:Body structure*’ already subsuming all morphological abnormalities as material aspects of pathology (in the sense of OGMS “disorders”), these would remain untouched, in a hierarchy

---

<sup>5</sup>See BFO\_SCT\_pattern\_4.owl

Table 2  
SNOMED CT findings and disorders sample, only interpretable as clinical occurments

Fully Specified name	Direct parents	Explanation
Dissociative neurological symptom disorder co-occurrent with symptoms of gait disorder (disorder)	Dissociative disorder (disorder); Abnormal gait (finding); Movement disorder (disorder)	Accompanying disorder expressed as taxonomic parent
Disorder of eye co-occurrent and due to Marfan syndrome (disorder)	Disorder of eye proper (disorder); Marfan's syndrome (disorder); Hereditary disorder of the visual system (disorder)	
Congenital spastic foot (disorder)	Congenital disease (disorder); Disorder of soft tissue of lower limb (disorder); Disorder of foot (disorder); Disorder of the central nervous system (disorder); Spastic foot (finding); Disorder of skeletal muscle (disorder); Developmental disorder (disorder)	Etiology and anatomic manifestation expressed as taxonomic parent
Familial porencephaly (disorder)	Autosomal dominant hereditary disorder (disorder); Porencephalic cyst (disorder); Hereditary disorder of nervous system (disorder)	
Gas gangrene caused by clostridium perfringens (disorder)	Infection caused by Clostridium perfringens (disorder); Gas gangrene (disorder)	
Cystic fibrosis (disorder)	Inherited mucociliary clearance defect (disorder); Autosomal recessive hereditary disorder (disorder)	
Kleine-Levin syndrome (disorder)	Sleep disorder (disorder); Disorder of brain (disorder)	
Community resource finding (finding)	Environmental finding (finding); Health management finding (finding)	Parents represent different healthcare aspects
Aortic valve overriding ventricular septum (disorder)	Congenital abnormality of cardiac ventricle (disorder); Ventricular septal abnormality (disorder); Congenital anomaly of aortic valve (disorder); Abnormal position of cardiac valve (disorder)	Conjunction of abnormalities of spatially but not mereologically related body structures expressed as common child
Continuity between mitral valve and pulmonary valve (disorder)	Congenital anomaly of mitral valve (disorder); Congenital abnormality of cardiac connection (disorder); Congenital pulmonary valve abnormality (disorder)	
Intracranial and intraspinal abscesses (disorder)	Intracranial abscess (disorder); Spinal cord abscess (disorder)	
Talipes valgus of bilateral feet (disorder)	Disorder of right lower extremity (disorder); Disorder of left lower extremity (disorder); Deformity of left foot (finding); Deformity of right foot (finding); Talipes valgus (disorder)	
Benign neoplasm of bilateral ovaries (disorder)	Benign neoplasm of right ovary (disorder) Benign neoplasm of left ovary (disorder)	

Table 2  
(Continued)

Fully Specified name	Direct parents	Explanation
Sepsis-associated myocardial dysfunction (disorder)	Cardiac complication (disorder); Myocardial dysfunction (disorder); Sepsis (disorder)	
Autosomal recessive progressive external ophthalmoplegia (disorder)	Hereditary disorder of nervous system (disorder); Autosomal recessive hereditary disorder (disorder); Mitochondrial cytopathy (disorder); Hereditary disorder of the visual system (disorder); Hereditary disorder of musculoskeletal system (disorder); Progressive external ophthalmoplegia (disorder); Chronic metabolic disorder (disorder)	
Cyclin-dependent kinase-like 5 deficiency (disorder)	Developmental delay (disorder); Seizure disorder (disorder)	Facets of a syndromic modeled as taxonomic parents
Modified measles (disorder)	Intellectual disability (disorder) Vascular disease of the skin (disorder); Exanthem caused by measles virus (disorder); Viral cardiovascular infection (disorder) Erythema of skin (finding); Vascular lesion of skin (finding)	Taxonomic parents express necessary signs of a disease
Leukonychia totalis, trichilemmal cysts, ciliary dystrophy syndrome (disorder)	Leukonychia totalis (disorder); Multiple system malformation syndrome (disorder); Trichilemmal cyst (disorder)	Elements of a syndrome-like disorder modeled as taxonomic parents
Intellectual disability, hyperkinetic movement, truncal ataxia syndrome (disorder)	Truncal ataxia (finding); Autosomal recessive hereditary disorder (disorder); Developmental hereditary disorder (disorder); Hereditary ataxia (disorder); Global developmental delay (disorder); Intellectual disability (disorder); Movement disorder (disorder)	
Foreign body in skin wound (disorder)	Wound of skin (disorder); Foreign body left in wound (disorder)	Different pathological entities are modeled as taxonomic parents
	Foreign body in skin (disorder)	

