Aligning an administrative procedure coding system with SNOMED CT

Stefan SCHULZ[a,c,1], Johannes STEFFEL[a], Peter POLSTER[a],
Matvey PALCHUK[b], Philipp DAUMKE[a]
[a] *Averbis GmbH, Freiburg, Germany*
[b] *TriNetX, Cambridge, MA, USA*
[c] *Institute for Medical Informatics, Statistics and Documentation, Medical University of Graz, Austria*

Abstract. OPS, the German coding system for therapeutic and diagnostic procedures, is a large and complex classification system. Its main purpose is to provide codes for billing. Like other systems of this type (e.g. ICD-10) it follows the principle of class disjointness and exhaustiveness. SNOMED CT, on the other hand, aims at providing standardised terms, together with logic-based descriptions, and pursues the goal to make the electronic health record (EHR) computable and interoperable across languages and jurisdictions. We investigated the feasibility of aligning OPS with SNOMED CT, based on the 1000 most frequently used OPS codes. A team of three terminologists performed the mapping (partially overlapping), using the first hundred codes to determine guidelines. From the work, which is currently being extended, we can draw the following conclusions: (i) for less than half of the OPS codes, a semantically equivalent SNOMED CT code can be found; (ii) many maps require SNOMED CT post-coordination but remain approximate; (iii) the mapping work is impaired by imprecise descriptions in either terminology system.

Keywords. Medical Procedures, Medical Classifications, SNOMED CT

1. Introduction

Most artefacts that provide a semantic reference – generically referred to as terminology systems [1, 2] – for the organisation of biomedical data are restricted to a well-defined scope regarding the types of referents these data denote. Examples are drug terminologies, which offer codes and definitions for pharmaceutical products and chemicals (e.g., ATC, RxNorm, ChEBI), terminologies for everything that can be observed and measured (LOINC), bodily conditions like disorders and injuries (ICD-10), cell components, molecular functions and biological processes (Gene Ontology), just to name some of the most important ones. There is an increasing momentum towards international standardisation and cross-border use. It has a long tradition in the case of ICD for health statistics, promoted by the WHO in 42 language versions, and is more recent in the case of ongoing standardisation efforts for drug products (IDMP),

[1] Corresponding Author: Stefan Schulz, Averbis GmbH, Salzstraße 15, 79098 Freiburg im Breisgau; E-mail: stefan.schulz@averbis.com

responding to a worldwide demand for internationally harmonized medicinal product specifications. LOINC is being translated into more and more languages. Regarding terminology support for biomedical research, resources like the Gene Ontology, ChEBI and other bio-ontologies have been created from the very beginning for semantic annotations of bio-molecular data across international research communities.

There is one remarkable exception to this trend, viz. terminologies for medical procedures. Procedure terminologies encompass operations, drug and other therapies, as well as imaging and other diagnostic procedures. Here, a multitude of national terminologies coexists. Most of them are better described as catalogues than as terminologies or ontologies, i.e., flat lists of medical interventions, often completed by weight values from which the price of each procedure can be computed. Rather than terminologies or ontologies (which systematise terms and/or their referents), they resemble classification systems (like ICD) with mutually disjoint classes, supported by rules that clarify under which conditions a given procedure belongs to a given class.

Billing is therefore a major driver for procedure coding systems, and the heterogeneity of national health systems (enhanced by different systems co-existing within jurisdictions) explains the special status of these artefacts. While there are good examples that semantic harmonisation of data across national borders is possible when the data is captured by standards that enjoy global adoption, semantic harmonization of medical procedures remains a problem [3], as they stand out as a large data domain that does not have a globally adopted standard.

Medical procedures also constitute an important portion of SNOMED CT (58,213 concepts) [4], the "common global language for health terms", an international standard that enjoys increasing acceptance around the globe. Best described as an ontology-based terminology, SNOMED CT aims at providing standardised terms, together with logic-based descriptions, and pursues the goal to make electronic health records (EHR) computable and interoperable across languages and jurisdictions. The focus of the SNOMED procedure hierarchy (about one sixth of all SNOMED CT concepts) is to provide fine-grained, standardised descriptions of clinical procedures.

Therefore, SNOMED CT is often seen as a strong candidate to represent medical procedures, but capturing procedures with SNOMED CT has never been current practice. An important step towards having procedures coded in SNOMED CT, could consist in the construction of mappings to SNOMED CT from the local procedure coding systems, as there does not seem to be any other suitable standard. ICD-10-PCS, developed for use in the U.S. and adopted in few additional countries, is typically considered to be very difficult to map, ICHI (International Classification of Health Interventions [5]), currently under development by the WHO is far too coarse-grained, which is also true for the Procedures section of the Medical Subject Headings (MeSH) primarily targeting biomedical literature indexing.

