JMIR Publications
The Leading eHealth Publisher

PCEtoFHIR: Decomposition of post-coordinated SNOMED CT expressions for storage as HL7 FHIR resources

Tessa Ohlsen, Josef Ingenerf, Andrea Essenwanger, Cora Drenkhahn

Submitted to: JMIR Medical Informatics
on: February 28, 2024

Disclaimer: © The authors. All rights reserved. This is a privileged document currently under peer-review/community review. Authors have provided JMIR Publications with an exclusive license to publish this preprint on it's website for review purposes only. While the final peer-reviewed paper may be licensed under a CC BY license on publication, at this stage authors and publisher expressively prohibit redistribution of this draft paper other than for review purposes.

Table of Contents

Preprint JMIR Publications

PCEtoFHIR: Decomposition of post-coordinated SNOMED CT expressions for storage as HL7 FHIR resources

Tessa Ohlsen[1, 2] MSc; Josef Ingenerf[1, 2] Dr rer nat, Prof Dr; Andrea Essenwanger[3] MSc; Cora Drenkhahn[2] MSc

[1]IT Center for Clinical Research University of Luebeck Luebeck DE
[2]Institute of Medical Informatics University of Luebeck Luebeck DE
[3]mio42 GmbH Berlin DE

Corresponding Author:
Tessa Ohlsen MSc
IT Center for Clinical Research
University of Luebeck
Ratzeburger Allee 160
Luebeck
DE

Abstract

Background: To guarantee interoperability, structural and semantic standards must be adhered to. For exchanging medical data between information systems, the structural standard FHIR (Fast Healthcare Interoperability Resource) is recently gaining traction. In terms of semantic interoperability, the reference terminology SNOMED CT as a semantic standard enables a post-coordination in comparison to many other vocabularies. These post-coordinated expressions (PCEs) result in SNOMED CT being an expressive and flexible interlingua, which enables precise coding of medical facts, but at the expense of increased complexity and challenges with storage and processing. In addition, the boundary between the scope of semantic (terminology) and structural (information model) standards blurs, the so-called TermInfo problem.

Objective: Although often discussed critically, the TermInfo overlap can also be explored for its beneficial potential by enabling a flexible transformation of parts of the PCEs. In this paper, an alternative solution for the storage of PCEs is presented, i. e., in combination with the FHIR data model. In the end, all components of a PCE should be expressible exclusively by pre-coordinated concepts that are linked to suitable elements of the information model.

Methods: The approach is based on storing PCEs and/or parts of them conforming to FHIR resources. Using the Web Ontology Language for generating an OWL ClassExpression in combination with an external reasoner and semantic similarity measures, a pre-coordinated SNOMED CT concept that describes the PCE most precisely is determined as a superconcept. In addition, the non-matching attribute relationships between the superconcept and the PCE are determined as the so-called delta. Once SNOMED CT attributes have been mapped to FHIR elements manually, FHIRPath expressions can be determined for the superconcept and the delta, which in turn enable the identified pre-coordinated codes to be stored in FHIR resources.

Results: A web application called PCEtoFHIR was developed to implement this approach. In a validation with 600 randomly selected pre-coordinated concepts, the correctness of the generated OWL ClassExpression could be confirmed. Additionally, two different approaches for calculating semantic similarities were validated, with the approach by Sanches et al. demonstrably yielding more precise results. In a validation of the entire approach, considering 33 already existing PCEs, the correct functionality could also be demonstrated.

Conclusions: PCEtoFHIR provides services to decompose PCEs for storing them in FHIR resources. When creating structure mappings concerning certain subdomains of SNOMED CT concepts (e.g. allergies) to desired FHIR profiles, the use of SNOMED CT Expression Templates has proven to be very useful. Domain experts can prepare templates with suitable mappings that can be reused in a constrained way by end users more easily.

(JMIR Preprints 28/02/2024:57853)
DOI: https://doi.org/10.2196/preprints.57853

Preprint Settings

1) Would you like to publish your submitted manuscript as preprint?

- ✓ **Please make my preprint PDF available to anyone at any time (recommended).**
- Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users.
- Only make the preprint title and abstract visible.
- No, I do not wish to publish my submitted manuscript as a preprint.

2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?

- ✓ **Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).**
- Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain v
- Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in <a href="http

Preprint JMIR Publications

Original Manuscript

Preprint
JMIR Publications

PCEtoFHIR: Decomposition of post-coordinated SNOMED CT expressions for storage as HL7 FHIR resources

Abstract

Background: To guarantee interoperability, structural and semantic standards must be adhered to. For exchanging medical data between information systems, the structural standard FHIR (Fast Healthcare Interoperability Resource) is recently gaining traction. In terms of semantic interoperability, the reference terminology SNOMED CT as a semantic standard enables a post-coordination in comparison to many other vocabularies. These post-coordinated expressions (PCEs) result in SNOMED CT being an expressive and flexible interlingua, which enables precise coding of medical facts, but at the expense of increased complexity and challenges with storage and processing. In addition, the boundary between the scope of semantic (terminology) and structural (information model) standards blurs, the so-called TermInfo problem.
Objective: Although often discussed critically, the TermInfo overlap can also be explored for its beneficial potential by enabling a flexible transformation of parts of the PCEs. In this paper, an alternative solution for the storage of PCEs is presented, i.e., in combination with the FHIR data model. In the end, all components of a PCE should be expressible exclusively by pre-coordinated concepts that are linked to suitable elements of the information model.
Methods: The approach is based on storing PCEs and/or parts of them conforming to FHIR resources. Using the Web Ontology Language for generating an OWL ClassExpression in combination with an external reasoner and semantic similarity measures, a pre-coordinated SNOMED CT concept that describes the PCE most precisely is determined as a Superconcept. In addition, the non-matching attribute relationships between the Superconcept and the PCE are determined as the so-called Delta. Once SNOMED CT attributes have been mapped to FHIR elements manually, FHIRPath expressions can be determined for the Superconcept and the Delta, which in turn enable the identified pre-coordinated codes to be stored in FHIR resources.
Result: A web application called PCEtoFHIR was developed to implement this approach. In a validation with 600 randomly selected pre-coordinated concepts, the formal correctness of the generated OWL ClassExpression was verified. Furthermore, 33 PCEs were used for two different validations. Based on these, it was shown that the calculation of the semantic similarity by Sánchez et al. is appropriate for the determination of the Superconcept. Secondly, these 33 PCEs were used to demonstrate the correct functioning of the entire approach. In addition, the FHIR StructureMaps were validated meaningful by FHIR experts.
Conclusion: PCEtoFHIR provides services to decompose PCEs for storing them in FHIR resources. When creating structure mappings concerning certain subdomains of SNOMED CT concepts (e.g. allergies) to desired FHIR profiles, the use of SNOMED CT Expression Templates has proven to be very useful. Domain experts can prepare templates with suitable mappings that can be reused in a constrained way by end users more easily.

Keywords: SNOMED CT; HL7 FHIR; TermInfo; post-coordination; semantic interoperability; terminology; OWL; semantic similarity

Introduction

Background

The increasing digitization of medical records has resulted in a larger amount of patient data available for analysis in healthcare. There is a need to utilize this data to enhance medical care and

provide more targeted treatments. However, to achieve this, it is necessary to have the ability to exchange and process data automatically between different systems. This not only requires technical compatibility, but also semantic interoperability, which ensures that the meaning of the data is retained when importing it into another computer system. The ability to transfer and work with data across systems is essential for leveraging the full potential of digital medical records and enhancing patient care [1].
To guarantee semantic interoperability, structural and semantic standards must be adhered to. Structural standards define the syntax for accessing data fields in information models. In recent years, the newly developed HL7 standard FHIR (Fast Healthcare Interoperability Resources) is gaining international popularity in this regard due to its focus on simplified implementation and up-to-date technologies [2]. Semantic standards, on the other hand, involve terminologies with language-independent codes to represent the data's meaning in an interoperable way. Here, SNOMED CT is known as the most expressive terminology in medicine that is used to improve semantic interoperability [3]. In 2021, Germany obtained a national license for SNOMED CT leading to increased interest and usage of the terminology. SNOMED CT concepts are being incorporated into data modeling, such as the Medical Information Objects of the German National Association of Statutory Health Insurance Physicians (NASHIP) [4] and the core data set of the Medical Informatics Initiative (MII) [5]. The use of SNOMED CT – over 350.000 concepts exist here – is to provide a machine-usable interlingua that minimizes coding issues specific to countries and medical fields. Due to the complexity of natural language, not all medical circumstances can be precisely coded using SNOMED CT's large number of pre-coordinated concepts. To avoid a rapid increase in the number of new concepts, and in contrast to many other vocabularies, SNOMED CT enables post-coordination, which allows the combination of pre-coordinated concepts into new expressions using a formal grammar.

Therefore, post-coordination is a unique feature that is of great value for the precise recording of medical facts and thus largely increases SNOMED CT's expressive power. However, the use of post-coordination is currently lagging due to various difficulties. Some of these hurdles have already been addressed [6–9], yet the integration of post-coordinated SNOMED CT expressions (PCEs) into the electronic health records (EHRs) of legacy hospital information systems remains challenging due to several reasons:

1. *Adherence to familiar data structures*
   Medical circumstances are typically described using individual codes, and there are established methods for storing and processing these. Restrictions such as length-limited data types can be tolerated when using simple codes, but not with arbitrarily large formal expressions. Because of this, there are concerns about the use of PCEs [10].
2. *Lack of technical support*
   The technical handling of PCEs is inherently complex, requiring at least a description logic reasoner and the implementation of several formal specifications (for example: Concept Model, Compositional Grammar, Expression Constraint Language) by SNOMED International [3]. While specialized terminology servers may reduce the implementation burden in general, only the CSIRO Ontoserver [11] supports post-coordination as of now [12].
3. *Difficulties with FHIR Search*
   When searching for information in FHIR resources, post-coordination is only supported if the exact same PCE is explicitly and previously defined in a FHIR CodeSystem supplement. Such supplements allow for the extension of the standard FHIR CodeSystem for SNOMED CT with a collection of PCEs (e.g., for value set definition) but don't facilitate the recording of PCEs in a patient's EHR.

Because these challenges will likely not be solvable in the short term, a different approach is required to ensure interoperability between systems that support post-coordination and those that don't. Therefore, an alternative representation of post-coordinated SNOMED CT expressions will be presented in this paper.

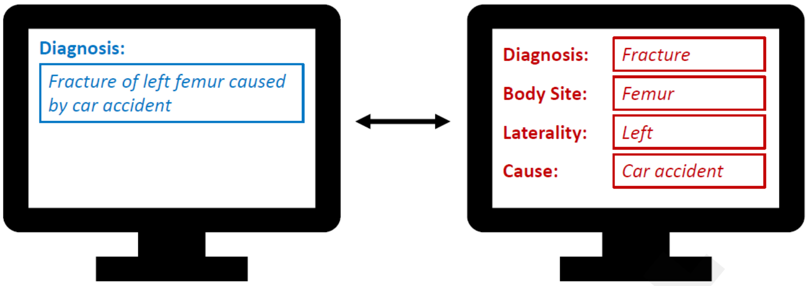


Figure 1. TermInfo: The same medical fact can be represented employing either the terminology (left side) or the information model (right side) more heavily.

