FinJeHeW Finnish Journal of eHealth and eWelfare

SCIENTIFIC PAPERS

VERTAISARVIOITU
KOLLEGIALT GRANSKAD
PEER-REVIEWED
www.tsv.fi/tunnus

Manual annotation of narrative patient charts – Finnish experiences related to a multilingual text corpus

Päivi Mäkelä-Bengs[1], Päivi Hämäläinen[2], Virpi Kalliokuusi[2], Riikka Vuokko[3]

[1] City of Tuusula Health Care, Tuusula, Finland; [2] National Institute of Health and Welfare, Finland; [3] Ministry of social affairs and health, Helsinki, Finland

Riikka Vuokko, Ministry of social affairs and health, Department for Steering of Healthcare and Social Welfare (OHO), Unit for Digitalisation and Information Management (DITI), P.O. Box 33, 00023 Valtioneuvosto, FINLAND. Email: riikka.vuokko@stm.fi

Abstract

ASSESS CT project evaluated SNOMED CT use for patient information exchange in EU. Finland was one of the six participating EU states. The Finnish part of the research was conducted in the National Institute for Health and Welfare. The Finnish experiences and results are interesting from the perspective of a minority language.

Research purpose was to compare SNOMED CT to two alternative terminology scenarios, a UMLS terminology set and a value set of national codes. The Finnish research team participated in the UMLS scenario. Clinical text samples were gathered from the six states resulting in a corpus of 60 texts. All texts were translated to six research languages. The annotators' task was to identify clinically relevant concepts of a corpus text, add respective codes using a term browser, and evaluate concept and term coverages. The Finnish team conducted annotations as two pairs. The annotators chunked text samples covering 23 % of corpus texts by the first annotator and 35 % by the second. For clinical concepts, the annotators added 818 codes in total, of which 270 (33 %) were exact matches and 548 (66 %) different ones. Main issues affecting the Finnish results were corpus translation quality in a multilingual context and vagueness of annotation guidelines contributing to different interpretations of included semantic groups. Consequently, limited terminology content in Finnish affected results. However, the annotation bridges a path towards more comparable evaluation results of international reference terminologies such as SNOMED CT. The experiences can be used to inform a national level implementation decisions.

Keywords: data annotation, clinical coding, data accuracy, terminology, systematized nomenclature of medicine

Introduction

ASSESS CT project was launched to evaluate large-scale use of SNOMED CT in clinical context for cross border information exchange in the EU [1]. SNOMED CT is a comprehensive healthcare terminology used in electronic health record systems (EHRs) to improve information exchange and retrieval, decision support and use of other computational tools available in EHRs by providing standardized clinical meaning based on defined concepts and terms [2]. Consequently, SNOMED CT could increase efficiency of data exchange and interoperable registry data.

Reference terminologies enhance re-use and transmission of patient data with increased data quality, thus reducing costs and saving time in documenting and in processing of data [3,5]. In EHRs, terminologies provide

meaning structures that are a prerequisite for structuring clinical content and supporting workflows. However, semantic interoperability is challenging as various terminologies are used [5]. Clinical terminologies are diversified based on specialty specific tasks [6]. Reconciling clinical terminologies requires data content analysis and concept mapping that is costly and time-consuming. Studies to determinate quality and accuracy of reference terminologies have not provided well transferable results [4,7,8].

Strengths and weaknesses of a terminology can be inspected by its structural formalism, domain coverage, comprehensiveness and functionality. Cornet et al. [3] state that content shortcomings (e.g. concept coverage vs. gaps) can be solved relatively easy, whereas formalism issues (e.g. how term relations are arranged, what structure a terminology has) or lack of functionality (e.g. how terms are retrievable, how terms are included or excluded hierarchically) are harder to solve.