2. Background

There are ongoing efforts in some countries to harmonize procedure standards. In the US, the National Library of Medicine is working on a way to map ICD-10-PCS to SNOMED. Instead of a map [6] they chose to focus on a tool that requires human intervention to generate an equivalence mapping from an ICD-10-PCS procedure to SNOMED. In the UK, the OPCS standard has been mapped to SNOMED CT by the NHS. However, efforts requiring the mapping of very large (tens of thousands of codes) and very complex terminologies such as the ones used for procedures (as well as SNOMED itself) are being undertaken by governmental agencies more so than private companies or collaborations. Governments tackle these projects because the effort required is high and because supporting and promoting interoperability is being understood as a global need.

Compared to other biomedical terminology systems, SNOMED CT is unique not only due to its scope and size, but also regarding its ontological foundation, based on the description logics [7] profile OWL-EL [8]. This allows for logically defining its representational units (SNOMED CT concepts), e.g. in

Tonsillectomy equivalent to *Procedure* and
has-part some ((**method** some *Excision - action*) and
(‘**procedure site direct**’ some ‘*Tonsillar structure* (palatine)’))

This is the case of a so-called pre-coordinated concept (identified by the code 173422009 and the label ‘Tonsillectomy (procedure)’). The syntax also allows for constructing more detailed expressions for which no code is available, e.g. tonsillectomy with an ultrasonic scalpel.

Procedure and **has-part** some
((**method** some *Excision - action*) and
(‘**procedure site direct**’ some ‘*Tonsillar structure* (palatine)’) and
(‘**using device**’ some ‘*Ultrasonic scalpel*’))

This mechanism, called post-coordination, allows for an increased level of coverage and detail, however, at the price of increased complexity, which means that documentation systems have to deal with logical expressions instead of just codes.

A typical classification system also provides a code for tonsillectomy, but often comes with additional instructions, such as using different codes for tonsillectomy with or without adenoidectomy, or demanding additional codes for, e.g. haemostasis after tonsillectomy, or requiring different codes for different age groups. Table 1 provides a juxtaposition of a procedure ontology with a procedure classification (see also [9]).

Table 1. Procedure ontologies compared to procedure classifications

	Procedure ontology	**Procedure classification**
Semantics	Open world, classes (extension of concepts) often overlap	Closed world (disjoint classes)
Structure	Multiple hierarchy, intensional	Single hierarchy, extensional
Constructors	Subclass, Equivalence, conjunction (“and”), existential quantification (“some”)	Subclass, excludes
Ontological commitment	Classes of medical procedures, i.e. (parts of) actions performed on a patient by a health professional	Purpose-oriented standardised information objects related to medical procedures (see left)
Purpose	Provision of representational units within a formal account of the electronic health record	Provision of correlates to medical procedures and their parts, from which the monetary value of a procedure can be derived

Several scenarios of use justify an alignment between classification-type coding systems and SNOMED CT:

1. Primary effort put into manual administrative encoding using a procedure classification or catalogue. Then, the alignment resource can be used to infer SNOMED codes for adding semantic annotations to the EHR, supporting a broad range of primary or secondary use scenarios (decision support, prediction, cohort building, health statistics), capitalizing on the ontological structure of SNOMED CT.
2. Primary effort put into manually annotating EHR content with SNOMED CT codes. However, inferring the full meaning of a procedure classification code would then require representing also the disjointness conditions and exclusions. For several reasons this exceeds the power of SNOMED CT post-coordination. As an alternative, each procedure code could be expressed as a query on a SNOMED-CT annotated record.
3. The same as 2, but using natural language processing (NLP) for annotating clinical narratives with SNOMED CT codes. This scenario, as well as the previous one, would realistically require additional human encoding efforts, given the high quality required for codes that are used for billing.

Our work described in this paper pursues the first of these three goals, i.e. the direction from the classification to the ontology. Out of a ranked list of approximately 24,000 codes from the German procedure classification system OPS [10] (Versions 2004 - 2019) it takes the most frequent 1000 and attempts to map them to SNOMED codes or post-coordinated expressions. The main criterion underlying this effort is the following: given an OPS procedure code p_i attached to an EHR: for which SNOMED CT concepts c_{i1}… c_{in} (or OWL class-like post-coordinated expressions), instance(s) can be assumed to exist in the health care episode described by that EHR?