that naturally falls under *bfo:continuant*. E.g., the tumor as a mass would get under *bfo:continuant*; having the tumor would get under *bfo:occurent*.

- The proposal would not be affected by a future re-interpretation of morphological abnormalities as qualities. In such a case the material correlate of a disorder such as the tumor mass would then be expressed as the bearer of tumor mass morphological quality, and ‘**sct:Associated morphology**’ as the concatenation of the three BFO relations **bfo:occurs in**, ‘**bfo:has participant**’, and ‘**bfo:bearer of**’.
- Interpreting ‘**sct:Role group**’ as equivalent to the reflexive and transitive ‘**bfo:has occurrent part**’ would support current queries, and would particularly clarify the semantics of role groups once and for all, at least for the CF hierarchy.
- Expressing the SNOMED CT object property ‘**sct:Has realization**’ by the concatenation of three BFO object properties: ‘**bfo:occurs in**’ ◦ ‘**bfo:bearer of**’ ◦ ‘**bfo:has realization**’.

Thus, all instances of ‘*Clinical occurrent*’ exist in reality (because all the organisms, i.e. the patients whose health records SNOMED CT claims to represent exist), even those to which a controversial or obsolete term is ascribed now or was ascribed in the past.

The relevance of the fact that this modification does not require any structural redesign of SNOMED CT CF hierarchy and related axioms is emphasized by our analysis of a sample of classes in this hierarchy. Here, roughly between 8,200 and 18,200 classes would cause problems if their parents were interpreted literally. This would affect much more SNOMED CT content due to the placement of these classes in a tightly woven multi-hierarchical network.

According to the clinical committees advising SNOMED CT an accompanying disorder is placed as a taxonomic parent of the disorder it accompanies, e.g., ‘*sct:Disorder of eye co-occurent and due to Marfan syndrome (disorder)*’ subsumed by both ‘*sct:Disorder of eye proper (disorder)*’ and ‘*sct:Marfan’s syndrome (disorder)*’: “Having Marfan syndrome with an eye disorder” implies “having Marfan syndrome” and “having an eye disorder”. Only the *Clinical occurrent* interpretation provides an acceptable model for this.

An equally typical and frequent pattern is the expression of both etiology and manifestation as taxonomic parents, e.g., in ‘*sct:Cystic fibrosis (disorder)*’, the parents being ‘*sct:Inherited mucociliary clearance defect (disorder)*’ and ‘*sct:Autosomal recessive hereditary disorder*’. In a similar vein, constituting elements of syndrome-like disorders are represented as taxonomic parents, such as ‘*sct:Leukonychia totalis, trichilemmal cysts, ciliary dystrophy syndrome (disorder)*’, which inherits from ‘*sct:Leukonychia totalis (disorder)*’, ‘*sct:Multiple system malformation syndrome (disorder)*’, and ‘*sct:Trichilemmal cyst (disorder)*’. Finally, the case of ‘*sct:Modified measles (disorder)*’ shows how necessary signs of a disease are expressed as taxonomic parents, viz. ‘*sct:Exanthem caused by measles virus (disorder)*’, ‘*sct:Erythema of skin (finding)*’, and ‘*sct:Vascular lesion of skin (finding)*’, among others. In any evaluation of a clinical retrieval system, if the system failed to make these inferences, it would be considered faulty.