The EU funded ASSESS CT project was carried out during 2015-2016 by experts from six EU member states with three SNOMED CT use scenarios; adopt, alternative and abstain. The scenarios used different terminology resources: SNOMED CT for adopt, UMLS for alternative and national code sets for abstain scenario. UMLS (The Unified Medical Language System Metathesaurus) is a multi-purpose thesaurus that contains biomedical and health related concepts, synonyms and concept relationships arranged as a semantic network. It is used in documenting patient care and in billing, statistical work, research and indexing. [9,10] Although UMLS is available in several languages, its terminology content is less extensive for smaller languages than for English. The Finnish team participated in the alternative scenario with the UMLS terminology. In the ASSESS CT alternative scenario, a Finnish UMLS terminology set consisted of three international classifications: International Classification for Diseases (ICD-10), Anatomical Therapeutic Chemical (ATC) classification for medication, Logical Observation Identifiers Names and Codes (LOINC) for medical laboratory observations, and additionally MeSH terms for medical subject headings. National code sets were not included in the UMLS terminology. Finnish UMLS set included 31622 concepts and 40675 terms.

ASSESS CT aimed to explore mapping of clinically relevant terms in different languages. Both the SNOMED CT and UMLS scenarios were limited to pre-defined semantic groups. The semantic groups covered relevant medical concepts: anatomy, chemicals and drugs, concepts and ideas, devices, disorders, genes and molecular sequences, living beings, objects and procedures [1]. Semantic groups excluded patient demographic information, phases of care etc.

In the annotation of clinical texts, both reference terminologies proved similar performance; with SNOMED CT 86 % and with UMLS 88 % of the clinical texts could be coded. Inter-annotator agreement levels comparing coding coverages between two or more annotators with same task were moderate; 30 % for UMLS and 49 % for SNOMED CT. [1] Inter-annotator agreement rates describe how similarly annotators choose a semantic category to determine reliability of annotation results. These rates were low for the Finnish results and we set out to analyze reasons for this. Overall analysis results conducted with statistical methods are reported in the ASSESS CT project deliverables [e.g. 1,7]. Instead, our emphasis is on what can be learned from the annotation experience and what kind of observations the annotators did from a perspective of a minority language. During annotation process, our focus was on information quality and re-usability of patient documentation [8], which suffers from inconsistent practices that in turn result in coding disagreements.

Material and methods

The ASSESS CT corpus was built up of clinical text samples in six languages (Dutch, English, French, Finnish, German and Swedish). Ten texts were selected for each language. Professional translators translated the texts into other five languages. This resulted in a parallel corpus of 60 clinical text samples. A sample consisted of 400-600 characters structured by document type and clinical discipline [1]. In this paper, our corpus text includes four discharge summaries, five outpatient sum-

maries, one findings/visit clinical report, three findings documents, five autopsy reports, one toxicology report, and one medication documentation.

Our team worked in two groups: a medical expert with an assisting terminology expert and another with a coding expert. Both medical experts received a spreadsheet with 40 text samples translated in Finnish with an overlap of 20 texts. We base our in-depth inspection on the overlapping 20 texts. The parallel texts cover all six source languages; 3 annotations are of Dutch origin, 3 English, 6 Finnish, 2 French, 2 German and 4 Swedish.

Annotation guidelines were produced by the ASSESS CT group, which arranged training as a webinar to work through the unfinished guidelines. Finnish annotators started right after the webinar. Averbis terminology browser was used for searching terms. The search function allowed entering both text strings and codes, if known. The search function did not work with synonyms or related concepts. The search result was displayed with synonyms and concept hierarchies.

The annotation workflow had four primary steps: analysis of clinical content, chunking identified terms to form clinical concepts related to the semantic groups, adding codes, and lastly, reviewing concept and term coverage. Using spreadsheets, the annotators first identified clinically relevant concepts, i.e. "chunks" of a given clinical text, and then proceeded to add codes found in the UMLS to map with the chunk's concept information and finally rate concept and term coverages for the coding. A chunk was defined to cover one clinical concept, but chunk delineations varied a lot based on individual ideas and preferences. Below, in Figure 1, is one example of different types of chunk demarcations by annotators A1 and A2 resulting in different coding.

Concept coverage was used to indicate the degree of successful representation of conceptual matches between reference terminologies and the chunks identified in the text samples. Concept coverage was rated according to five degrees; full, inferred, partial, no coverage and out of scope. Term coverage was used to measure the degree by which the linguistic forms of the terms used in the text samples showed matches with the respective term forms of the reference terminologies.