3. Materials and Methods

3.1. OPS and SNOMED CT

OPS, used in Germany for encoding therapeutic and diagnostic procedures, is a fine-grained classification system with 35,641 codes distributed across seven hierarchical levels. Its purpose is to provide codes for billing. Like other systems of this type (e.g. ICD-10) it follows the principle of class disjointness and exhaustiveness. The "Systematic Version" PDF file (2019 release) was used as a reference. In this version, formatted like a book, certain naming principles had to be considered. E.g., the code 5-790.26: "Geschlossene Reposition einer Fraktur oder Epiphysenlösung des distalen Radius mit Osteosynthese unter Verwendung eines intramedullären Drahts" [Closed reduction of a fracture or slipped epiphysis of the distal radius by internal fixation using an intramedullary wire]. This label is not pre-synthesised; it has to be constructed by

5-790	*Closed reduction of a fracture or slipped epiphysis by internal fixation*
****5-790.2**	*By intramedullary wire*
6 ↔	*Distal radius*

The sixth digit is taken from a list with anatomical sites that can be combined with several 5-character codes. However, completely pre-synthesised texts are also available. Nevertheless is the inspection of the Systematic Version indispensable in order to get access to exclusions, inclusions, and scope notes at all hierarchical levels. E.g., the subchapter **5-79** *Reduction of Fracture and Dislocation* is preceded by nearly one page of such additional information (e.g. excludes therapy of pseudarthrosis, requires separate encoding of nerve sutures). In addition, under the heading the subsubchapter **5-790**, more additional information is given, e.g. that child fractures are included, closed reductions dislocations of joints are excluded, or that arthroscopic assistance requires an additional code. The level of detail is often only fully understandable by specialist surgeons. Expertise and a certain degree of subjective interpretation are also needed in order to understand the meaning of certain terms, which lack precise definition, such as "Epiphyseolyse".

The SNOMED CT *Procedures* hierarchy provides formal concept definitions, which can be interpreted as description logics axioms (cf. Background). In contrast, scope notes or text definitions are completely missing, which is particularly challenging where formal definitions refer to undefined primitives, e.g. from the SNOMED CT *Qualifier Value* hierarchy, e.g. *Preperitoneal approach* without connection to any anatomy reference. Another peculiarity is what SNOMED CT calls *role grouping*. This can most straightforwardly be interpreted as asserting a mereological order between procedures and their processual parts. However, these role groups also occur solely in, e.g. in the concept *Sigmoidoscopy*, which is therefore classified as a taxonomic parent of, e.g. *Sigmoidoscopy with biopsy*. Therefore, the precise meaning of *Sigmoidoscopy* would then be "Procedure with sigmoidoscopy". The background of this is to optimise term retrieval: searching for "sigmoidoscopy" would then retrieve also data annotated with the concept *Sigmoidoscopy with biopsy.* Negation cannot be expressed by the SNOMED CT syntax. However, we find "reified" negations in several concepts, such as *Computed tomography of head without contrast*.

These examples demonstrate, in addition to the basic distinctions exposed in Tab. 1, the wide discrepancy of SNOMED CT and OPS in particular (which can also be extended to classification-like coding systems in general). Whatsoever alignment between OPS and SNOMED CT has to be aware of this. Simple lexical mapping is not sufficient. It would lead to numerous wrong equivalence statements: an OPS code, with its meaning restricted by numerous exclusion rules and with the underlying close-value assumption is rarely fully semantically equivalent to any SNOMED CT concept.

3.2. Dataset for OPS code ranking

The dataset was provided by TriNetX. They harvested it from two German hospitals, where OPS is established as the official coding system (together with ICD-10) for the German DRG (Diagnosis-related groups) payment system [12]. The datasets consists of 6,892,330 single codes out of 23,985 different OPS code types. This corresponds to a coverage of 67.3%, i.e. about one third of OPS codes were never used. For our mapping project we selected the 1,000 most frequent codes, which still correspond to 5,580,702 code assignments, i.e. 80.9%.

3.3. Coders, pilot SNOMED mappings and mapping schema

Two coders were recruited (2nd and 3rd authors), both of them final year medical students, one of which already held a degree in nursing. They were trained supervised by the first author, MD and experienced terminologist / ontologist. Each coder was hired for 26 hours a month over three months. The supervisor had the same time budget. This period covered the whole cycle from guideline creation, training, mapping, validation to the delivery of the map and the final report (this paper).