One could argue that this is the result of a long-lasting negligence of ontological principles among SNOMED CT developers, disregarding proposals already aligned with BFO such as OGMS with its tripartition between (i) clinical dispositions, (ii) clinical material entities and (iii) clinical processes (despite the often questioned labeling with “disease”, “disorder”, and “disease course”). However, from a pragmatic, clinical documentation point of view, a dissection of, e.g. *Fracture* into ‘*Fracture structure*’ and ‘*Fracture process*’ would be considered as redundant without any benefit. It seems that ‘*sct:Fracture of bone (disorder)*’, interpreted as “having a fracture” provides enough detail, because that a patient having a fracture also was the location of a *Fracture process* with the outcome of a *Fracture structure*, seems rather trivial and would not justify any documentation overhead beyond the assertion of ‘*sct:Fracture*

of bone (disorder)’. The fact that necessary signs and symptoms of a disease often appear as taxonomic parents could also be criticized from an ontological point of view (particularly some of them are continuants such as an exanthema and others are occurrents like a pathological gait pattern), but again, if the meaning of the disease  $D$  is “having  $D$ ” and its necessary sign  $S$  is “having  $S$ ”, then both are clinical occurrents and  $S$  subsumes  $D$ . The ontological nature of both  $D$  and  $S$  in a strict interpretation is irrelevant.

What is, in contrast, not irrelevant in clinical documentation is the difference between dispositions and manifestations. Recently, SNOMED CT has strengthened this distinction, particularly with regard to allergy, where ‘*sct:Allergic Condition*’ subsumes both ‘*sct:Allergic Disposition*’ and ‘*sct:Allergic Reaction*’. We have shown that our *Clinical occurrent* interpretation is able to support this distinction, which would be better characterized as *Clinical occurrent state*, and the meaning of dispositions in CF would also benefit from interpreting as “disposition states”. However, ‘**sct:Has realization**’ does not map to ‘**bfo:has realization**’, because the source class is a descendant of ‘*Clinical Occurrent*’. It therefore has to be interpreted as a concatenation of BFO object properties as shown in Formula (15), with the bearers of the dispositions and the dispositions themselves not being explicit. This is less straightforward compared to the other proposals, but nevertheless compatible with BFO, and its implementation would not have any impact on SNOMED CT maintenance and use.

It would be interesting to compare the SNOMED CT CF pattern with patterns found in other ontologies that represent clinical occurrents. The Human Phenotype Ontology (HPO) represents phenotypes as qualities, e.g. fractured is the quality of a fractured object like a bone:

‘*obo:fractured radius*’ EquivalentTo:

‘**obo:has part**’ some

(‘*obo:fractured*’

and (‘**obo:characteristic of**’ some ‘*obo:radius bone*’)

and (‘**obo:has modifier**’ some *obo:abnormal*))

(16)

Although a combined forearm fracture is not contained in the ontology, it is assumed that this would be expressed as a conjunction of two clauses beginning with ‘**obo:has part**’. Thus, the pattern is very similar to our re-interpretation of SNOMED CFs as descendants of ‘*Clinical occurrent*’, and the distinction between the “quality proper” (the last three lines of the axiom) and “having that quality” (the whole axiom) follows a very similar pattern as that the grouping done in our approach. Note also that the interpretation of phenotypes as qualities – in the BFO sense – is problematic, because BFO restricts qualities as properties that depend on continuants, whereas in HPO we also find qualities that inhere in occurrents, e.g. ‘*obo:behavior abnormality*’, which ‘**obo:inheres in**’ some ‘*obo:behavior process*’. As long as the HPO qualities inhere in continuants, a linking to SNOMED CT clinical occurrent could be done by the following equivalence statement:

(‘**sct:Finding site**’ some ‘*sct:Bone structure of radius*’) and

(‘**sct:Associated morphology**’ some ‘*sct:Fracture*’)

EquivalentTo

‘**sct:Finding site**’ some (‘*sct:Bone structure of radius*’ and  
 (‘**bfo:bearer of**’ some ‘*obo:fractured radius*’)) (17)

For HPO qualities that inhere in occurrents, HPO – BFO compatibility has to be established before discussing the harmonization of HPO with SNOMED CT.

Another interesting source is the openGALEN ontology (Rector et al., 2003), which grew out of the pioneering GALEN project (Rogers et al., 2001) in the 1990s.