	A1	UMLS				A2	UMLS			
TOKENS	CHUNK	CODE UMLS CUI	termit	CONCEPT COVERAGE SCORE	TERM COVERAGE Y/N	CHUNK	CODE UMLS CUI	termit	CONCEPT COVERAGE SCORE	TERM COVERAGE Y/N
tuotiin						10	C1136313	Päivystysaikainen hoito	Partial cov	no
päivystykseen						10	C1136313		Partial cov	no
kaaduttuaan	5	C0085639	Määrittämätön kaatuminen tai putoaminen	Inferred cov	yes	11	C0478694	Kaatuminen samalla tasolla	Partial cov	no
.										
Hänet						12	C0019993	Sairaalahoito	Full cov	no
sijoitettiin						12	C0019993		Full cov	no
sairaalaan						12	C0019993		Full cov	no

Figure 1. Example of an annotation spreadsheet showing a part of a clinical text sample "who was brought into emergency room after she had fallen. She was placed in hospital" arranged in token rows. A1 has identified one chunk (number 5) and added a code "Unspecified fall" with inferred concept coverage and full term coverage. A2 has identified three chunks (numbered 10-12) and added three codes "Emergency care", "Fall on same level" and "Hospital care" with partial and full concept coverage but no term coverage.

The annotation results were post-processed to reduce errors due to trivial mistakes, such as missing coding or coverage ratings. Inter-annotator agreement was measured using Krippendorff's alpha that compares codes assigned more than once within a subset of 20 text samples across the languages [1,11]. The inter-annotator agreement on concepts was rated based on a hypothesis stating the following: "The more codes coincide, which the annotators propose for representing the same piece of meaning, the more suitable is the terminology setting." [1, p. 24] For our analysis, we reviewed concept and term coverages to inspect how much two annotators had similar vs. different coverages.

Results

The Finnish annotators conducted the annotations from July 2015 to February 2016. The annotators chunked the texts differently; chunks defined by A1 covered 23 % of a text and by A2 35 %. Within the chunks, in total 818 codes were added. Of the codes, 270 codes (33 %) were exact matches between the two annotators, whereas 548 codes (66 %) were chosen differently by A1 and A2. We illustrate different reasons contributing to a systematic bias apparent in the Finnish annotation results.

Results of the corpus and translation

Translation quality left a lot for the annotators to interpret. ASSESS CT had 3 translation principles that caused loss of a text sample's original meaning: Acronyms and abbreviations were not translated and no full term was provided. Drug names were replaced with their active ingredients to define not language specific ATC coding. Spelling or grammar errors of the original texts were supposed to be introduced similarly into the translations causing further interpretation issues. The translations were not culturally or professionally sensitive. This caused expressions that were hard to interpret or clinically out of scope or could not be fully interpreted by the annotator.

Both Finnish annotators reported frustration when they could not receive sufficient contextual information for interpreting the meanings of the texts. They had no way of ensuring, for example, the phase of care. Table 1 illustrates how interpreting the phase of care resulted in different coding results, for example when choosing between symptoms (observations) and findings (diagnosis).

Table 1. Different coding interpretations. Corpus text "acceptably controlled type 2 diabetes mellitus" was coded by A1 with "Adult-onset diabetes, type 2 diabetes" and by A2 with "Adult onset diabetes without complications".

Annotator / *Text*	***A1***					***A2***				
	Chunk	**UMLS code**	**Term**	**Concept cov**	**Term cov**	**Chunk**	**UMLS code**	**Term**	**Concept cov**	**Term cov**
hyväksyttävästi						20	C0494290	Aikuistyypin diabetes ilman komplikaatioita	Partial cov	No
säädelty						20	C0494290		Partial cov	No
tyypin	12	C0011860	Aikuistyypin diabetes, Tyypin II diabetes	Full cov	Yes	20	C0494290		Partial cov	No
II	12					20	C0494290		Partial cov	No
diabetes	12					20	C0494290		Partial cov	Yes

Choosing a level of granularity seemed to produce different coding results and often partial concept coverage (see Table 2). The coding granularity refers to the level of identifying a clinically relevant concept. Depending on how much inferred meaning the annotators attached to a text sample and how loose or precise semantic equivalences they picked up from the UMLS, the annotators produced different coding and concept coverages. Table 2 illustrates the variation of semantic accuracy between A1 and A2. A1 coded the chunk with a UMLS concept lacking some semantic elements of the intended meaning of the original text chunk (the information of the anatomic structure "neck" is lacking). A2 coded the chunk with a UMLS concept having some extra semantic elements (the information of "crushing injury").