The 1000-code set OPS_{1000} was ordered by random. The first author, together with the coders analysed the first 100 codes regarding their alignment with SNOMED CT, i.e. the mapping of one OPS code to one or more SNOMED CT codes. The main purpose of this initial step was to reach a consensus regarding a meaningful, simple and reusable mapping scheme. First, the underlying assumption of the mapping process was formulated: According to assumption 3 in the Background chapter, we defined the mapping task as follows: Given a patient record annotated with the code OPS_i, which SNOMED CT expression(s) $SCT_{1..j}$ can be reliably assumed to be instantiated. The range of the mapping should contain one or more SNOMED CT codes or SNOMED CT post-coordinated expressions belonging to one of the semantic types "Procedure", "Regime/Therapy" or "Situation". In addition, a scoring system is used to distinguish either the quality of the mapping or “no mapping”.

For each map, the OPS code is analysed in its hierarchical context, taken into consideration scope notes, inclusion and exclusion statements. Elements of the OPS label are translated into English, eliminating doubts regarding the appropriate translation in online sources whenever necessary. In case of doubt, a search with more general terms is done. Words or word stems are entered into the SNOMED browser [13]. In order to find the best matching term, also sibling, super and subconcepts are inspected.

As a rule of thumb, maps are preferred that were as close as possible to the original wording. In case no map is achieved, a compositional approach is pursued to approximate the meaning. Full post-coordination using the SNOMED CT compositional grammar is not aimed at, due to its complexity, its experimental status (especially regarding its closeness to description logics) and its irrelevance for current implementations. Post-coordination is therefore restricted to logical conjunction (AND), disjunction (OR) and addition (ADD). The latter is preferred in case the OPS code stands for clearly distinct entities, of which a conjunction (even given the large tolerance how SNOMED CT handles logical conjunctions) is considered inappropriate. E.g., if there is no SNOMED CT concept for an OPS code ‘Procedure P on body site B’, then the post-coordination *P* AND ‘*Procedure on B*’ would be an appropriate representation, because it is still one procedure. In contrast, if an OPS code stands for ‘Procedure P followed by Procedure Q’, the preferred SNOMED map would then be *P* ADD *Q*, which means that it is represented by actually two separate procedures.

OPS codes are often defined by numerical values or value ranges such as number of therapy units, dosage, frequency, or implicitly by age groups (e.g. adults, children). SNOMED CT procedure concepts never include such criteria, i.e. an exact mapping cannot be expected in these cases. Table 2 gives an overview of the mapping scores we elaborated.

Table 2. Scoring of OPS – SNOMED CT mappings

Score	Meaning regarding source OPS code (S) and target code or expression (T)
Exact	T holds for the same (individual) procedures as S
Exact-Q	T holds for the same (individual) procedures as S, when quantitative restrictions on S are neglected
Broader	The individual procedures denoted by S are a (still significant) subset of those denoted by T
No mapping	There is no code or expression T that allows any of the above judgement.

For practical purposes, there is still the preliminary category “revisit”, which is set in case a coder is not sure about the decision and wants to mark the code for a group discussion. Each OPS code is seen three times, with the following roles: C: coder - the person who does the OPS-SNOMED mapping; R1: first reviser - the person who checks the decision taken by C, R2: second (senior) reviser: the person (mostly the supervisor) who takes the final mapping decision, sometimes as a result of a group discussion. Comment fields are available for each of the three experts; whenever a map is changed, this is documented by an entry. Each expert uses a different colour for his comments.
The pilot phase also yielded the following exclusion recommendations:

1. OPS Codes containing the administration of medicines for which no SNOMED CT procedure codes exists
2. Codes that contain extremely detailed descriptions of a "complex therapy"
3. Planning phase A of a procedure B if only SNOMED codes for B are present
4. Procedures in cases of doubt
5. Supplementary OPS codes (“Zusatzcodes"), unless containing significant information of the type procedure
6. Retired OPS codes

The exclusion criteria were reassessed at the end of the mapping phase.

3.4. Mapping process

As a collaborative environment, a Google spreadsheet was created and filled by the coders with OPS codes, texts, logical operators, and comments. Table 3 provides the stepwise approach on a randomised list of OPS codes, identified as OPS_1 - OPS_{1000}.