As known for a long time, there is significant overlap between the scope of structural and semantic standards resulting in unclear responsibilities and therefore ambiguous representations of medical facts and possibly inconsistent redundancies. This so-called TermInfo problem [3,13–15] largely emerged due to the independent development of these standards leading to mutual coverage of the same data elements on both sides (see Figure 1). SNOMED CT‘s ability for post-coordination further blurs the line between structure and semantics, and may thus be interpreted to worsen the existing problem. Post-coordinated expressions are fully interpretable though, so that the information components they are comprised of may be identified and flexibly taken apart. Based on this idea, extensive knowledge gained in previous projects [6,16–18] and the current synergy of using SNOMED CT with FHIR, the authors propose PCEtoFHIR, an application to decompose post-coordinated SNOMED CT expressions for storage as FHIR resources in a meaning-preserving way.

Related work

While an increasing number of publications address the post-coordination of SNOMED CT concepts [6,16,19–21], a literature search revealed no existing publications on the storage of post-coordinated SNOMED CT expressions in FHIR resources. Nevertheless, a HL7 International working group addresses the topic and provides several resources: For selected FHIR resources, such as *Condition* and *Observation*, a mapping of some SNOMED CT attributes to FHIR is currently available (see [22] and [23]). In addition, the Confluence pages "SNOMED on FHIR" ("Bindings to FHIR Clinical Resources") [24] show various options for mapping SNOMED CT attributes to FHIR while avoiding semantic overlaps. The information contained within these documents is considered in our work.
Apart from this, there are some papers addressing the use of SNOMED CT in combination with standardized information models based on HL7 Reference Information Model (RIM), HL7 Clinical Document Architecture (CDA), HL7 FHIR resources. A project by Perez-Rey et al. [25] is generally concerned with linking the normal form of pre-coordinated SNOMED CT concept definitions. SNOMED CT concepts are normalized and then bound to the HL7 RIM classes. In general, this approach could be extended to PCEs. However, the HL7 v3 standard has not been widely adopted due to its complexity [1]. A project by Arguello-Casteleito et al. [26] deals with a mapping of pre-coordinated SNOMED CT concepts or post-coordinated SNOMED CT expressions from Consolidated Clinical Document Architecture (C-CDA) to FHIR resources. The objective of this approach resembles the approach taken in our work. However, the approach by Arguello-Casteleito et al. intensely focuses on an ontology. In contrast, our work primarily utilizes the widely adopted FHIR and SNOMED CT native specifications. Furthermore, the specific ocular diseases used by Arguello-Casteleito et al. only pertain to a very specific subset of SNOMED CT expressions. Additionally, the publication by Arguello-Casteleito et al. does not provide data regarding to what extent all information included in the PCEs could be transferred into the FHIR representation.

Methods

Overview

The aim of this work is to develop and implement an approach that enables the storage of a SNOMED CT post-coordinated expression (PCE) within FHIR resources by using only pre-coordinated codes. For this alternative representation, the PCE shall be decomposed into pre-coordinated concepts, which can then be stored in suitable elements of matching FHIR resources. An overview of the envisioned approach is shown in Figure 2.
A PCE, which is first checked for syntactic and semantic correctness, serves as input. This can be classified within SNOMED CT by using OWL and a reasoner, whereby the direct supertype ancestors can be determined. Of these concepts, the most similar concept to the PCE is determined (Superconcept). The Delta can then be calculated between the Superconcept and the PCE, which includes all the information of the PCE that the Superconcept does not represent. In the final step, appropriate elements of matching FHIR resources must be identified to store the information of the Superconcept and the Delta. Therefore, FHIR StructureMaps defining these associations on a general level need to be created beforehand.

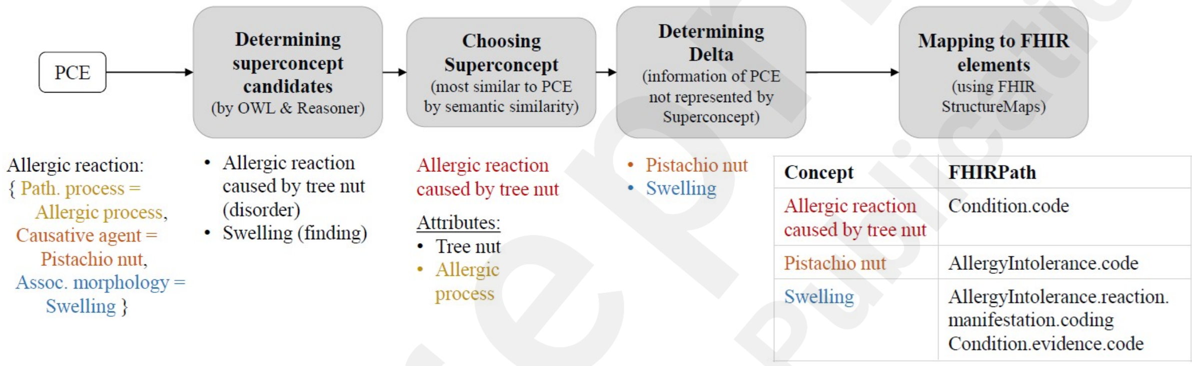


Figure 2. The decomposition of a PCE into elements of (profiled) FHIR resources consists of four steps.

Validation of PCE

To ensure that the flexible PCEs can be interpreted and evaluated, they must comply with the syntactic requirements of the Compositional Grammar and the semantic rules of the Concept Model defined by SNOMED International. In the first step, the input PCE is checked for syntactic and semantic correctness, using the HL7 FHIR service *$validate-code* and Ontoserver. Ontoserver is a FHIR-based terminology server provided by the Australian company CSIRO. Ontoserver supports and facilitates working with important coding systems such as SNOMED CT or LOINC [11]. It furthermore provides an integrated description logic reasoner and full support of SNOMED CT post-coordination, so that PCE validation can be carried out on the Ontoserver directly [11]. Only if a PCE is syntactically and semantically correct, the process will automatically continue with the next steps.

Determining Superconcept candidates

The Web Ontology Language (OWL) is an exchange format for ontologies. Here, it is used to determine the ancestors of a PCE. By using OWL, complex knowledge about concepts and relationships between them can be represented [27,28]. SNOMED CT can also be represented as an OWL ontology, and the definitions of individual concepts of SNOMED CT can be represented as OWL expressions, which are moreover contained in the monthly release packages of SNOMED CT

Preprint JMIR Publications

besides other formats [29].

To obtain SNOMED CT as an OWL ontology, the SNOMED OWL Toolkit [30] was used on the Release Format 2 of the International Edition, 2022-04-30 of SNOMED CT [31]. The generated ontology contains, among other components, pre-coordinated SNOMED CT concepts and their definitions in functional syntax.

PCEs on the other hand are based on the syntax of the Compositional Grammar. Thus, a PCE needs to be transformed into an OWL ClassExpression for further processing. For each component of the PCE the respective OWL counterpart is determined according to the existing concept definitions and as shown in Table 1. An algorithm using the Java library OWL API [32] is developed to automatically carry out the transformation. The result of the algorithm is an OWL ClassExpression that is structured in the same way as the OWL ClassExpression created by SNOMED International for the definitions of SNOMED CT concepts. An exemplary OWL expression for a PCE is shown in Figure 3, with Fully Specified Names added for improved readability. The symbol ":" is a placeholder for the defined namespace, here for "*http://snomed.info/id/*" [29].

Table 1. Various OWL constructs are used for the representation of the components of a PCE. While the components in the first two rows are exclusively employed in the native OWL ontology of SNOMED CT, the OWL constructs below are also used for the transformation of PCEs.

PCE component	**OWL construct**
SNOMED CT concept	OWL Class
SNOMED CT attribute, attribute value: SNOMED CT concept	OWL ObjectProperty
Linking of ungrouped attribute relationships	OWL ObjectIntersectionOf
Linking between individual attribute relationships in a Role Group	
Linking between focus concept and all attribute relationships of ungrouped attributes and Role Groups	
Attribute relationship of attribute and SNOMED CT Identifier as attribute value	OWL ObjectSomeValuesFrom
All grouped attribute relationships and Role Group Identifier	

In the ... eated PCE-specific OWL ... E into the existing SNO... ed through logical infere... superrelationships betwe... reasoners capable of pro...
The ... Java library elk-reaso... ated SNOMED CT conce... andidates.

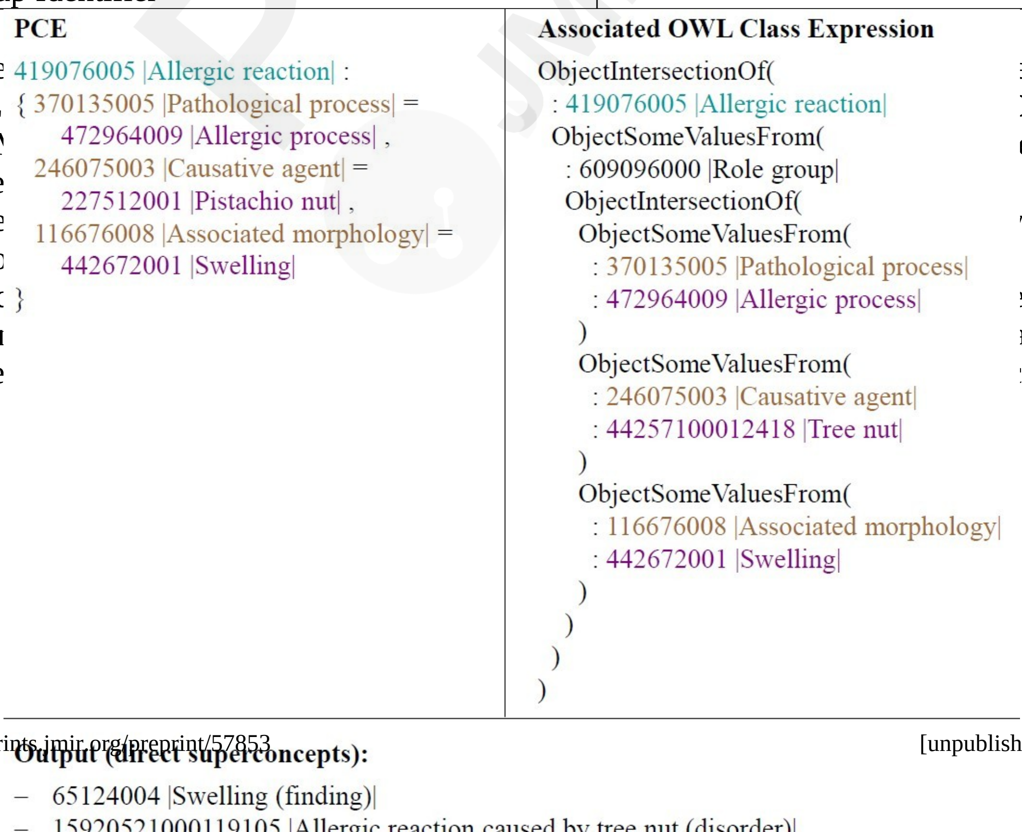


Figure 3. PCE on the left side and the associated OWL ClassExpression based on functional syntax on the right side.

Choosing Superconcept

One of the previously determined Superconcept candidates needs to be selected as the Superconcept. The Superconcept is the pre-coordinated SNOMED CT concept, which is most similar to the PCE, meaning that it covers the largest share of information contained within the PCE compared to any other pre-coordinated SNOMED CT concept.