*PulmonaryAtresiaWithVentricularSeptalDefect* EquivalentTo

*ClinicalSituation*

and (**isMainlyCharacterisedBy** some  
 (*presence* and  
 (**isExistenceOf** some *PulmonaryAtresia*)))  
 and (**isMainlyCharacterisedBy** some  
 (*presence* and  
 (**isExistenceOf** some *VentricularSeptalDefect*))) (18)

A problem that may have prevented a widespread use of this large ontology may be its idiosyncratic naming conventions and object properties. It contains very few combined disorders, but the definition of this one also reveals a modular structure that is very close to SNOMED CT clinical findings (the openGALEN relation **isMainlyCharacterisedBy** would then correspond to ‘**obo:has occurrent part**’).

Regarding BFO compatibility of other SNOMED CT hierarchies, in many cases the effort seems to be rather straightforward. As investigated by Schulz and Martínez-Costa (2015), however with regard to the BTL2 ontology, the ones that require most scrutiny are probably *Qualifier value*, and *Situation with explicit context*, *Observable entity*, *Social context*, and *Staging and scales*.

## 5. Conclusion and outlook

The early versions of SNOMED had started out working to provide structure based on representations of organ systems and pathophysiology that were intended to be understandable, reproducible, and useful (Spackman and Reynoso, 2004). With SNOMED’s evolution, the intent was to incrementally add structure in order to provide incremental value. This bottom-up development is in contrast to the top-down approach of BFO and other foundational ontologies, where a more comprehensive model is defined in advance.

In this paper, we reviewed the *Clinical finding* (CF) hierarchy, guided by the question whether this branch of SNOMED CT can be harmonized with the Basic Formal Ontology (BFO). The task was challenged by the intuition that diseases, disorders, signs and symptoms form a homogeneous upper-level class, because this is utterly incompatible with BFO’s upper-level distinction into continuants and occurrents.

We concluded that *Clinical finding* (CF) is rather an umbrella for many kinds of entities of clinical interest, which belong to different upper-level classes in BFO. *Clinical finding* would therefore be unsuited as an upper-level class that is compliant with BFO and probably most other foundational ontologies.

Analyzing this phenomenon more in detail, led us to the conclusion that *Clinical finding* classes in SNOMED CT (as clinicians understand them and current terminologies and ontologies have – mostly implicitly – interpreted them) actually do not refer to the entities proper (e.g. fractures, allergies, tumors, seizures etc., literally), but rather to the conditions of *patients having* a fracture, allergy, seizure etc. Only this gives sense to the striking and frequent characteristic of the *Clinical finding* hierarchy, viz. that complex classes are, taxonomically, children of their constituents, i.e. that they are linked to them by subclass expressions. In SNOMED CT, most of these taxonomic links are inferred, as the consequence of the ‘role group’ design pattern, which is ubiquitous in SNOMED CT and has often been subject of controversy regarding its semantics. Our analyses resulted in the proposal of (i) equating SNOMED CT’s ‘**sct:Role group**’ property with the BFO relation ‘**bfo:has occurrent part**’; and (ii) reinterpreting ‘*sct:Clinical finding*’ as ‘*Clinical occurrent*’, i.e. temporally extended entities in an organism, having one or many occurrents as temporal parts that occur in continuants. However, this proposal does not address usability aspects, particularly Gruber’s (Gruber, 1995) criterion of ontology clarity, viz. that an ontology “should effectively communicate the intended meaning of defined terms”. Whether the labeling of ontology classes suggest the right interpretation can only be ascertained in real-world experiments. Obviously the “clinical finding” hierarchy tag has not removed the controversy whether ‘*sct:Tetralogy of Fallot*’ is correctly placed under ‘*sct: Pulmonic valve stenosis*’. Reinterpreting it to something like ‘*sct:Clinical occurrent with Pulmonic valve stenosis*’, might bring more clarity, but it would require that “*Clinical occurrent*” is introduced and understood by the users.

A result that reaches beyond the SNOMED CT use case, is that this work provided evidence that the harmonization of a terminology system that grew over decades in a bottom-up manner with a principled foundational ontology, developed top-down, does not necessarily require major redesign, but rather a thorough ontological analysis of the implicit assumptions of its curators and users. The fact that the phenomenon of logical polysemy (Pustejovsky and Bouillon, 1995), which pervades domain terminologies, poses problems to ontologists but usually not to terminology users such as clinicians is an important factor in the harmonizing process. The work also suggests that a common ground is possible between those who insist in the application of ontological rigor and formal methods and those who do the practical work of producing artifacts that fulfill concrete requirements of clinical users and use cases. The experience that with some effort a top-down and a bottom-up approach can be harmonized, resulting in a shared understanding and representation, is a powerful validation of both approaches.