Table 3. Steps for mapping the most frequent OPS codes to SNOMED CT

OPS codes	Step
OPS_1 - OPS_{100}	Collaborative, explorative (C, R1, R2) . Consolidation of the mapping scheme and mapping guidelines.
OPS_{101} - OPS_{300}	Students play the role of C and R1 for half of the codes Thereafter group discussion including R2, adjudication of controversial decisions
OPS_{301} - OPS_{400}	Performed in separate spreadsheet (and without communication) for first reliability testing. Both C and R1 play the C role. Thereafter, calculation of inter-coder agreement, then adjudication between students and with R2 for controversial cases. Adding dataset to main table
OPS_{401} - OPS_{800}	Like in step 2, students play the role of C and R1 for half of the codes Thereafter group discussion including R2, adjudication of controversial decisions
OPS_1 - OPS_{800}	Reordering of list by order of codes. Comparison of similar codes and related mapping decisions by all C, R1, R2. Chat and phone discussions in case of inconsistent mappings of similar codes. Revising and completing R2 decisions
OPS_{801} - OPS_{1000}	Second reliability testing. Adjudication between students and with R2 for controversial cases. Reassessment of the exclusions. Mapping of the re-included codes. Decision for all codes marked as “revisit”

Coders were also asked to skip the mapping of a code whenever this takes more than ten minutes. These codes were tagged as “revisit”. The revisiting of these codes is scheduled to take place once all other codes are consolidated.

3.5. Quality assessment of mappings

Inter-coder agreement was measured at two points (Tab. 3): for the codes OPS_{301} - OPS_{400}, in order to achieve a preliminary estimation and at the end, for the codes OPS_{801} - OPS_{1000}.

3.6. Prototypical cases of mapping issues

During the whole process, cases of difficult or controversial mappings were picked out and discussed. Priority was given to those mapping problems that can be seen as prototypical issues not only with regard to OPS, but also to classification-like coding systems in general.

3.7. Final workup of top 1000 map

For a final quality check, the OPS codes were re-arranged from a random order to the numeric order of OPS. This revealed many inconsistencies regarding the mapping of similar codes. The mapping guidelines were adjusted in the sense that also supplementary codes were mapped (as long as this yielded significant clinical meaning). In addition, retired codes were mapped. As a matter of principle, codes that only consisted of administration of drugs were not mapped, assuming that in EHRs there are other, more complete and reliable sources of medication information.

4. Results

4.1. Metrics

The complete time spent amounted to 3 months x 3 experts x 26 hours per expert and month, totalling 234 hours. This corresponds to an average effort of approx. 4.3 OPS codes per hour (14 minutes per code). The mapping of the first 100 codes, including guideline development and documentation required approx. one fourth of the total time. A descriptive analysis of the mappings of the 1000 OPS codes is provided by Table 4. Each code was seen which seen by the three experts and revisited in their original order by at least one expert.

Table 4. Descriptive analysis of mappings

Cardinality of map (SNOMED CT codes per OPS codes)				
0	1	2	3	4
48	617	282	42	11
Quality of mapping				
Broader	Exact	Exact-Q	No mapping	Revisit
610	310	32	48	0
Type of logical combination				
None	AND	ADD	OR	Complex
665	178	79	56	22

The results of inter-coder agreement are provided in Table 5. We compute simple percentage agreements, because agreement by chance is negligible.

Table 5. Inter-coder agreement in percent (100 mappings evaluated: 301-400; 200 mappings: 801-1000)

Type of agreement	**Agreement [95% CI]**	
	OPS_{301} - OPS_{400}	**OPS_{801} - OPS_{1000}**
Coders agree on at least one core SNOMED CT concept per OPS code	68% [58%; 76%]	65% [58%; 71%]
Coders agree on the same set of SNOMED CT concepts per OPS code	54% [44%; 63%]	46% [38%; 52%]
Coders agree on the same set of SNOMED CT concepts per OPS code and agree regarding the mapping quality	41% [31%; 50%]	36% [30%; 43%]

4.2. Typical cases

In Table 6, typical mapping phenomena are presented. All examples are instances of frequently recurring phenomena. The OPS labels, which are only available in German, were translated to English for better understanding.

Table 6. Instances of recurring mapping phenomena. Left column: OPS codes, central columns: mapped SNOMED CT codes, right column: logical connection between SNOMED CT codes

1	**Procedure with finding: finding is not represented in the map (would require complex post-coordination)**			
	1-265.4 Electrophysiological examination of the heart, catheter-assisted: In tachycardia with narrow QRS complex or atrial tachycardia		175131000 Percutaneous transluminal electrophysiological studies on conducting system of heart	
2	**Procedure with device: device is not represented in the map (would require complex post-coordination)**			
	1-266.1 Electrophysiological examination of the heart, not catheter-assisted: implanted cardioverter defibrillator (ICD)		252425004 Cardiac electrophysiology	
3	**Procedure with body part: body part is not represented in the map (would require complex post-coordination)**			
	1-268.3 Cardiac Mapping: Right Ventricle		21032000 Cardiac mapping	
4	**Part of the procedure requires separate coding**			
	1-430.1 Endoscopic biopsy of respiratory organs: bronchus 312849006	312849006 Biopsy of bronchus	10847001 Bronchoscopy	ADD