Semantic similarity measure

To identify the most similar concept, the semantic similarity between concepts is determined. It is a measure for calculating the taxonomic proximity between two elements within a knowledge base, such as SNOMED CT. If two elements are more similar, the semantic similarity is higher [34].
While several semantic similarity measures are proposed in the existing literature, a path-based approach by Sánchez et al. is employed in our work as it was specifically developed for large knowledge bases with a subtype-relationship-based polyhierarchy [34], such as SNOMED CT. The equation for the calculation is as follows:

$$\text{¿}(c_1, c_2) = -\log_2\left(\frac{|T(c_1) \cup T(c_2)| - |T(c_1) \cap T(c_2)|}{|T(c_1) \cup T(c_2)|}\right).$$

Here, $T(c_i)$ is the set of all ancestors of a concept c_i and the concept c_i itself. To calculate the semantic similarity between two concepts c_1 and c_2, the ratio between the set of non-shared ancestors (nominator) and the union of all ancestors of both concepts (denominator) is considered [34].
In our work, the semantic similarity between a classified PCE and each of its Superconcepts needs to be calculated. Therefore, the equation is modified as

$$\text{¿}(pce, c_i) = -\log_2\left(\frac{|T(pce) \cup T(c_i)| - |T(pce) \cap T(c_i)|}{|T(pce) \cup T(c_i)|}\right).$$

with c_i being one of the Superconcepts. The calculated similarities for the exemplary PCE are shown

Preprint JMIR Publications

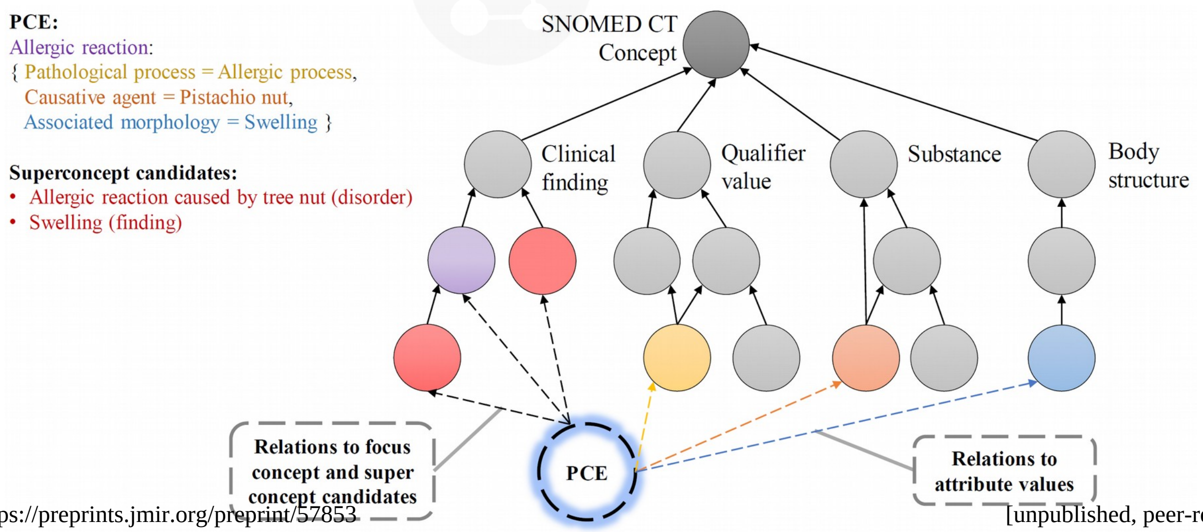
in Table 2.

Table 2. To choose the most fitting Superconcept, the semantic similarity between the PCE and each of its Superconcept candidates is calculated using a measure by Sánchez et al. The previously introduced exemplary PCE shares a larger ratio of ancestors with *Allergic reaction caused by tree nut* than with *Swelling*. Thus, the former leads to a higher semantic similarity and is determined as Superconcept.

Superconcept candidates	**Non-shared ancestors**	**Union of all ancestors**	**Semantic similarity**
Allergic reaction caused by tree nut	8	30	2,10
Swelling	29	30	0,10

Implementation

An algorithm was developed which selects the most suitable, i.e. the most semantically similar SNOMED CT concept from the Superconcept candidates. The calculation of the semantic similarity is based on a graph-based approach by Sánchez et al. as described above. For this, SNOMED CT must be transformed into a directed acyclic graph (DAG) once beforehand. For constructing the DAG, the Release Format 2 (RF2), version 2023-04-30 [31] of SNOMED CT was algorithmically processed using the Python library NetworkX [35]. In the resulting graph, SNOMED CT concepts are represented via nodes while their relations form the connecting edges, including the relation types.
To now enable semantic similarity calculation, the respective PCE must be temporarily inserted into the DAG. For this purpose, a node called “pce” is introduced into the graph as a subnode of its focus concept and the previously determined Superconcept candidates (see Figure 4). The algorithm iterates over the individual attribute values of the PCE and inserts the edges between these and the node “pce”.
Next, the semantic similarity between the PCE and all possible Superconcept candidates is calculated based on the DAG. The Superconcept candidate with the highest semantic similarity describes the PCE most accurately of all pre-coordinated SNOMED CT concepts. This concept is defined as the Superconcept.



Preprint JMIR Publications

Figure 4. The PCE is represented as a sub-concept of its focus concept and the super-concept candidates. It also has edges to its attribute values.

Determining Delta

While it is the most similar pre-coordinated concept available in SNOMED CT, the Superconcept does not cover all the information contained within the PCE. Therefore, the missing parts of information shall now be determined which is referred to as the "Delta" between the PCE and the Superconcept. A graph-based approach is employed once again, representing both the attribute relations of the PCE and those of the Superconcept in separate graphs. To allow merging later on, a dummy root node of the same name is introduced into both graphs. Thereupon, the attribute relations of the PCE or the Superconcept are respectively added as further nodes and connected via edges, as shown on either side of Figure 5. The Delta is then calculated by subtracting the edges of the Superconcept graph from the PCE graph. As a result, all equivalent components available on both sides are eliminated, leaving only those attribute relationships not or not precisely represented in the Superconcept graph (see the Delta graph in the middle of Figure 5).

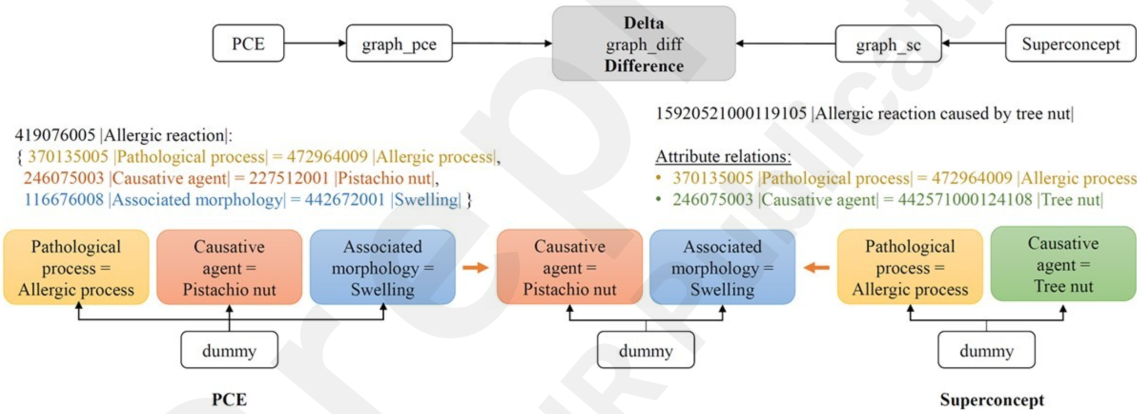


Figure 5. When calculating the Delta, the graph of the Superconcept is subtracted from that of the PCE. The determined edges contain attribute relations that were not covered by the Superconcept but are necessary for the precise representation of the PCE.

Mapping to FHIR elements

In a final step, the Superconcept and the Delta must be stored as pre-coordinated concepts in suitable FHIR elements. Either the FHIR base resources from HL7 International or more specific FHIR profiles can be used for this. FHIR profiles are certain resources that are customized to national characteristics and legislation as well as special use [35].

In our work, two sets of FHIR profiles were considered as target representations for the mapping:

- profiles of the German National Association of Statutory Health Insurance Physicians (NASHIP, version 1.4.0, based on FHIR R4) [4]
- profiles of the core data set of the German Medical Informatics Initiative (MII, version 1.0, based on FHIR R4) [5].

To accurately map a decomposed PCE to these profiled FHIR resources, mapping rules are required to conjoin its Superconcept and Delta with the correct FHIR elements. While PCEs are often highly individualized expressions, they may be collated into categories according to their focus concept, attribute relationships, and the SNOMED CT Expression Templates. Therefore, it is not necessary to define a mapping rule for each individual PCE; instead, PCEs with similar content can be mapped with a uniform set of rules.

To demonstrate the applicability of our approach, five content categories of PCEs from the Top-level hierarchies *Procedure* and *Clinical finding* are considered here, including “General procedures” for the former and “Allergies”, “Diseases due to allergy”, “Allergic reactions”, and “General clinical findings” for the latter. For the unambiguous assignment of PCEs later on, SNOMED CT’s Expression Constraint Language (ECL) is used to formally define the scope of each category and therefore form pairwise disjoint partitions of the two hierarchies.

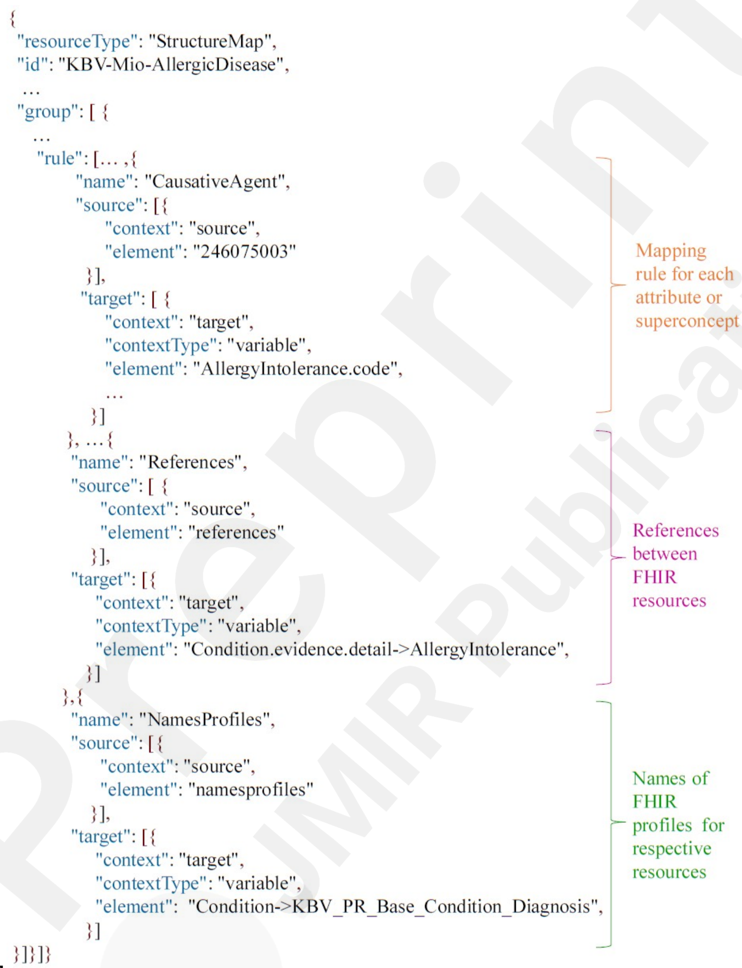


Figure 6. Extract of the FHIR StructureMap for allergic reactions in JSON format. It contains an entry for each SNOMED CT attribute and the Superconcept as well as references and names of the profiles.

For each combination of one of the five content categories and either of the two sets of target representations, mapping rules are manually defined based on the respective Expression Template [36], concept definitions of related pre-coordinated concepts, existing documents (as described in “Related work”), and the author’s expertise. These rules are formally transcribed via FHIR StructureMaps (R4) [37] using the Java library HAPI FHIR [38]. The resulting ten FHIR StructureMaps are stored on a local HAPI FHIR server [39].