These are, however, still hypotheses, which require more validation effort. In this sense, we recommend the following investigations:

- Demonstrate the practical use of the harmonization with one or more of the BFO-compliant ontologies like HPO and OGMS, as well as with several disease ontologies that use BFO;
- Perform a similar scrutiny of other SNOMED CT hierarchies known as heterogeneous, particularly *Situation with Explicit Context*, *Qualifier Value*, and *Observable Entity*;
- Scrutinize all SNOMED CT object properties, together with the constraints from the SNOMED CT concept model against the BFO object properties and constraining axioms.
- Provide evidence for the benefit of BFO-SNOMED harmonization and integration for the communities of Applied Ontology and Biomedical Informatics as well as for healthcare and biomedical research in the context of health data analytics and interoperability.

Future SNOMED CT content development decisions should be informed by the output of these investigations, regarding naming of SNOMED CT components, particularly hierarchy labels, but also regarding updates in the documentation for users and content developers.

### Supplementary data

BFO\_SCT\_Pattern1a.owl, BFO\_SCT\_Pattern1b.owl, BFO\_SCT\_Pattern3.owl, BFO\_SCT\_Pattern3.owl, BFO\_SCT\_Pattern4.owl is available at: <http://dx.doi.org/10.3233/AO-230018>.

### Acknowledgements

We would like to give our appreciation to the reviewers Jim Campbell, Keith Campbell and Alan Rector. Their exceptional dedication and invaluable insights have greatly enriched the quality of this work. Their meticulous attention to detail, constructive criticism, and unwavering support have played an instrumental role in shaping this article into its final form. We extend our thanks to Barry Smith and Werner Ceusters who provided valuable feedback to an earlier version of the manuscript. We are immensely grateful for the expertise of all of them and the amount of time they invested. Their commitment to advancing knowledge in our field is truly commendable, and we sincerely thank them for their remarkable contributions.

### References

- Arp, R., Smith, B. & Spear, A.D. (2015). *Building Ontologies with Basic Formal Ontology*. MIT Press.
- Ayaz, M., Pasha, M.F., Alzaharani, M.Y., Budiarto, R. & Stiawan, D. (2021). The fast health interoperability resources (FHIR) standard: Systematic literature review of implementations, applications, challenges and opportunities. *JMIR Medical Informatics*, 9(7), e21929. doi:10.2196/21929.
- Baader, F., Horrocks, I. & Sattler, U. (2008). Description logics. *Foundations of Artificial Intelligence*, 3, 135–179. doi:10.1016/S1574-6526(07)03003-9.
- Basic Formal Ontology v.2020. <https://github.com/BFO-ontology/BFO-2020>. Accessed: August 20, 2023.
- Borgo, S., Galton, A. & Kutz, O. (2022). Foundational ontologies in action, 1, 1–16.
- Campbell, K.E., Das, A.K. & Musen, M.A. (1994). A logical foundation for representation of clinical data. *Journal of the American Medical Informatics Association: JAMIA*, 1(3), 218–232. doi:10.1136/jamia.1994.95236154.
- Chakravartty, A. (2017). Scientific realism. In E.N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy (Summer 2017 Edition)*. <https://plato.stanford.edu/archives/sum2017/entries/scientific-realism>. Accessed: August 30, 2022.
- Cheetham, E., Gao, Y., Goldberg, B., Hausam, R. & Schulz, S. (2015). Formal representation of disorder associations in SNOMED CT. In *Proc. of 3rd International Conference on Biomedical Ontology (ICBO)*. <http://ceur-ws.org/Vol-1515/regular6.pdf>. Accessed: August 20, 2023.
- Choi, S. & Fara, M. (2021). Dispositions, *the Stanford Encyclopedia of Philosophy*. E.N. Zalta (Ed.), <https://plato.stanford.edu/entries/dispositions>. Accessed: August 20, 2023.
- Cornet, R. & de Keizer, N. (2008). Forty years of SNOMED: A literature review. *BMC medical informatics and decision making*, 8(Suppl 1), S2. doi:10.1186/1472-6947-8-S1-S2.
- Cornet, R. & Schulz, S. (2009). Relationship groups in SNOMED CT. *Studies in Health Technology and Informatics*, 150, 223–227.
- Fitting, M. (2020). *Intensional Logic*, *the Stanford Encyclopedia of Philosophy*. E.N. Zalta (Ed.), <https://plato.stanford.edu/entries/logic-intensional/>. Accessed: August 20, 2023.
- Galton, A. (2016). The ontology of time and process. In *Third Interdisciplinary School on Applied Ontology*, Bolzano, Italy 2016. <https://isao2016.inf.unibz.it/wp-content/uploads/2016/06/bolzano-notes.pdf>. Accessed: August 20, 2023.
- Goossens, N., Nakagawa, S., Sun, X. & Hoshida, Y. (2015). Cancer biomarker discovery and validation. *Translational Cancer Research*, 4(3), 256–269.