5	**Logical conjunction of specific procedure and anatomy-related procedure**				
	1-490.6 Biopsy without incision on skin and subcutaneous tissue: lower leg	287538006 Non-surgical skin biopsy	118714000 Procedure on lower leg		AND
6	**Different granularity in SNOMED requires post-coordination in one case but not in another**				
	3-825 Magnetic resonance imaging of the abdomen with contrast		432369004 Magnetic resonance imaging of abdomen with contrast		
	3-826 Magnetic resonance imaging of the musculoskeletal system with contrast agent	58713006 Magnetic resonance imaging of musculoskeletal structures	51619007 Magnetic resonance imaging with contrast		AND
7	**Coordination needed to add the feature that a procedure is a diagnostic one:**				
	1-631 Diagnostic Esophagogastroscopy	392153002 Esophagogastroscopy	103693007 Diagnostic procedure		AND
8	**Missing of aggregations at the level "vessel" (regardless of whether artery or vein)**				
	3-611.x Phlebography of cervical and thoracic vessels: Other	4008007 Phlebography of neck	60006002 Intrathoracic phlebography		OR
9	**Exclusion statements for OPS codes (cannot be expressed by SNOMED CT semantics)**				
	1-207.2 Video-EEG (10/20 Electrodes). Excl.: Video-EEG during pre and intraoperative epilepsy assessment		252738008 Video electroencephalogram		
10	**Residual class "other", i.e. logical complement (cannot be expressed by SNOMED CT semantics)**				
	1-273.x Right heart catheterization: Other		40403005 Catheterization of right heart		
11	**Explicit definition "without" (cannot be expressed by SNOMED CT semantics)**				
	1-275.0 Transarterial Left Heart Catheter Examination: Coronary angiography without further action	33367005 Coronary angiography	67629009 Catheterization of left heart		AND
12	**Distinction between logical conjunction "AND" and addition (more than one instance in the target representation ("ADD"))**				
	1-650.2 Diagnostic Colonoscopy: Total, with Ileoscopy	174184006 Diagnostic endo-scopic examination on colon	235150006 Total colonoscopy	265387003 Diagnostic endo-scopic examination of ileum	(X AND Y) ADD Z

5. Discussion

Given the size of the two terminologies and the fact that the mapping was done only with the most frequent 1000 codes (i.e. the most frequent medical procedures, covering 80.9% of the procedure coding results used in German university hospitals), the relatively low amount of exact mappings and the frequent need of post-coordination of two or more SNOMED CT concepts may be surprising. However, knowing the large structural differences between these two coding systems and their distinct scenarios of use explains the differences. For instance, many frequent OPS codes contain numeric criteria (number of treatment sessions, duration of interventions, number of sites where a complex intervention takes place, dosage of drugs). This is mainly because complexity and the treatment costs grow with these numeric values. By including them in the definition of codes, the use of OPS as a tool for billing becomes more convenient, since multiple assignments of the same code, e.g. for each single application of a drug, are not necessary. A similar case is the use of single codes for complex treatments, e.g. stroke. Here the assignment of one single code depends on fine-grained rules (for stroke, comprising 592 words), including the frequency of monitoring, the required diagnostic measures and the specialty of the clinicians involved. In rheumatology, complex therapies with integrated function-oriented and pain-therapeutic treatment sections, often lasting a week or more, are prerequisite for an efficient acute care of chronically ill patients. The combination of a multitude of "small" clinical procedures requires appropriate codes that represent the overall effort without coding each single procedure [14]. There are also cases, in which the exact OPS code depends on the computation of a score that estimates the overall effort spent in complex treatments (32 among 1000 OPS codes).

On the other hand, many of the simple procedures, which are mentioned as constituents of complex therapies, like blood pressure measurement or blood sampling, as well as most lab procedures are missing in OPS. The reason is simple: the effort needed for these actions, in isolation, is just too insignificant. As much a single procedure may be relevant for clinical documentation, if it is cheap, there is no OPS code.

For SNOMED CT, such an overloading of procedure concepts would contradict its main destination as a standard for fine-grained clinical documentation, where reimbursement is not the focus. Although concepts for combined procedures exist, the focus is on encoding every single procedure.