Results of the term browser

One crucial issue was the terminology browser usability. The terminology browser did not support search with synonymous concepts. Both annotators used other tools during the annotation process, such as a Finnish guidebook for ICD-10 coding and a web-based medical terminology resource supporting several languages and providing synonymous terms. Search did not support predictive text functions either, which was frustrating from the Finnish viewpoint with a language that is highly inflected and has numerous word form variants. When working with limited time resources, it was also disappointing that the terminology browser did not support continuation from a previous session after crashing.

The terminology browser limitations meant that the annotators learned a different approach to chunking during the process, making the annotation flow more fluently. The annotators started partially to pre-define chunks based on what could be searched in the terminology browser. This meant that the UMLS content started to guide decisions on what should be identified as a chunk in a text sample. Relying on the guidelines did not help with its vague notions on how clinically relevant concepts or semantic groups should be identified. This resulted in different chunking styles at the beginning vs. near the end of annotating process.

Results of the annotation process

Examining the annotation spreadsheets illustrates how differently two or more annotators might work based on the same guidelines. One reason for this may be the medical experience of the annotators. Compared to medical students working as annotators for other languages, both Finnish annotators were clinically experienced physicians with a deep understanding of health care data structures and classifications.

A1 attempted to delineate clinical concepts individually, more according to the principle of minimum differential conceptual unit, and to identify concepts in named semantic groups. A2 emphasized that the chunks should reflect the overall clinical meaning of whole text or a text passage. For A2, the coded concepts were selected for the purposes of reliable patient data exchange (see Table 3). When concept identification remains loosely covering longer parts of clinical text, choosing respective codes can result in major variations of coding. It would be even more challenging to choose exact terms covering loosely defined clinical meanings. This is a crucial issue, as defining clinically relevant concepts directly contributes to the overall coding results and concept and term coverages.

Table 2. Example of coding granularity. Corpus text "neck impression of hanging" was coded by A1 with "Hanging, strangulation or suffocation, intentionality unclear" and by A2 with "Crushing injury of neck, location unspecified".

Annotator / *Text*	***A1***					***A2***				
	Chunk	**UMLS code**	**Term**	**Concept cov**	**Term cov**	**Chunk**	**UMLS code**	**Term**	**Concept cov**	**Term cov**
kaulalla	1	C0480929	Hirttäytyminen, kuristuminen, tai tukehtumi-nen, tahallisuus epäselvä	Inferred cov	Yes	1	C0273433	Sijanniltaan määrittelemätön kaulan murska-vamma, Kaulan murskavamma	Partial cov	No
hirttouurre	1					1	C0273433		Partial cov	No

Table 3. Different chunking styles with a sample text "Examination at appointment was scheduled at the end of the year." A1 has chunked "control in a year" with no coding available. A2 has three chunks for "appointment", "control" and "planned in a year" with full concept coverage coding for contact type and content but no coverage for point of time. The translated codes for A2 are: "appointment" and "general clinical examination".

Annotator / *Text*	***A1***					***A2***				
	Chunk	**UMLS code**	**Term**	**Concept cov**	**Term cov**	**Chunk**	**UMLS code**	**Term**	**Concept cov**	**Term cov**
Vastaanotolla						20	C0028900	Vastaanottokäynnit	Full cov	No
tehtävän						21	C0260860	Lääketieteellinen yleistarkastus	Full cov	No
tarkastuksen	8			No cov	No	21	C0260860		Full cov	No
ajankohta						22			No cov	No
on						22			No cov	No
sovittu						22			No cov	No
vuoden	8					22			No cov	No
loppuun	8					22			No cov	No

Table 4. Identifying semantic groups with a text sample "Patient is 80 years old women that was brought to emergency department after she had fallen". A1 coded "Unspecified fall" and A2 "Emergency care" and "Fall on same level".