An excerpt of the FHIR StructureMap for the category “Allergic reaction” and the NASHIP profiles is shown in Figure 6. Each StructureMap contains exactly one “rule” entry for the Superconcept and for each attribute relationship relevant to the category. The mapping rule’s “source” is thus either the

Preprint JMIR Publications

string “Superconcept” or the SNOMED CT identifier of the respective attribute. For the “target” element, FHIRPath is used to traverse the FHIR resources and explicitly identify the correct sub-element [2]. Illustrated in orange in Figure 6, the conjunction of the attribute “Causative agent” (“246075003”) and the FHIRPath “*AllergyIntolerance.code*” is exemplified. This specifies that the associated attribute value of the PCE shall be stored in the element *code* of the resource *AllergyIntolerance*.
Apart from the central mapping rules, some further information is included in the FHIR StructureMaps.
If necessary, references between FHIR elements and resources are (Figure 6, in pink). Furthermore, the FHIR resources of the FHIRPath entries are referred to the names of the respective profiles (Figure 6, in green).

With these general preparations concluded, singular decomposed PCEs can now be mapped to appropriate FHIR resources. Therefore, the PCE is firstly assigned to one of the five content categories by determining the subsuming subset via the ECL definitions and Ontoserver. According to the category and the desired target representation, the correct FHIR StructureMap is allocated. Its mapping rules are automatically executed on the PCE’s Superconcept and Delta, determining the combinations of FHIR element and pre-coordinated SNOMED CT concept necessary for an alternative representation.

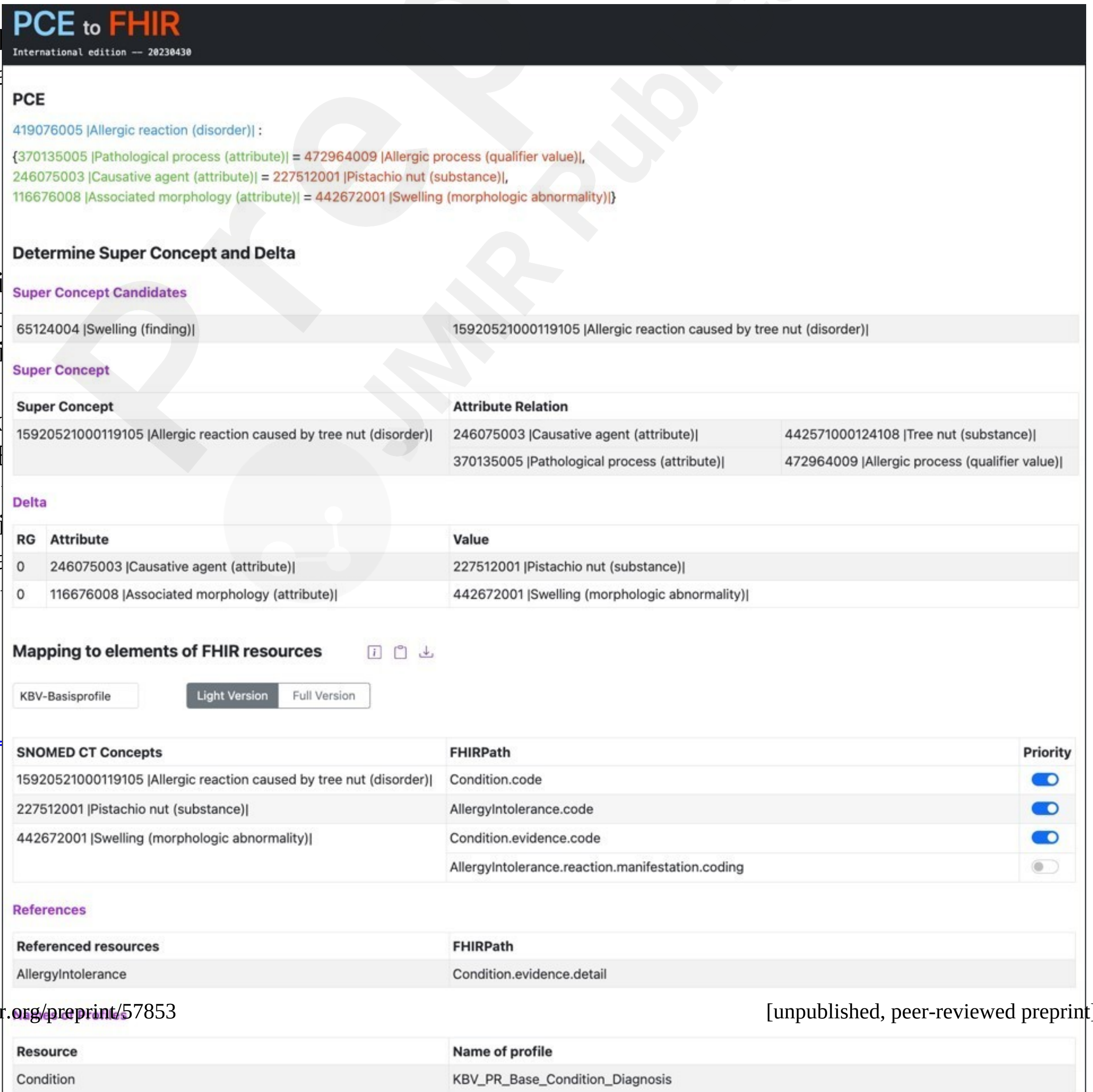
Results

Web application

Bringing al
web applica
framework
version 17)

An excerpt
syntactical
are determi
Super Conc
as defined i
determined
combo box
English: KI
Health Ins
correspondi
displayed, a
7 in the sec

The source
described
https://githu



Preprint JMIR Publications

Figure 7. Excerpt of the web application PCEtoFHIR.

Determining Superconcept candidates

To ensure the correct functionality of the algorithm for generating the OWL ClassExpressions, a validation with 600 randomly selected concept definitions of pre-coordinated SNOMED CT concepts was performed. These concept definitions in compositional grammar, published in the monthly release 2023-04-30 [31], were first examined for being of diverse structure and subsequently transformed into OWL ClassExpressions by PCEtoFHIR's regular algorithm. Afterwards, these generated OWL Class Expressions were compared to the concepts' official OWL ClassExpressions, available as part of the same SNOMED CT release, as reference data.

The comparison involved checking whether both OWL ClassExpressions match syntactically and semantically. Hereby, the focus concepts needed to be excluded from the semantic validation because the OWL ClassExpressions of the reference data is based on the Stated Concept Definitions, while the Inferred Concept Definitions are used in our approach. Therefore, the validation only considered whether a focus concept was syntactically in the correct position in the OWL ClassExpression.

In summary, the comparison revealed no discrepancies between the OWL ClassExpressions generated via PCEtoFHIR and those of the reference data.

Choosing Superconcept

As described previously, a path-based measure by Sánchez et al. [34] is applied to calculate semantic similarity in our work. A preliminary analysis was conducted to ensure both the theoretical as well as the practical applicability of this measure.
SNOMED CT's most striking characteristics include its reliance on subtype relationships and the resulting, heavily interwoven polyhierarchy. By utilizing a path-based approach which incorporates each concept's ancestors, these central features are prioritized. While there are several path-based semantic similarity measures available [34], they mostly rely on the shortest path between concepts [40–43]. Sánches et al. argue that in large knowledge bases, such as SNOMED CT, a concept inherits information from several hierarchies and is connected to many concepts simultaneously, and that considering only the shortest path is therefore insufficient.
In their proposed measure, the ratio between the set of non-shared ancestors and the set of all ancestors of both concepts is considered in contrast. Adapted to our approach, both for the PCE and each of its Superconcept candidates, all ancestors in the SNOMED CT hierarchy may be determined and the ratio calculated. Thus, this measure is applicable in principle.

To evaluate if the calculated values are reasonable beyond that, a practical validation by means of an exemplary sample was done in succession. For 33 PCEs taken from a publication by Kate et al. [44] (see detailed explanation in "Overall evaluation") the most similar Superconcept candidate was determined manually and compared to the Superconcept calculated by PCEtoFHIR via Sánchez et al.'s semantic similarity measure. In 76% (25 in 33) of all cases, the same Superconcept was chosen. An analysis of the remaining 24% revealed that different information components contained within the PCE were prioritized during Superconcept selection (e.g., favoring localization over procedure type) but that the divergent choice of the algorithm was arguably plausible. As a result, the measure of Sánchez et al. was found feasible for Superconcept determination and is thus used to calculate semantic similarity in this work.

Mapping to FHIR elements

As previously described, the Basisprofile of the NASHIP (version 1.4.0) [4] and the profiles of the Core Data Set of the MII (version 1.0) [5] were used for mapping to the FHIR profiles. To illustrate the mapping, five categories were considered. For each category and profile type, a FHIR StructureMap was created, resulting in a total of ten FHIR StructureMaps. Table 3 shows the FHIR resources that were used for the mapping.

Table 3. Depending on the category and profile type, different FHIR resources are used, which are listed here.

Category	**Profile type**	**FHIR resources**
Allergies	NASHIP	AllergyIntolerance
	MII	Condition, Observation
Disease due to allergies	NASHIP	AllergyIntolerance, Condition
	MII	Condition, Observation
Allergic reaction	NASHIP	AllergyIntolerance, Condition
	MII	Condition, Observation
Clinical finding (general)	NASHIP	AllergyIntolerance, Condition
	MII	Condition, Observation
Procedure	NASHIP	Procedure
	MII	Procedure

For the category of general *Clinical findings*, the FHIR StructureMaps include mappings for the Superconcept and the following five SNOMED CT attributes:

- *Causative agent, Finding site, Associated morphology, Pathological process, Clinical course.*

In addition to the five SNOMED CT attributes listed above, the following attributes were considered for the remaining *Clinical finding* categories focusing on allergies:

- *Has realization, Occurrence, Due to.*

Furthermore, for hierarchy *Procedure*, mapping rules for the Superconcept and the following 12 additional SNOMED CT attributes are established:

- *Method, Procedure site – Direct, Procedure site – Indirect, Direct substance, Direct morphology, Using substance, Using device, Using access device, Has intent, Access, Surgical approach, Has focus.*

Table 4 shows exemplary the mapping of the SNOMED CT elements to the FHIRPath for *Allergic reactions*. The complete set of mapping rules and the associated StructureMaps are available at: https://itcr-uni-luebeck.github.io/pce-to-fhir/.

Table 4. The SNOMED CT elements and the associated FHIRPath for the category ”Allergic reaction” based on the profiles of the NASHIP.

SNOMED CT element	NASHIP
Super concept	Condition.code
Causative agent	AllergyIntolerance.code
Finding site	Condition.bodySite
Associated morphology	AllergyIntolerance.reaction. manifestation.coding:snomed Condition.evidence.code
Pathological process	AllergyIntolerance.reaction.manifestation.coding:snomed Condition.evidence.code
Has realization	AllergyIntolerance.reaction.manifestation.coding:snomed Condition.evidence.code
Occurrence	AllergyIntolerance.extension: abatement-lebensphase-von [45]
Clinical course	Extension of HL7 International: Condition.condition-diseaseCourse [46]
Due to	Extension of HL7 International: Condition.condition-dueTo [47]

To ensure their correctness, the created FHIR StructureMaps were validated by author AE who was not involved in the PCEtoFHIR project up to this point. AE has in depth knowledge of FHIR and SNOMED CT. Due to her former and current work, she is profoundly familiar with the MII and NASHIP profiles. Based on the definitions of the profiles used and her expertise, the FHIR StructureMaps were validated, including the correct choice of profiles, the mapping of SNOMED CT elements and associated FHIRPaths as well as possible references. The validation yielded the following results: The choice of profiles and the possible references were found to be entirely correct. The mapping rules between SNOMED CT elements and FHIR paths were largely considered correct as well, but here a total of eight suggestions for improvement were provided. These suggestions were reviewed by the other authors and upon agreement, the FHIR StructureMaps were changed accordingly.
Lastly, the finalized FHIR StructureMaps were analyzed for their coverage of the SNOMED CT attributes listed above. Table 5 illustrates the quantity of attributes per category that could be successfully mapped in the respective profile (“mappable”). Depending on the category, up to four attributes could not be mapped to the native profiles (“unmappable”). For some of these, existing FHIR extensions, such as those of HL7 International, can be introduced to the profiles to achieve a more complete mapping (“with extension”). Overall, between 55.6% and 92.3% of attributes (76,1% on average) can be mapped without modifications, depending on category and profile. By introducing the aforementioned extensions to the profiles, coverage could be increased to an average of 93,5% (between 66,7% and 100% for individual combinations).