This explains why our map required so many SNOMED CT co-ordinations, even for apparently simple concepts, and why the mapping could often not be considered exact, given the exclusion rules that assure the non-overlapping of OPS classes.

The ontological structure of SNOMED CT allows for complex logical expressions for concept refinement by its so-called post-coordination mechanism. Because such expression cannot be processed by any routine implementation, we decided to restrict ourselves to simple post-coordination patterns. This explains inexact mappings, e.g. when an indication, a body part of a device was missing for a perfect semantic match between an OPS code and a SNOMED CT expression (examples 1 – 3 in Table 6). In other cases, the meaning of an OPS code could be represented by the logical combination of two or three SNOMED CT classes (example 4 - 7). The same meaning could also be achieved by using the SNOMED compositional syntax. The semantic equivalence between such different syntactic forms could be ascertained by a description logics reasoning engine like SNOROCKET [15]. Example 6 shows variation in the degrees of pre-coordination: "Magnetic resonance imaging of the abdomen with contrast" maps to one SNOMED CT concept, whereas "Magnetic resonance imaging of the musculoskeletal system with contrast" requires post-coordination.

Pre-coordinated SNOMED CT concepts often skip certain anatomical hierarchy levels, e.g. they require a distinction between arteries and veins, whether OPS often just refers to "vessels" (Example 8). In other cases, the anatomic delineation of an OPS code (e.g. knee + thigh for certain procedures on skin) has no correlate in SNOMED CT. Post-coordination by simple conjunction is here not possible; the way out would be a disjunctive expression ('procedure X on skin of knee' or 'procedure X on skin of thigh'), which is, however, not supported by SNOMED CT logics.

A tricky issue is the distinction between one single procedure and a set of procedures. This was the reason for the "ADD" operation, which in contrast to "AND" just lumps codes together. Given the special semantics of SNOMED CT procedures (see comment to "role groups"), one could argue that even complex procedures could be expressed as logical conjunctions. This might be an argument in favour of substituting all "ADD" statements by "AND" statements in order to simplify implementations.

The limits of SNOMED CT's post-coordination power are also reached when it comes to negation, such as in explicit exclusion rules, residual classes ("others") and implicit negations in labels including "without" (examples 9 - 11). This is explained by the OWL EL profile used for SNOMED CT, which lacks negation.

Finally, a complicating factor was the language gap (German vs. English) and the lack of clear term definitions. The coders, who were not specialists in any surgical or diagnostic discipline, depended on medical textbooks and online references, in order to make clear whether a German term meant the same as its supposed English translation. In many cases, this was difficult.

In the light of all these factors, the relatively low inter-coder agreement rates were not surprising. Especially the distinction between "exact" mapping and "broader" was not easy. That the inter-coder agreement in the last phase was even lower than in the early phase (see Tables 3, 5) is explainable by the fact that in the early phase many complex mappings had been left out and encoded as "revisit".

Our results reinforce the large difference between medical coding systems even between those that cover the same domain. The dependency of coded information on the specifics of the vocabulary used and the purpose of the codes cannot be emphasized enough. The problem of re-use of administrative codes for other scenarios has been repeatedly addressed [16]. So far, idiosyncratic procedure codes are normally the only source of procedure information, beyond EHR narratives. Mappings to an international standard as SNOMED CT is, in theory, a partial solution to this problem; but to be of high quality it will require leveraging its complete post-coordination mechanism, supported by description logics reasoning, in order to obtain a high coverage and convincing retrieval results. Another way to achieve interoperable EHR information would be the application of natural language processing technology to clinical narratives in which medical procedures are referenced. The quality and comprehensiveness of such data depend, however, on natural language resources like lexicons and annotated corpora for training. Both, however, still constitute a major bottleneck, given the dynamics of medical language on the hand and the difficulties to share clinical real-world data due to privacy issues, on the other hand.

Once the content of EHRs is comprehensively represented by standardised information models and terminologies, a new mapping challenge will arise, viz. inferring administrative codes like OPS from SNOMED CT codes in context. This could put a new task on the agenda, viz. the construction of mappings in the inverse direction. This has already been discussed in the context of ICD-11; a suggested formalism was to express classification codes as queries on SNOMED CT coded EHR data [17]. Thus, a new generation of medical classifications could be rooted in and maintained by using international EHR standards like SNOMED CT.