- Gruber, T.R. (1995). Toward principles for the design of ontologies used for knowledge sharing? *International Journal of Human-Computer Studies*, 43(5), 907–928. doi:10.1006/ijhc.1995.1081.
- Guarino, N. & Musen, M. (2015). Applied ontology: The next decade begins. *Applied ontology*, 10(1). doi:10.3233/AO-150143.
- Hitzler, P., Krötzsch, M., Parsia, B., Patel-Schneider, P.F. & Rudolph, S. (2012). *OWL 2 Web Ontology Language Primer*. W3C Recommendation 11 December 2012 (2nd ed.). <http://www.w3.org/TR/owl-primer>. Accessed: August 20, 2023.
- HorrIDGE, M. & Patel-Schneider, P.F. (2012). *OWL 2 Web Ontology Language Manchester Syntax* (2nd ed.). W3C Working Group Note. <http://www.w3.org/TR/owl2-manchester-syntax/>. Accessed: August 20, 2023.
- International Standards Organisation (2023). *Standards*. <https://www.iso.org/standards.html>. Accessed: August 20, 2023.
- ISO 1087-1:2000, 3.2.1
- ISO/IEC 21838-2 Information technology – Top-level ontologies (TLO) – Part 2: Basic Formal Ontology (BFO). <https://www.iso.org/standard/74572.html>. Accessed: August 20, 2023.
- Lowe, J. (2006). *The Four-Category Ontology*. Clarendon Press.
- Otte, J.N., Beverley, J. & Ruttenberg, A. (2022). *BFO: Basic Formal Ontology* (pp. 17–43).
- OWL 2 Web Ontology Language Document Overview (Second Edition) <https://www.w3.org/TR/owl2-overview/>. Accessed: August 20, 2023.
- Pustejovsky, J. & Bouillon, P. (1995). Aspectual coercion and logical polysemy. *Journal of Semantics*, 12(2), 133–162. doi:10.1093/jos/12.2.133.
- Rector, A.L., Bechhofer, S., Goble, C.A., Horrocks, I., Nowlan, W.A. & Solomon, W.D. (1997). The GRAIL concept modelling language for medical terminology. *Artificial Intelligence in Medicine*, 9(2), 139–171. doi:10.1016/s0933-3657(96)00369-7.
- Rector, A.L., Rogers, J.E., Zanstra, P.E., Van Der Haring, E. & OpenGALEN (2003). OpenGALEN: Open source medical terminology and tools. *Annual Symposium proceedings. AMIA Symposium, 2003*, 982.
- Robinson, P.N. & Mundlos, S. (2010). The human phenotype ontology. *Clinical Genetics*, 77(6), 525–534. doi:10.1111/j.1399-0004.2010.01436.x.
- Rogers, J., Roberts, A., Solomon, D., van der Haring, E., Wroe, C., Zanstra, P. & Rector, A. (2001). GALEN ten years on: Tasks and supporting tools. *Studies in Health Technology and Informatics*, 84(1), 256–260.
- Scheuermann, R.H., Ceusters, W. & Smith, B. (2009). Toward an ontological treatment of disease and diagnosis. *Summit on translational bioinformatics, 2009*, 116–120.
- Schulz, S., Boeker, M. & Martínez-Costa, C. (2017). The BioTop family of upper level ontological resources for biomedicine. *Studies in Health Technology and Informatics*, 235, 441–445.
- Schulz, S., Brochhausen, M. & Hoehndorf, R. (2011). Higgs bosons, Mars missions, and unicorn delusions: How to deal with terms of dubious reference in scientific ontologies. In *Proceedings of the 2nd International Conference on Biomedical Ontology*. <http://ceur-ws.org/Vol-833/>. Accessed: August 20, 2023.