Table 5: The number of SNOMED CT elements that can be mapped directly to the profiles are represented by "mappable", whereas elements that cannot be mapped are shown as "ummappable". FHIR offers extensions to map items that are not mappable by default, which could reduce the number of unmappable elements. The number of unmappable elements to represent by extension are shown in the last column.

Category	**#elements Total**	**Profile type**	**#elements mappable Total**	**#elements unmappable**	
				Total	***Could be mapped using extensions***
Allergies	9	NASHIP	5	4	*1*
		MII	5	3	*2*
Disease due to allergies	9	NASHIP	7	2	*2*
		MII	7	2	*2*
Allergic reaction	9	NASHIP	7	2	*2*
		MII	7	2	*1*
Clinical finding (general)	6	NASHIP	4	2	*2*
		MII	4	2	*2*
Procedure	13	NASHIP	12	1	*1*
		MII	12	1	*1*

Overall evaluation with existing PCEs

After validating several steps on their own, the whole process of PCEtoFHIR shall be evaluated. To achieve a rather realistic scenario, 35 existing PCEs were used, which are provided as part of the publication "Automatic full conversion of clinical terms into SNOMED CT concepts" by Kate [44,48]. This publication presents a method for the conversion of clinical texts into post-coordinated SNOMED CT expressions in which the 35 PCEs were manually created from clinical terms for a small-scale evaluation (see "4.3. Evaluation methodology").

To use this dataset as input for PCEtoFHIR, the 35 provided PCEs were first reviewed manually as well as automatically checked for syntactic and semantic correctness as described before. As a result, two of the PCEs were excluded from further processing: One for violating the cardinality restrictions of the Concept Model, and one for being equivalent to another. The remaining 33 PCEs consist of 23 PCEs of the Top-level hierarchy *Clinical finding* and 10 PCEs of the Top-level hierarchy *Procedure*. The following SNOMED CT attributes are used:

- *Clinical finding*: *Associated morphology*, *Finding site*, *Causative agent*, *Clinical course*, *Finding method*, *Pathological process*
- *Procedure*: *Method*, *Procedure site – Direct*, *Using device*, *Using substance*.

These 33 PCEs were imported into PCEtoFHIR in bulk, bypassing the web frontend but using the developed algorithm as usual. Each PCE was therefore decomposed into Superconcept and Delta, for which the FHIRPaths were determined using the appropriate FHIR StructureMap. Based on the FHIRPaths, the corresponding attribute values were fed into FHIR resources which had been prepopulated with further required data elements beforehand (like reference to a FHIR resource "Patient"). The resulting FHIR resources for each PCE were stored on the local HAPI FHIR server [39].

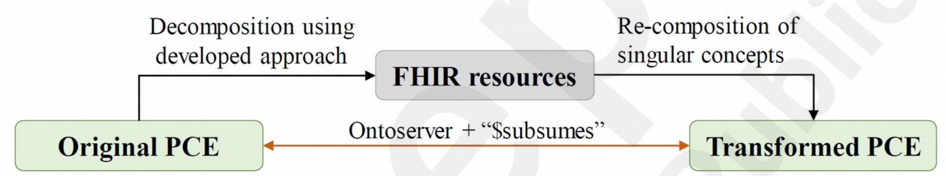


Figure 8. Validation process: The original PCE is decomposed by PCEtoFHIR as usual, and accordingly stored in FHIR resources. Based on this representation, the singular concepts are recomposed into a second PCE. The original and the recomposed PCE are compared.

To determine whether all information contained within the original PCE has been successfully translated into the FHIR resources, the process shall now be reversed as shown in Figure 8. Based on the separate pre-coordinated SNOMED CT concepts spread across multiple FHIR elements, a PCE shall be recomposed and compared to the original. Therefore, the values of the stored FHIR resources were extracted and the respective StructureMap applied in reverse direction. The recomposed PCE uses the Superconcept as the focus concept, while its refinement consists of attribute relationships with the extracted SNOMED CT concepts as attribute values. For each of the extracted concepts, the corresponding SNOMED CT attribute must be determined according to the StructureMap's mapping rules. Like most mappings, the created StructureMaps are inherently unidirectional [3], and using them in reverse direction leads thus to several possible SNOMED CT attributes for some FHIRPaths. In these circumstances, the following three rules are applied successively until a distinction of cases is achieved:

1. *By attribute value range:* For each of the attributes in question, a value range is defined by the SNOMED CT Concept Model. This allows to check if the extracted concept is contained in the respective value range and is therefore a feasible value for the corresponding attribute. If only one value range matches, the correct attribute is found.
2. *By attribute hierarchy:* Some SNOMED CT attributes are arranged hierarchically (e.g., *Procedure device – Using device – Using access devices*). If several attributes of the same

hierarchy are available, the most general attribute of these is used (here: *Procedure device*).

3. *By occurrence heuristic:* The concept definitions of pre-coordinated concepts are analyzed to determine which SNOMED CT attribute is statistically most likely used for the concept in question.

Finally, each recomposed PCE was compared to the corresponding original PCE as reference by testing their subsumption relationship via Ontoserver and the FHIR service *$subsumes*. For 32 of the 33 comparisons, both PCEs were evaluated as equivalent, signifying that no information was lost at any time during the process. For the remaining comparison, the recomposed PCE was found to be a subclass of the original. Further analysis showed that some semantic precision was lost due to the needed case distinction but that the FHIR representation achieved through PCEtoFHIR is still semantically equivalent to the original PCE. The correct functioning of the approach could therefore be proven including the preservation of a PCE’s content.

Discussion

Principal Results

The aim of this work was to develop an algorithm that enables an alternative representation for post-coordinated SNOMED CT expressions. The introduction discussed various reasons for the reluctance to use PCEs. One reason was the familiar approach of editing and saving individual codes. With pre-coordinated concepts, a single code is available for a medical circumstance as well as a human-readable description. These two aspects are seen as advantages when using pre-coordinated concepts but are not given when using post-coordination. However, by identifying a Superconcept based on OWL and semantic similarity, it is possible to identify a pre-coordinated SNOMED CT concept that most accurately describes the PCE. The PCE is a subconcept of the determined Superconcept that can be displayed and analyzed. In addition, the decomposition of the PCE leads to individual codes that are familiar to the user from practical experience. FHIR is also regarded as one of the central interoperable data standards in the healthcare sector and is also becoming increasingly important in Germany [49] and worldwide [49–51].

The calculation of semantic similarity is essential for identifying the concept most similar to a PCE. Here, the measure by Sánchez et al. [34] was chosen as it suits SNOMED CT’s polyhierarchies and multiple inheritances. While this measure was evaluated to be a workable solution, other approaches to semantic similarity calculation are available and may be considered alternatively. In this regard, the choice of an appropriate Superconcept is obviously important, but any remaining information is represented through the Delta anyway and thus a different semantic similarity measure will ultimately have little effect on the overall functionality of PCEtoFHIR. However, PCEtoFHIR could be extended by enabling the user to change the Superconcept manually.

As described above, the Superconcept and the Delta are stored in the respective FHIR profiles and there in the matching elements. The profiles of the National Association of Statutory Health Insurance Physicians (NASHIP) [4] and the profiles of the core dataset of the Medical Informatics Initiative (MII) [5] were used for this purpose. However, the approach can also be adapted to any other profile. As shown in Table 5, depending on the content category and profile type, between 55.6% and 92.3% of attributes could be mapped directly to the respective profile without modifications, indicating a large coverage of content but also some gaps throughout.

An analysis of the unmappable attributes revealed that these relations mostly concern very specific details that only occur in very few definitions of pre-coordinated concepts. As PCEs are generally built adjacent to existing concept definitions, they will likely neither draw upon these attributes regularly. SNOMED CT however offers the possibility to include medical facts to such a granular

level. HL7 FHIR on the other hand was designed with a pragmatic paradigm in mind, focusing on the most prevalent information. Therefore, this fundamental discrepancy in design accounts for the majority of mapping hindrances.
Nevertheless, and as intended by the FHIR standard, there are suitable extensions available from HL7 International that would appropriately represent some of these attribute relations (see Tables 4 and 5) and could therefore be integrated into the profiles to extend their coverage of PCE content. In a similar case, the core data set does not include a profile for the FHIR resource *AllergyIntolerance* as of now, hindering the representability of this content category. With the CDS being an ongoing modeling initiative, such a profile may be added in the future though.

Apart from the unmappable attributes, the proposed mapping rules cannot always ensure an exact translation. In some cases, there is no precisely matching FHIR element for a specific attribute, or several attributes must be mapped to the same element (e.g., *Using access device* and *Using device* can both only be recorded via the FHIR element *Device.type*, which is then referenced via *Procedure.usedReference)*. As a result, some semantic precision is threatened to be lost. Therefore, the original PCE should always be stored in the metadata of the FHIR representation to assuredly preserve the original meaning.
Nevertheless, the outlined difficulties of achieving a semantically equivalent representation through FHIR elements once again illustrate the precision achievable through SNOMED CT's post-coordination and thus its importance for describing medical data in detail.

Several validations were conducted to ensure the correctness of both the individual processing steps as well as the overall functionality of PCEtoFHIR. While a large dataset could easily be taken from SNOMED International's releases to firmly validate the OWL expression generation, other evaluations necessitated pre-existing real-world PCEs of specific content categories which are not readily available. The employed set of 35 PCEs meets these criteria and could therefore essentially facilitate our validation approaches. However, both the amount and the semantic variance of these exemplary PCEs is limited so that the significance of the results could be enhanced by incorporating further reference data.

Another validation included the manual review of the FHIR StructureMaps by a FHIR expert. This revealed only minor inaccuracies, which were remedied in the current version of the StructureMaps. By using the FHIR StructureMaps, the mapping could be successfully solved for the considered categories and can be extended to other sub-hierarchies of SNOMED CT without any difficulties. Hence, further possibilities of application may be considered, e.g., regarding the TermInfo problem: As stated above, the decision to choose either terminology or information model as the place to represent medical facts is often variable and depends on the intended use. For example, "Fracture of the left femur" can either be represented by using a single PCE as the FHIR element *Condition.code* like

- *Condition.code*: *71620000*:{*363698007=722738000}*
  (Fracture of femur : {Finding site = Structure of bone of left femur})

or by splitting the semantic meaning up into two pre-coordinated SNOMED CT concepts using further element-code-combinations of the FHIR resource *Condition*:

- *Condition.code*: *71620000 |Fracture of femur (disorder)|*
- *Condition.bodySite: 722738000 |Bone structure of left femur (body structure)|*.