Annotator / *Text*	***A1***					***A2***				
	Chunk	**UMLS code**	**Term**	**Concept cov**	**Term cov**	**Chunk**	**UMLS code**	**Term**	**Concept cov**	**Term cov**
Potilas						1			No cov	No
on										
80-						2			No cov	No
vuotias						2			No cov	No
nainen						3			No cov	No
tuotiin						10	C1136313	Päivystysaikainen hoito	Partial cov	No
päivystykseen						10	C1136313		Partial cov	No
kaaduttuaan	5	C0085639	Määrittämätön kaatuminen tai putoaminen	Inferred cov	Yes	11	C0478694	Kaatuminen samalla tasolla	Partial cov	No

Table 5. Example of different coding styles with a text sample "we detected that weight was 76 kg and height 159 m. Blood pressure…" A1 coded "Weight and units" and "Blood pressure measurement". A2 coded "Body weight and measurements", "Weight and units" and "Blood pressure measurements".

Annotator / Text	A1					A2				
	Chunk	**UMLS code**	**Term**	**Concept cov**	**Term cov**	**Chunk**	**UMLS code**	**Term**	**Concept cov**	**Term cov**
totesimme						4, 6	C0005912	Kehon painot ja mitat	Partial cov	No
painon	2	C0043101	Painot ja mittayksiköt	Parial cov	No	4	C0005912		Partial cov	No
olevan						4	C0005912		Partial cov	No
76	2					5	C0043101	Painot ja mittayksiköt	Partial cov	No
kg	2					5	C0043101		Partial cov	No
ja										
pituuden	3			No cov	No	6	C0005912	Kehon painot ja mitat	Parial cov	No
1,59	3					7	C0043101	Painot ja mittayksiköt	Partial cov	No
metriä.	3					7	C0043101		Partial cov	No
Verenpaine	4	C1313910	Verenpaineen mittaus	Full cov	Yes	8	C1313910	Verenpaineen mittaus	Full cov	No

The guidelines listed semantic groups for concepts that should be included vs. excluded. A2 often chose to cover concepts belonging to other semantic groups. In Table 4, A1 has chosen to code only "falling" and A2 "falling" with additional demographic information (patient, age and sex) and specialty (emergency care). From the clinical point of view, all information is relevant.

Additional reason for different outcomes was the style of coding (see Table 5). A1 started working on annotations right away, when the annotation guidelines were still under modification. She entered a code or codes only on the first row of an identified chunk. A2 followed the finished annotation guidelines stating that codes should be entered for each row of a chunk.

Both annotators were worried about the UMSL content leaving out clinically relevant concepts. They hesitated to interpret meaning of clinical concepts: for example, is a text about previous procedures and laboratory tests vs. findings and/or future planning. Due to poor translation quality, verb tenses did not help interpreting. Procedure codes were missing in the UMLS content as well as some morphology codes. Likewise, qualifiers for examination results or their interpretation were missing.

Results of the concept and term coverage

Finnish results were different when compared to other languages in the ASSESS CT. With the Finnish alternative terminology, resulting concept coverage 36 % was lower than the study average. In the total ASSESS CT results, the concept coverage for UMLS was slightly higher than for SNOMED CT; 60-64 % with UMLS and 43-45 % with SNOMED CT. [1]. Term coverage results would suggest that it is not sufficient to provide just one term per concept when a terminology is being translated and localized. For minor languages, lack of terminology content was apparent and for the Finnish UMLS especially limiting. The included terminology did not cover even all semantic groups that were mentioned in the study guidelines. This caused a lot of identified chunks with "no coverage" or "partial coverage" in the annotations.

Table 6. Sums for concept and term coverage ratings.

Annotator:	A1		A2	
	Concept coverage			
Full coverage	92	26 %	202	21 %
Inferred coverage	56	15 %	0	
Partial coverage	98	28 %	338	35 %
No coverage	106	30 %	415	43 %
Out of scope	3	0,1 %	1	
	Term coverage			
Yes	146	41 %	123	13 %
No	206	58 %	838	87 %

When reviewing the annotations, it became apparent that there was no reconciliation over concept and term coverage types. Table 5 illustrates this with annotation resulting in the same code C1313910 for measuring blood pressure by both A1 and A2. As the original text has only "blood pressure" both annotators have full concept coverage, but results of term coverages are different. A2 has no coverage, as the term is not exactly the same, while A1 has full term coverage based on search term used in Averbis. Rest of Table 5 indicates similar interpretation. A1 searched concepts "weight" and "height" resulting in only one code for "weight and measurements". A2 has additional code with partial concept coverage for "body weight and measurements" to compensate the lack of separate code for "height". Both annotators have "no" for term coverage.