- Schulz, S., Cornet, R. & Spackman, K. (2011). SNOMED CT's ontological commitment. *Applied Ontology*, 6(1), 1–11. doi:10.3233/AO-2011-0084.
- Schulz, S., Hanser, S., Hahn, U. & Rogers, J. (2006). The semantics of procedures and diseases in SNOMED CT. *Methods of information in medicine*, 45(4), 354–358. doi:10.1055/s-0038-1634088.
- Schulz, S. & Martínez-Costa, C. (2015). Harmonizing SNOMED CT with BioTopLite: An exercise in principled ontology alignment. *Studies in Health Technology and Informatics*, 216, 832–836.
- Schulz, S., Spackman, K., James, A., Cocos, C. & Boeker, M. (2011). Scalable representations of diseases in biomedical ontologies. *Journal of Biomedical Semantics*, 2, S6. doi:10.1186/2041-1480-2-S2-S6.
- Schulz, S., Stegwee, R. & Chronaki, C. (2018). Standards in healthcare data. In P. Kubben et al.(Eds.), *Fundamentals of Clinical Data Science* (pp. 19–36). Springer.
- Schulz, S., Suntisrivaraporn, B., Baader, F. & Boeker, M. (2009). SNOMED reaching its adolescence: Ontologists' and logicians' health check. *International journal of medical informatics*, 78(Suppl 1), S86–S94. doi:10.1016/j.ijmedinf.2008.06.004.
- Simons, P. (2000). Continuants and occurrents: Peter Simons. *Aristotelian Society Supplementary Volume*, 74, 59–75.
- Smith, B. (2004). Beyond concepts: Ontology as reality representation. *FOIS – Formal ontology in information systems, 2004*, 73–84.
- Smith, B. (2018). Response to John Sowa. *Ontology Summit, 2018*.
- Smith, B., Ashburner, M., Rosse, C., Bard, J., Bug, W., Ceusters, W., Goldberg, L.J., Eilbeck, K., Ireland, A., Mungall, C.J., OBI Consortium, Leontis, N., Rocca-Serra, P., Ruttenberg, A., Sansone, S.A., Scheuermann, R.H., Shah, N., Whetzel, P.L. & Lewis, S. (2007). The OBO foundry: Coordinated evolution of ontologies to support biomedical data integration. *Nature biotechnology*, 25(11), 1251–1255. doi:10.1038/nbt1346.
- Smith, B., Ceusters, W., Klagges, B., Köhler, J., Kumar, A., Lomax, J., Mungall, C., Neuhaus, F., Rector, A.L. & Rosse, C. (2005). Relations in biomedical ontologies. *Genome biology*, 6(5). doi:10.1186/gb-2005-6-5-r46.
- SNOMED CT. <https://www.snomed.org/>. Accessed: August 20, 2023.
- SNOMED International SNOMED CT Starter Guide. <http://snomed.org/sg>. Accessed: August 20, 2023.
- SNOMED International (2023). SNOMED CT OWL Guide: <http://snomed.org/owl>. Accessed: August 20, 2023.

- Spackman, K. & Reynoso, G. (2004). Examining SNOMED from the perspective of formal ontological principles: Some preliminary analysis and observations. In *First International Workshop on Formal Biomedical Knowledge Representation KR-MED 2004*. CEUR-WS (Vol. 102, pp. 72–80).
- Spackman, K.A., Campbell, K.E. & Côté, R.A. (1997). SNOMED RT: A reference terminology for health care. In *Proceedings: A Conference of the American Medical Informatics Association. AMIA Fall Symposium* (pp. 640–644).
- Spackman, K.A., Dionne, R., Mays, E. & Weis, J. (2002). Role grouping as an extension to the description logic of ontolog, motivated by concept modeling in SNOMED. In *Proceedings of AMIA Symposium* (pp. 712–716).