This variability of expressing medical facts was put to use in the presented approach, allowing a flexible transformation between terminology- and information model-focused representations and therefore an alternative where post-coordinated expressions need to be replaced. Likewise, this transformation may help to solve some of the difficulties arising due to the TermInfo problem itself. By allowing to flexibly switch between different ways of expression, semantic interoperability is preserved independently of the paradigm used in an EHR's representation. Furthermore, it becomes possible to check the plausibility of recorded medical circumstances if their disjointed parts of information (e.g., scattered across various FHIR elements) can be merged into a single *interpretable* expression.

Despite the broad ambiguity between the scope of semantic or structural standards, to some extent specific recommendations are established, such as "contextual meaning should rather be represented via the information model" [52]. Contrarily, the concepts of the SNOMED CT hierarchy *Situation with explicit context* include such contextual information that go beyond the ordinary scope of a terminology, e. g., suspected diagnoses, procedures not done or family history facts, and therefore put correct logical conclusions at risk (an explanation of the reasons (epistemological versus ontological components of meaning) is beyond the scope of this paper, see [3,13]). Based on the concepts' logical definitions, which are directly related to PCEs, the approach presented in this paper allows to extract the problematic pieces of information and to store them within separate elements of the information model. As a result, a concept like *165008002 |Allergy testing not done (situation)|* could be represented by separating the epistemological aspect "not done" into the suitable FHIR element *Procedure.status*:

- *Procedure.code: 252512005 |Allergy test (procedure)|*
- *Procedure.status: not-done* (according to the required HL7 FHIR ValueSet).

Like this, the integrity of SNOMED CT's hierarchies may be preserved.

Conclusions

The use of post-coordinated expressions largely expands SNOMED CT's ability to record medical circumstances in detail. Despite its undeniable benefits, post-coordination has yet to be established throughout routine data collection. To facilitate moving forward, PCEtoFHIR presents an approach to ensure semantic interoperability between post-coordination-proficient versus PCE-naive systems by employing the internationally proliferating HL7 FHIR standard to achieve an alternative representation. State-of-the-art techniques in description logic and terminology services are brought together for a largely automated web application that decomposes PCEs into their principal components. Several validations, both of the individual processing steps as well as the overall process have demonstrated the correct functionality of the approach. PCEtoFHIR was designed and implemented in a modular way, making it well-suited for future adaptations in the highly dynamic environment of current health informatics. Besides the uncomplicated extension to additional hierarchies of SNOMED CT or FHIR profiles by adding further FHIR StructureMaps, the algorithm can even be adapted to work with other information models, such as openEHR or relational databases. Furthermore, the developed approach can be explored as the foundation for overcoming further difficulties in the realm of semantic and structural standards, like the TermInfo problem. By reversing the direction of processing – from FHIR elements to PCE – meaningful SNOMED CT-based analyses could be enabled.

Author Contributions

Conceptualization, JI and CD; methodology, TO and CD; software, TO; validation, TO, AE, CD; writing—original draft preparation, TO; writing—review and editing, AE, IJ, CD; visualization, TO, CD; supervision, JI and CD; project administration, JI
All authors have read and agreed to the published version of the manuscript.

Acknowledgements

This work is funded by the German Federal Ministry of Education and Research (BMBF) as part of the Medical Informatics Initiative Germany, grant 01ZZ2312A.

Conflicts of Interest

The authors declare no conflict of interest.

Abbreviations

CSIRO	Commonwealth Scientific and Industrial Research Organization
C-CDA	Consolidated Clinical Document Architecture
CDA	Clinical Document Architecture
DAG	Directed Acyclic Graph
ECL	Expression Constraint Language
FHIR	Fast Healthcare Interoperability Resources
HL7	Health Level 7
JSON	JavaScript Object Notation
KBV	Kassenärztliche Vereinigung
LCA	Least Common Ancestor
MII	Medical Informatic Initiative
MRCM	Machine Readable Concept Model
NASHIP	National Association of Statutory Health Insurance Physicians
PCE	Post-coordinated expression
RIM	Reference Information Model
XML	Extensible Markup Language
OWL	Web Ontology Language
WASP	Web application supporting SNOMED CT postcoordination

References

1. Benson T, Grieve G. *Principles of Health Interoperability: FHIR, HL7 and SNOMED CT*. Fourth edition. Springer; 2021.

2. Braunstein ML. *Health Informatics on FHIR: How HL7's API Is Transforming Healthcare*. Springer International Publishing; 2022. doi:10.1007/978-3-030-91563-6

3. Ingenerf J, Drenkhahn C. *REFERENZTERMINOLOGIE SNOMED CT: Interlingua zur Gewährleistung semantischer Interoperabilität in der Medizin*. Springer; 2024.

4. Kassenärztliche Bundesvereinigung. KBV-Basis-Profile. Accessed November 29, 2023. https://simplifier.net/organization/kassenrztlichebundesvereinigungkbv

5. Medizininformatik Initiative. Der Kerndatensatz der Medizininformatik-Initiative. Accessed November 29, 2023. https://www.medizininformatik-initiative.de/de/der-kerndatensatz-der-medizininformatik-initiative

6. Drenkhahn C, Ohlsen T, Wiedekopf J, Ingenerf J. WASP—A Web Application to Support Syntactically and Semantically Correct SNOMED CT Postcoordination. *Applied Sciences*. 2023;13(10):6114. doi:10.3390/app13106114

7. Karlsson D, Nyström M, Cornet R. Does SNOMED CT post-coordination scale? *Stud Health Technol Inform*. 2014;205:1048-1052.

8. Cornet R, Nyström M, Karlsson D. User-directed coordination in SNOMED CT. *Stud Health*

Technol Inform. 2013;192:72-76.

9. Lopetegui M, Mauro A. A Novel Approach to Create a Machine Readable Concept Model for Validating SNOMED CT Concept Post-coordination. *Stud Health Technol Inform*. 2015;216:1087. doi:10.3233/978-1-61499-564-7-1087

10. Wardle M. Health informatics and information technology; Semantic interoperability: SNOMED CT, post-coordination and the model. Accessed November 29, 2023. https://wardle.org/terminology/2018/10/27/snomed-postcoordination-1.html

11. Metke-Jimenez A, Steel J, Hansen D, Lawley M. Ontoserver: a syndicated terminology server. *J Biomed Semant*. 2018;9(1):24. doi:10.1186/s13326-018-0191-z

12. CSIRO. Ontoserver: SNOMED CT Post-Coordination support. Accessed November 29, 2323. https://ontoserver.csiro.au/docs/6/postcoordination.html

13. Schulz S, Schober D, Daniel C. Bridging the semantics gap between terminologies, ontologies, and information models. *Studies in health technology and informatics*. 2010;160(2):1000-1004. doi:10.3233/978-1-60750-588-4-1000

14. Cheetham E, Dolin RH, Markwell D. Using SNOMED CT in HL7 Version 3; Implementation Guide, Release 1.5. Accessed April 22, 2024. https://www.hl7.org/v3ballotarchive/v3ballot/html/infrastructure/terminfo/terminfo.html

15. HL7 International. TermInfo Project. Accessed April 22, 2024. https://www.hl7.org/Special/committees/terminfo/overview.cfm

16. Ohlsen T, Kruse V, Krupar R, Banach A, Ingenerf J, Drenkhahn C. Mapping of ICD-O Tuples to OncoTree Codes Using SNOMED CT Post-Coordination. *Studies in Health Technology and Informatics*. 2022;294:307-311. doi:10.3233/SHTI220464

17. Vogl K, Ingenerf J, Kramer J, Chantraine C, Drenkhahn C. LUMA: A Mapping Assistant for Standardizing the Units of LOINC-Coded Laboratory Tests. *Applied Sciences*. 2022;12(12):5848. doi:10.3390/app12125848

18. Rinaldi E, Drenkhahn C, Gebel B, et al. Towards interoperability in infection control: a standard data model for microbiology. *Sci Data*. 2023;10(1):654. doi:10.1038/s41597-023-02560-x

19. Green JM, Wilcke JR, Abbott J, Rees LP. Development and Evaluation of Methods for Structured Recording of Heart Murmur Findings Using SNOMED-CT(R) Post-Coordination. *Journal of the American Medical Informatics Association*. 2006;13(3):321-333. doi:10.1197/jamia.M1973

20. Miñarro-Giménez JA, Martínez-Costa C, López-García P, Schulz S. Building SNOMED CT Post-Coordinated Expressions from Annotation Groups. *Stud Health Technol Inform*. 2017;235:446-450.

21. Castell-Díaz J, Miñarro-Giménez JA, Martínez-Costa C. Supporting SNOMED CT postcoordination with knowledge graph embeddings. *Journal of Biomedical Informatics*. 2023;139:104297. doi:10.1016/j.jbi.2023.104297

22. HL7 International. Resource Condition - Mappings. Accessed November 23, 2023.

https://hl7.org/fhir/R4/condition-mappings.html

23. HL7 International. Resource Observation - Mappings. Accessed November 23, 2023. https://hl7.org/fhir/R4/observation-mappings.html

24. SNOMED International. SNOMED on FHIR; Bindings to FHIR Clinical Resources. Accessed April 22, 2024. https://confluence.ihtsdotools.org/display/FHIR/Bindings+to+FHIR+Clinical+Resources

25. Perez-Rey D, Alonso-Calvo R, Paraiso-Medina S, Munteanu CR, Garcia-Remesal M. SNOMED2HL7: A tool to normalize and bind SNOMED CT concepts to the HL7 Reference Information Model. *Comput Methods Programs Biomed*. 2017;149:1-9. doi:10.1016/j.cmpb.2017.06.020

26. Arguello Casteleiro M, Martinez-Costa C, Des-Diu J, Maroto N, Fernandez-Prieto MJ. From SNOMED CT expressions to an FHIR RDF representation: Exploring the benefits of an ontology-based approach. *ODLS 2019: Ontologies and Data in Life Sciences 2019*. Published online July 11, 2019:1-13.

27. Hitzler P, Krötzsch M, Rudolph S, Sure Y. *Semantic Web*. Springer Berlin Heidelberg; 2008. doi:10.1007/978-3-540-33994-6

28. Staab S, Studer R, eds. *Handbook on Ontologies*. Springer Berlin Heidelberg; 2009. doi:10.1007/978-3-540-92673-3

29. International Health Terminology Standards Development Organisation. SNOMED CT OWL Guide. Published online January 28, 2022. Accessed November 29, 2023. https://confluence.ihtsdotools.org/display/DOCOWL/SNOMED+CT+OWL+Guide?preview=/64265998/142119608/SNOMED%20CT%20OWL%20Guide-v19-20220128.pdf

30. International Health Terminology Standards Development Organisation. Snomed OWL Toolkit. Published online February 2, 2016. https://github.com/IHTSDO/snomed-owl-toolkit

31. International Health Terminology Standards Development Organisation. SNOMED CT International Edition. Accessed January 4, 2024. https://www.nlm.nih.gov/healthit/snomedct/international.html

32. Horridge M, Bechhofer S. The OWL API: A Java API for OWL ontologies. *Semantic Web*. 2011;2(1):11-21. doi:10.3233/SW-2011-0025

33. Kazakov Y, Krötzsch M, Simancik F. ELK Reasoner: Architecture and Evaluation. *Proceedings of the 1st International Workshop on OWL Reasoner Evaluation (ORE-2012) CEUR Workshop Proceedings*. Published online July 2012.

34. Sánchez D, Batet M. Semantic similarity estimation in the biomedical domain: An ontology-based information-theoretic perspective. *Journal of Biomedical Informatics*. 2011;44(5):749-759. doi:10.1016/j.jbi.2011.03.013

35. Aric Hagberg, Pieter Swart, Daniel S Chult. Exploring Network Structure, Dynamics, and Function Using NetworkX. 2008;Proceedings of the 7th Python in Science Conference (SciPy2008):11-15.