Difference of coding is apparent when comparing the number of coverage ratings (see Table 6). A2 has a larger number of concepts and terms to cover, and the coverages illustrate different interpretations of guidelines, when A1 attempted to use all alternatives when searching for applicable terms resulting in inferred coverages.

Discussion

ASSESS CT study shows that SNOMED CT could provide a reference terminology for patient information exchange in Europe [1]. Finland joined ASSESS CT to learn about cross-border information exchange and to evaluate pros and cons of a reference terminology use. There is evidence that well-defined data structures support clinical work [e.g. 8]. However, the goal setting of ASSESS CT was not clear; was our primary task to identify relevant clinical concepts and code them or to code the texts with sufficient granularity to support patient information exchange without changes of clinical meaning. Structured data would be usable in cross-border care, but requires a common reference terminology or other ways to ensure semantic level mapping of national and local data structures. The overall ASSESS CT results show that SNOMED CT use might be complex as it requires coding experience [1]. Language might also limit its implementation as currently SNOMED CT has no extensive Finnish content. Use of SNOMED CT would require extensive concept mapping, translation or other terminology localization work to increase its presumed benefits.

What caused the Finnish annotators to choose different codes or evaluate even opposite concept and term coverages so often during the annotation? Our examples illustrate how contextual information is clinically relevant, thus causing deviation of interpreting the semantic groups. Other reason for different granularity of coding might emerge from medical specialties with slightly diverse notions of how to classify medical information or what terms should represent a concept. Although a full agreement between the annotators would be unrealistic due to complexness of the clinical

domains [1], we have inspected annotation as a learning process. The annotators learned to code concepts with terms available in the UMLS.

A clear limitation of ASSESS CT was the corpus translation quality. Translation quality was especially low for Finnish despite the corrections done during the translation validation. Partly same people validated the corpus texts and conducted the annotations. For example, short forms (abbreviations) and brand names were not translated or localized, which caused additional interpretation difficulties and resulted mostly in "no coverage". The annotators could not know which errors of a text sample were mimics of the original text to be accounted for and which ones were translation errors to be corrected. Text quality would have required more preparation and a translation style guide in the ASSESS CT project. With a style guide, it may have been possible to trace back the implications of poor quality original texts and those of poor quality translations.

Representativeness of clinical texts is another limitation as the text samples were not based on defined sampling approaches. Sampling was carried out based on what kind of texts were readily available and different types of texts are not necessarily represented evenly. All Finnish corpus texts are based on pathology records. Resourcing issues emerged strongly in the Finnish annotation team when the annotations required more effort than was originally planned. Due to Averbis browser limitations combined with the corpus translation quality, additional terminological references were used.

Inter-annotator agreement values were low or moderate in the overall results (30-49 %). There was no significant difference of concept coverage between SNOMED CT adopt and UMLS alternative scenarios [1]. However, inter-annotator agreement results illustrate challenges of choosing only one term from the reference terminology to represent a concept, when in fact two or more terms might be depicted. Diverse backgrounds might contribute to deviation in clinical terms chosen. Even if the annotators did not agree on a specific code, they were confident to have chosen a correct code. In practice, this would hinder interoperability by resulting in deviation of coding based on personal preferences. Even when annotators had the same code, they were often in disagreement about the coverage rating. An example of this is illustrated with coding for blood pressure in Table 5.

ASSESS CT results show that more detailed guidelines and training of annotators is needed in the future. The Finnish annotators recommend establishing uniform workflows for more comparable results. Guidelines should instruct how to deal with language specific features in a multi-language context. The guidelines did not instruct how to deal with structural language issues that affect the formation of concepts in general and more precisely the identification and delineation of clinically relevant concepts.