6. Conclusions and Outlook

The exchange of clinical real-world data is a major desideratum; however, no universal coding standards are currently used for medical procedures. We report on mapping codes from OPS, the German coding system for therapeutic and diagnostic procedures, to SNOMED CT concepts, under the hypothesis that this might be a route towards worldwide interoperability of clinical data. A team of three terminologists has mapped the 1000 most frequently used OPS codes to SNOMED CT. After analysing a pilot set of 100 codes, mapping guidelines were derived. An intermediate analysis of the mappings showed that about one third of OPS codes could precisely mapped to one SNOMED CT code or a conjunction of up to three codes. A higher degree of precise maps would require sophisticated post-coordination but even in these cases, maps are still approximate. About 5% of the codes could not be mapped mostly due to complex rules for codes that aggregate many elements of complex therapies or codes for medication administration. The mapping result is affected by several factors such as lack of precise definitions in either terminology system, translation problems, or the need of fine-grained specialist knowledge in some areas. This is also one reason behind the rather low inter-coder agreement.

The team is currently extending the mapping, covering the most frequent 2,125 OPS codes, which correspond to 90% of the encodings in the underlying clinical dataset. This map will be published in early 2020 by TriNetX.

References

[1] S. Schulz, P. Daumke, M. Romacker, P. López-García. Representing oncology in datasets: Standard or custom biomedical terminology? *Informatics in Medicine Unlocked* **15** (2019), 100186.

[2] S. Schulz, J. Ingenerf, S. Thun, P. Daumke. German-Language Content in Biomedical Vocabularies. In *CLEF (Working Notes)*, Valencia, Spain, 2013.

[3] C. Maier et al. Towards Implementation of OMOP in a German University Hospital Consortium. *Appl Clin Inform*. **9** (2018), 54–61.

[4] SNOMED International. *SNOMED CT - The global language of Health*, 2019, www.snomed.org .

[5] A. Zaiss, H.P. Dauben. Prozedurenklassifikation im Spagat zwischen Statistik und Abrechnung. [ICHI-International Classification of Health Interventions : A balancing act between the demands of statistics and reimbursement]. *Bundesgesundheitsblatt* **61** (2018), 778–786.

[6] MAGPIE – *Map Assisted Generation of Procedure and Intervention Encoding*. National Library of Medicine, 2019, https://magpie.nlm.nih.gov .

[7] F. Baader, D. Calvanese, D.L. McGuinness, D. Nardi, P.F. Patel-Schneider. *The Description Logic Handbook. Theory, Implementation and Applications. 2nd edition*. Cambridge University Press. Online publication July 2010.

[8] B. Motik, B. Cuenca Grau, I. Horrocks, Z. Wu, C. Lutz. *OWL 2 Web Ontology Language Profiles (2nd Edition)* W3C Recommendation 11 December 2012. https://www.w3.org/TR/owl2-profiles .

[9] J. Ingenerf, W. Giere. Concept-oriented standardization and statistics-oriented classification: continuing the classification versus nomenclature controversy. *Methods Inf Med* **37** (1998), 527–539.

[10] Deutsches Institut für Medizinische Dokumentation und Information (DIMDI). *Operationen- und Prozedurenschlüssel. Version 2019.* https://www.dimdi.de/static/de/klassifikationen/ops/kode-suche/opshtml2019 .

[11] S. Schulz, S. Hanser, U. Hahn, J. Rogers. The semantics of procedures and diseases in SNOMED CT. *Methods Inf Med.* **45** (2006), 354–358.

[12] S. Müller-Bergfort, J. Fritze. Diagnose- und Prozedurendaten im deutschen DRG-System [Extent and use of administrative hospital data in the German DRG system]. *Bundesgesundheitsblatt* **50** (2007), 1047–1054.

[13] SNOMED CT Browser. https://browser.ihtsdotools.org/.

[14] H.J. Lakomek et al. Die multimodale rheumatologische Komplexbehandlung (OPS 8-983). [The multimodal rheumatologic complex treatment (OPS 8-983)--challenges, solutions and perspectives]. *Z Rheumatol*. **64** (2005), 557–563.
[15] M.J. Lawley, C. Bousquet. Fast classification in Protégé: Snorocket as an OWL 2 EL reasoner. Proc. 6th Australasian Ontology Workshop (IAOA'10). *Conferences in Research and Practice in Information Technology* **122** (2010).
[16] W.R. Hersh et al. Caveats for the use of operational electronic health record data in comparative effectiveness research. *Medical care* **51(8 Suppl 3)** (2013), S30-7.
[17] S. Schulz. J.M. Rodrigues, A. Rector, C.G. Chute. Interface Terminologies, Reference Terminologies and Aggregation Terminologies: A Strategy for Better Integration. *Stud Health Technol Inform*. **245** (2017), 940–944.