36. International Health Terminology Standards Development Organisation. Template Syntax DRAFT Specification. 2020;(1.1.1). Accessed November 29, 2023. https://confluence.ihtsdotools.org/display/DOCSTS?preview=/45529301/115875508/doc_TemplateSyntax_v1.1.1-en-US_INT_20201020.pdf

37. HL7 International. Resource StructureMap - Content. Accessed November 23, 2023. https://hl7.org/fhir/R4/structuremap.html

38. University Health Network. HAPI FHIR - The Open Source FHIR API for Java. Accessed May 3, 2024. http://hapifhir.io/

39. Smile Digital Health. HAPI FHIR. Accessed January 4, 2024. https://hapifhir.io/hapi-fhir

40. Rada R, Mili H, Bicknell E, Blettner M. Development and application of a metric on semantic nets. *IEEE Trans Syst, Man, Cybern*. 1989;19(1):17-30. doi:10.1109/21.24528

41. Claudia Leacock, Martin Chodorow. Combining Local Context and WordNet Similarity for Word Sense Identification. *WordNet: an electronic lexical database MIT Press*. Published online January1998.

42. Yuhua Li, Bandar ZA, McLean D. An approach for measuring semantic similarity between words using multiple information sources. *IEEE Trans Knowl Data Eng*. 2003;15(4):871-882. doi:10.1109/TKDE.2003.1209005

43. Choi I, Kim M. Topic distillation using hierarchy concept tree. In: *Proceedings of the 26th Annual International ACM SIGIR Conference on Research and Development in Informaion Retrieval*. ACM; 2003:371-372. doi:10.1145/860435.860506

44. Kate RJ. Automatic full conversion of clinical terms into SNOMED CT concepts. *Journal of Biomedical Informatics*. 2020;111:103585. doi:10.1016/j.jbi.2020.103585

45. Kassenärztliche Bundesvereinigung. KBV-Basis-Allergie/Unverträglichkeit, 1.5.1.2 BIS. Published July 8, 2024. https://mio.kbv.de/display/BASE1X0/1.5.1.2+bis

46. HL7 International. StructureDefinition-condition-diseaseCourse. Published July 8, 2024. https://fhir.org/guides/stats2/structuredefinition-hl7.fhir.uv.extensions.r4-condition-diseasecourse.html

47. HL7 International. StructureDefinition-condition-dueTo. Published July 8, 2024. https://fhir.org/guides/stats2/structuredefinition-hl7.fhir.uv.extensions.r4-condition-dueto.html

48. Rohit J. Kate. Clinical Terms to SNOMED CT Concepts. Accessed May 13, 2024. https://sites.uwm.edu/katerj/conversion/

49. Lau, Tobias. Gesundheitsdaten: FHIR wird europaweiter Standard. *Ärzteblatt*. Accessed January 4, 2024. https://www.aerzteblatt.de/nachrichten/142159/Gesundheitsdaten-FHIR-wird-europaweiter-Standard

50. Vorisek CN, Lehne M, Klopfenstein SAI, et al. Fast Healthcare Interoperability Resources (FHIR) for Interoperability in Health Research: Systematic Review. *JMIR Med Inform*. 2022;10(7):e35724. doi:10.2196/35724

51. Ayaz M, Pasha MF, Alzahrani MY, Budiarto R, Stiawan D. The Fast Health Interoperability Resources (FHIR) Standard: Systematic Literature Review of Implementations, Applications, Challenges and Opportunities. *JMIR Med Inform*. 2021;9(7):e21929. doi:10.2196/21929

52. Schulz S. Wozu benötigen wir standardisierte Terminologien wie SNOMED CT? *Swiss Medical Informatics*. Published online January 2011:27-32.

Preprint JMIR Publications

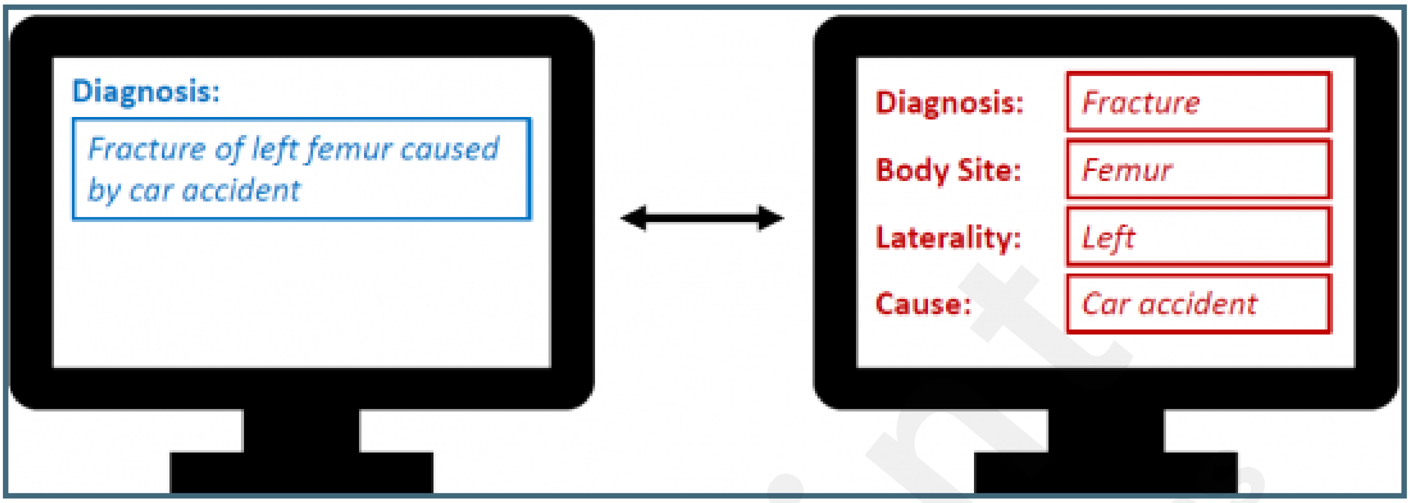
Supplementary Files

Preprint JMIR Publications

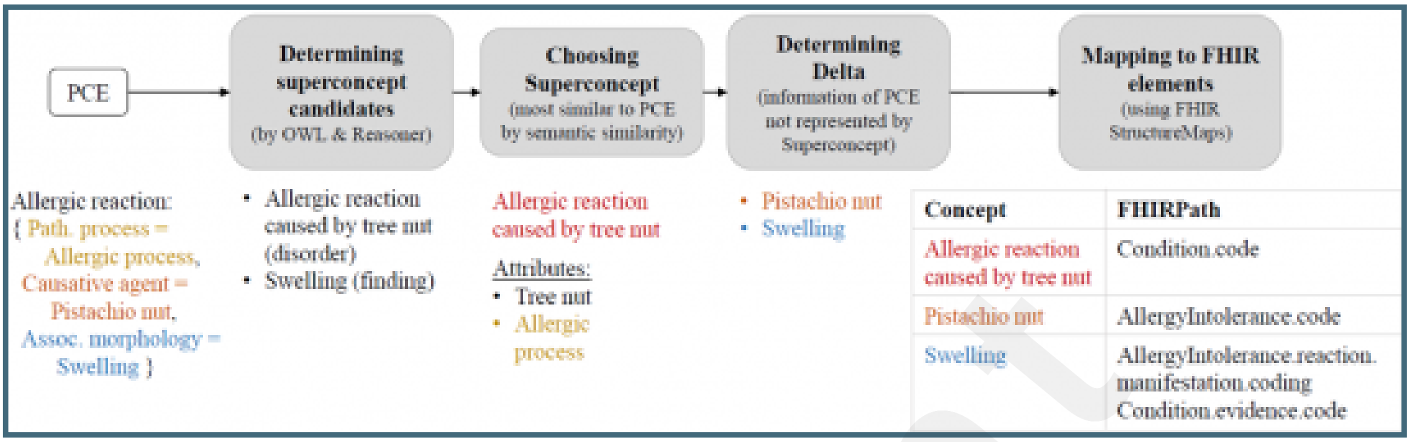
Figures

Preprint
JMIR Publications

TermInfo: The same medical fact can be represented employing either the terminology (left side) or the information model (right side) more heavily.



The decomposition of a PCE into elements of (profiled) FHIR resources consists of four steps.

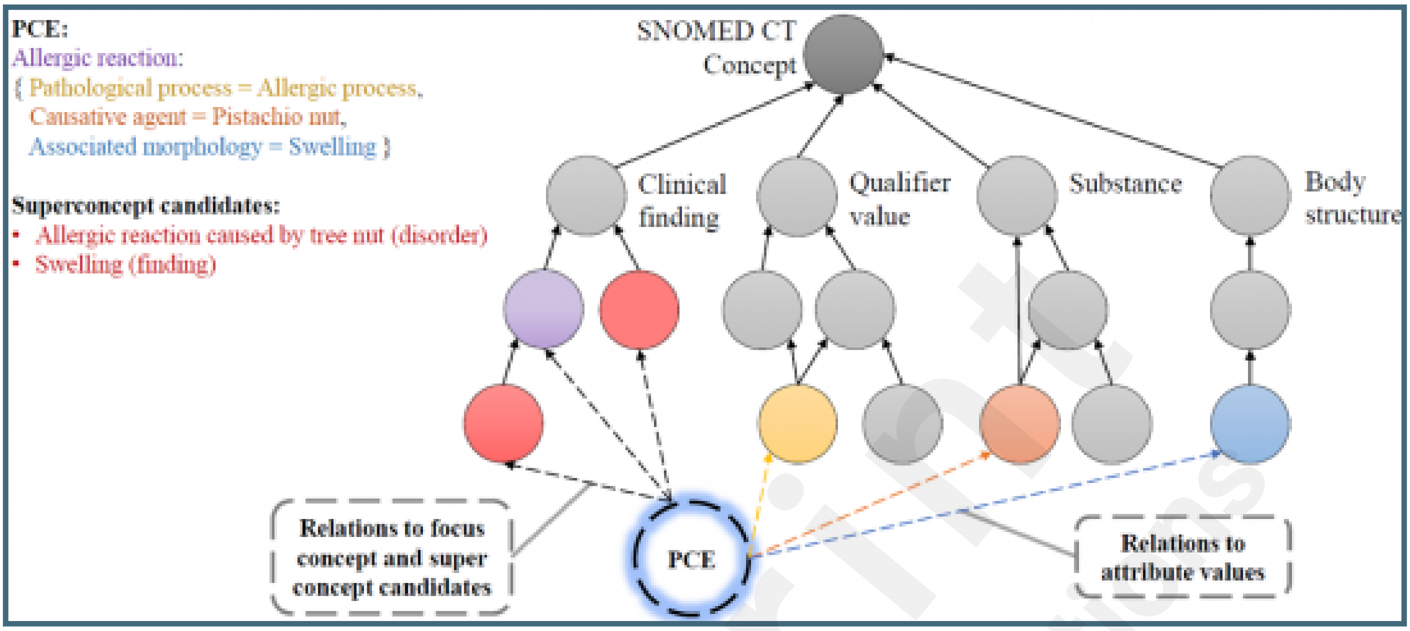


PCE on the left side and the associated OWL ClassExpression based on functional syntax on the right side.