Improvements of the structure and content of a reference terminology should be processed to ensure its better use [cf. 6]. Usability of terminologies can be increased by allowing use of synonyms in searches, supporting predictive phrase entry to suggest a term and revising a terminology structure and term relations in a system setting. Better term coverage in SNOMED CT or possible hybrid terminology, such as UMLS would improve terminology's fit for use. In the Finnish context, this would mean more in depth evaluation of SNOMED CT before possible implementation.

Conclusions

The lessons learned can guide new annotation studies and give insight for national level adaptation of an international terminology. The demand for national level and cross-border interoperability of medical information is rising fast, and mapping standardized reference terminologies to EHR specific interface terms has been suggested as one solution to support information exchange [e.g. 5, 12]. Patients travel more and have bigger expectations of information sharing. International collaboration in medical research requires interpretation of multilingual data in various registers. The ASSESS CT annotation experience has shown limitations but also possibilities for improving interoperability.

Acknowledgements

Producing this research paper did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors. The ASSESS CT project was funded by an EU Horizon 2020 research grant.

Annotation research was conducted in THL while all four authors were working there.

References

[1] ASSESS CT D2.3 – Multilingual and multidisciplinary study of terminology coverage and quality, ASSESS CT - Assessing SNOMED CT for Large Scale eHealth Deployments in the EU. Sep 2016. Available from: http://assess-ct.eu/fileadmin/assess_ct/deliverables_re/assess_ct_ga_643818_d2.3.pdf

[2] SNOMED CT. International Health Terminology Standards Development Organisation, IHTSDO. Sep 2016. Available from http://www.ihtsdo.org/snomed-ct

[3] Cornet R, de Kaizer NF, Abu-Hanna A. A Framework for Characterizing Terminological Systems. Method Inform Med 2006;45(3):253-266. https://doi.org/10.1055/s-0038-1634079

[4] Mullins HC, Scanland PM, Collins D, Treece L, et al. The Efficacy of SNOMED, Read Codes and UMLS in Coding Ambulatory Family Practice Clinical Records. Proceedings of AMIA Annual Fall Symposium. 1996:135-139.

[5] Rosenbeck Gøeg K, Chen R, Radorff Højen A, Elberg P. Content analysis of physical examination templates in electronic health records using SNOMED CT. Int J Med Inform 2014;83(10):736-748. https://doi.org/10.1016/j.ijmedinf.2014.06.006

[6] Ingenerf J, Reiner J, Seik B. Standardized terminological services enabling semantic interoperability between distributed and heterogeneous systems. Int J Med Inform 2001;64(2-3):223-240. https://doi.org/10.1016/S1386-5056(01)00211-8

[7] ASSESS CT D1.3 Current and Future Use of SNOMED CT Interim Report, ASSESS CT - Assessing SNOMED CT for Large Scale eHealth Deployments in the EU. Apr 2015. Available from http://assess-ct.eu/fileadmin/assess_ct/deliverables/assess_ct_d1.3_current_and_future_use_of_snomed_ct.pdf

[8] Vuokko R, Mäkelä-Bengs P, Hyppönen H, Lindqvist M, Doupi P. Impacts of structuring the electronic health record: Results of a systematic literature review from the perspective of secondary use of patient data. Int J Med Inform 2017;97:293-303. https://doi.org/10.1016/j.ijmedinf.2016.10.004

[9] Andrews JE, Richesson RL, Krischer J. Variation of SNOMED CT Coding of Clinical Research Concepts among Coding Experts. J Am Med Inform Assn 2007;14(4):497-506. https://doi.org/10.1197/jamia.M2372

[10] Unified Medical Language System (UMLS). U.S. National Library of Medicine, National Institute of Health. September 2016. Available from https://www.nlm.nih.gov/research/umls/

[11] Krippendorff K. Content Analysis: An Introduction to Its Methodology. 2nd Ed. Sage Publications, Inc.; 2003.

[12] ASSESS CT Recommendations. ASSESS CT - Assessing SNOMED CT for Large Scale eHealth Deployments in the EU Dec 2016. Available from http://assess-ct.eu/fileadmin/assess_ct/final_brochure/assessct_final_brochure.pdf.