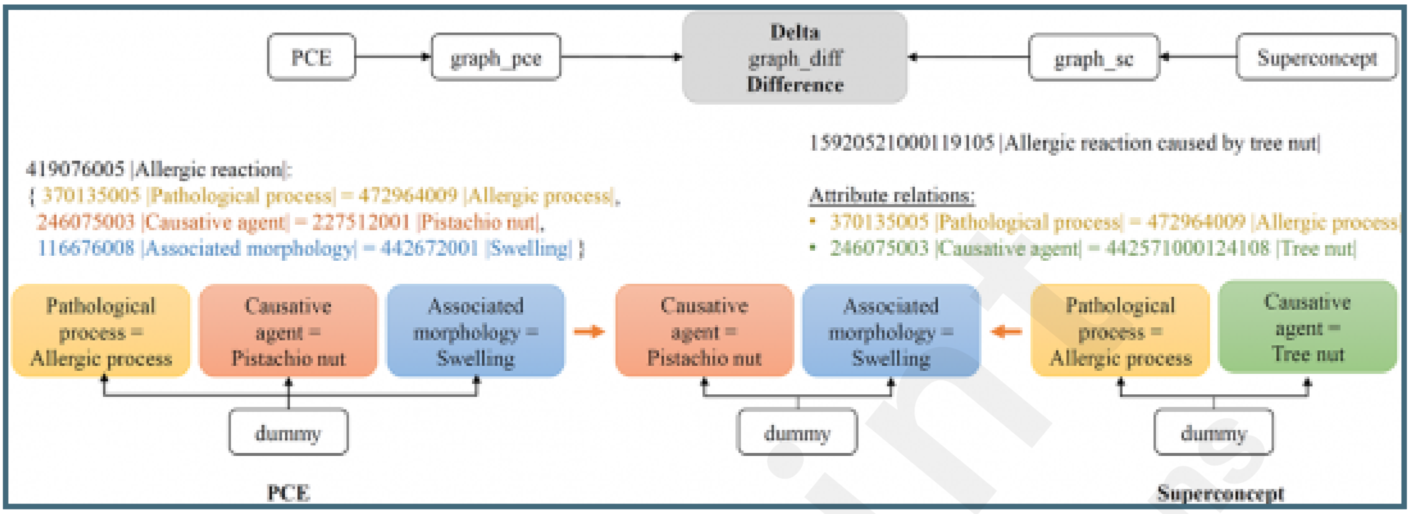
<table>
<tr><th>PCE</th><th>Associated OWL Class Expression</th></tr>
<tr><td><pre>
419076005 |Allergic reaction| :
{ 370135005 |Pathological process| =
    472964009 |Allergic process| ,
  246075003 |Causative agent| =
    227512001 |Pistachio nut| ,
  116676008 |Associated morphology| =
    442672001 |Swelling|
}
</pre></td><td><pre>
ObjectIntersectionOf(
  : 419076005 |Allergic reaction|
  ObjectSomeValuesFrom(
    : 609096000 |Role group|
    ObjectIntersectionOf(
      ObjectSomeValuesFrom(
        : 370135005 |Pathological process|
        : 472964009 |Allergic process|
)
      ObjectSomeValuesFrom(
        : 246075003 |Causative agent|
        : 44257100012418 |Tree nut|
)
      ObjectSomeValuesFrom(
        : 116676008 |Associated morphology|
        : 442672001 |Swelling|
)
)
)
)
</pre></td></tr>
<tr><td colspan="2">Output (direct superconcepts):
– 65124004 |Swelling (finding)|
– 15920521000119105 |Allergic reaction caused by tree nut (disorder)|</td></tr>
</table>

preprint
JMIR Publications

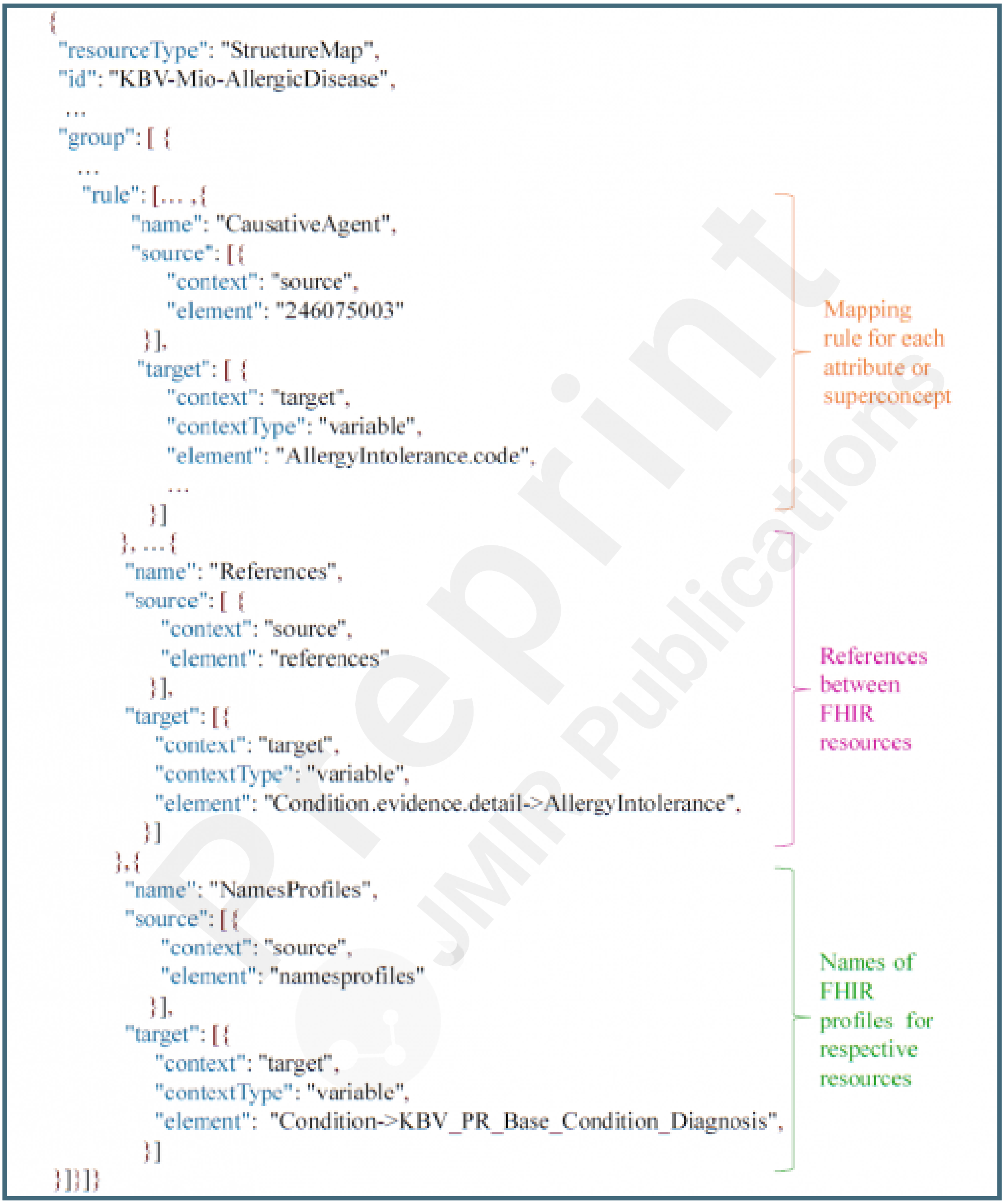
The PCE is represented as a sub-concept of its focus concept and the super-concept candidates. It also has edges to its attribute values.



When calculating the delta, the graph of the superconcept is subtracted from that of the PCE. The determined edges contain attribute relations that were not covered by the superconcept but are necessary for the precise representation of the PCE.

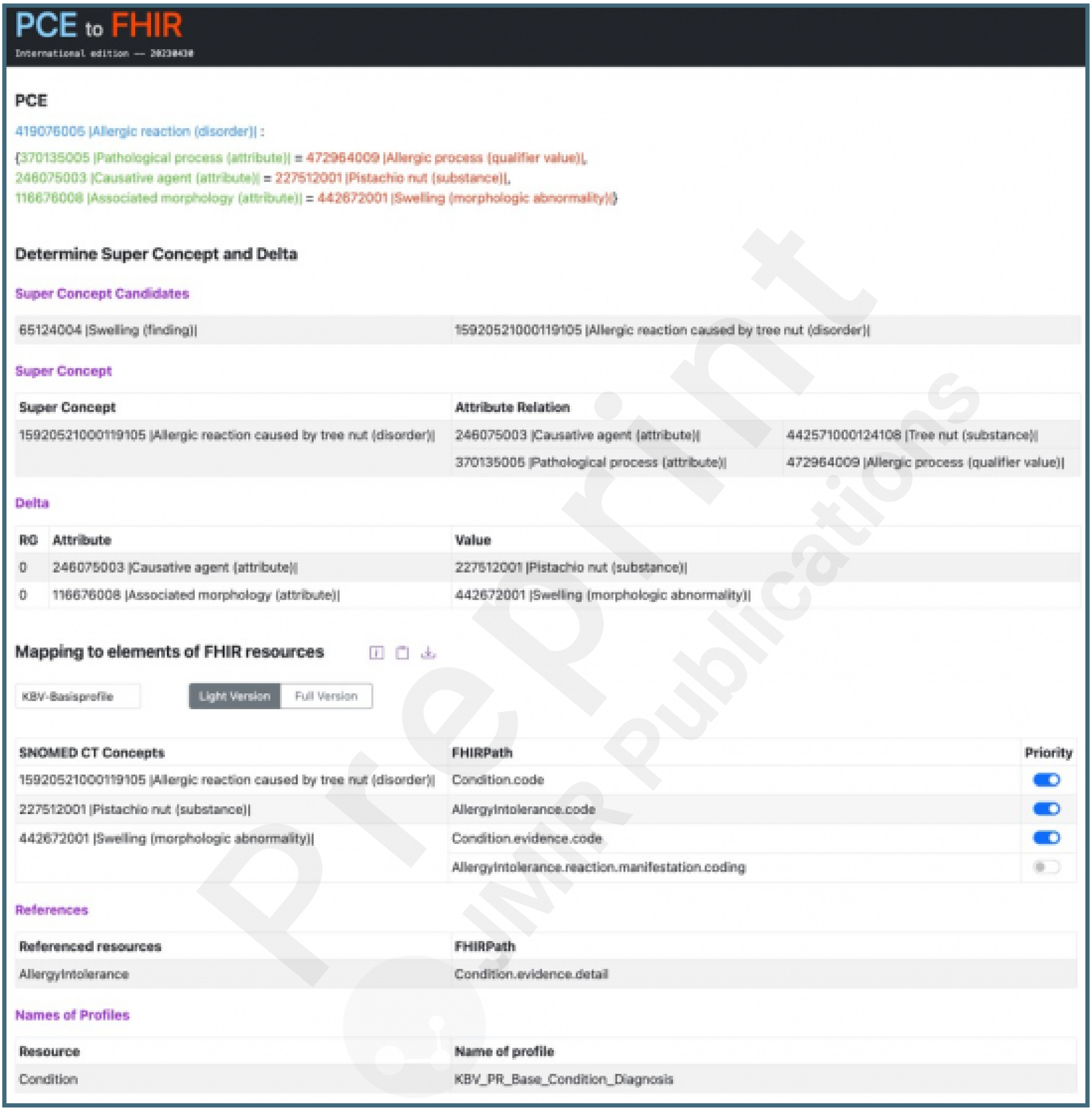


Extract of the FHIR StructureMap for allergic reactions in JSON format. It contains an entry for each SNOMED CT attribute and the superconcept as well as references and names of the profiles.



Preprint JMIR Publications

Excerpt of the web application PCEtoFHIR.



Validation process: The original PCE is decomposed by PCEtoFHIR as usual, and accordingly stored in FHIR resources. Based on this representation, the singular concepts are recomposed into a second PCE. The original and the recomposed PCE are compared.