Mapping between

international medical terminologies

Elena Cardillo

Institute of Informatics and Telematics UOS Rende (CS), National Research Council, Consenza, Itally

Annex 4
to SHN Work Package 3
Deliverable D3.3

Final version, May 29, 2015

Document description

Deliverable:	Annex 4 to SHN WP3 D3.3
Publishable summary:	In this Annex, the author first describes the available methods for mapping between international medical terminologies, including ontological alignment and the use of semantic web technologies. In addition, existing mapping efforts are reviewed.
Status:	Final draft
Version:	1.2
Public:	x No □ Yes
Deadline:	May 31, 2015
Contact:	Elena Cardillo elena.cardillo@iit.cnr.it Robert Vander Stichele robert.vanderstichele@ugent.be
Editors:	Elena Cardillo

Table of contents

Note :

This manuscript was commissioned by Prof. Dr. R. Vander Stichele, Workpackage Leader of SemanticHealth Net WP3, to Elena Cardillo, Institute of Informatics and Telematics UOS Rende (CS), National Research Council, Via P. Bucci, 17B, 7 floor, 87036, Rende, Cosenza, Italy

Elena Cardillo, PhD, Email: elena.cardillo@iit.cnr.it Tel. : +390984494957

Table of content

1 Introduction

Over the last two decades research on medical terminologies and classification systems has become a popular topic, above all in the last decade, after the institutional measures, in Europe and abroad, expressing the need for an integrated management of patients' healthcare data, and for guaranteeing semantic interoperability of these data. In fact, messaging standards employed in the healthcare domain use different terms for the same concept and the consequence is often clinical misinterpretation, wrong management of the registered knowledge, misdiagnosis of the patient's problem. Much work has already been done to resolve this problem through the development and use of coding systems and clinical terminologies, to be used by physicians and other health professionals during their daily practice for coding patients' clinical data in their electronic health records (EHRs), and a huge standardization effort has allowed to establish a number of classification systems. Existing medical terminologies and coding systems vary in their coverage and completeness, and are differentiated on the basis of their purpose (e.g. diagnostic, functional capabilities, procedural, pharmaceutical), and most of them are thought to cover a relatively narrow subset of healthcare data, such as nursing procedures, problem lists, etc.. Furthermore, some systems have been proprietary, custom-built, limited, or difficult for clinicians to use. In the perspective of data exchanges between different healthcare information systems, other issues are associated to existing terminologies and coding systems to be considered: the inconstant update of the resource, so versioning, the lack of updated and context-based translations, lack of updated mappings with other terminologies or coding systems used in the same medical subdomain. As mentioned in 2.3. of the Deliverable, the most known and used coding systems, nomenclatures and thesauri are respectively the International Classification of Diseases (so far at its 10th revision, with the 11th under construction) developed by the World Health Organization (WHO); the International Classification of Primary Care (at its 2nd revision, with the 3rd under construction) developed by the (The World Organization of Family Doctors) WONCA and maintained and distributed by the Wonca International Classification Committee (WICC); the Logical Observation Identifiers Names and Codes (LOINC), developed and distributed by the Regenstrief Institute (Indianapolis, USA); the Anatomical Therapeutic Chemical (ATC) classification system developed by WHO Collaborating Centre for Drug Statistics Methodology (WHOCC); the Systematized Nomenclature of Medicine Clinical Terms (SNOMED CT) initially developed by the UK National Health Service and its Centre for Coding and Classification and then its IPRs were transferred to the IHTSDO; the Medical Subject Headings (MeSH) theusarus and the Unified Medical Language System (UMLS) Metathesaurus developed by the United States National Library of Medicine. Other internationally used examples include the International Classification for Nursing Practice (ICNP), developed by the International Council of Nurses[1], and the International Classification of Functioning, Disability and Health (ICF) from the WHO[2], which is widely used in allied health.

However, the use of these terminological resources have proved unsuccessful in resolving semantic heterogeneity in healthcare. Ontologies have also been developed to resolve this problem by making explicit the meaning of terms used in healthcare. They provide a source of shared and precisely defined concepts, resulting in interoperability by knowledge sharing and reuse. Even though, the different ways in which healthcare domain can be conceptualized resulted in the creation of a high number of ontologies which often have contradicting or overlapping parts between them. Thus, even the use of ontologies introduces semantic heterogeneity to the

[1] http://www.icn.ch/what-we-do/about-icnpr/
[2] http://www.who.int/classifications/icf/en/

healthcare domain. Another area of semantic conflicts can occur when terminologies are applied in so called information models. Standards Developing Organizations (SDO's) create information models to represent medical concepts in databases, user interfaces, data exchange formats, decision support systems, data warehouses, registries and so on. For instance, the International Standards Organization (ISO) developed the ISO 13606 series on electronic health record communication, where the first standards depicts an information model for the electronic health record content[3]. Health Level 7 (HL7) created, in its 3rd version of the messaging standards, a Reference Information Model (RIM), consisting of data classes that represent all kind of medical, demographic and administrative data, from which various messages are constructed[4]. This RIM is intended to assure the consistency in models and data, and hence semantic interoperability. However, it is well known that certain constructs in the terminology and in information modeling can conflict with each other. Examples include in particular the negation. If the terminology states "no heart infarct present", and the information class uses a negation indicator that is said to be true, then the double negation leads to dangerous errors. Health Level 7 has a specific guideline - Terminfo (available at http://www.hl7.org/Special/committees/terminfo/docs.cfm) - for the use of Snomed CT in the HL7 v3 messages to prevent such conflicts, that obviously can lead to serious errors in patient care.

In this context, an effective solution to this problem is the introduction of methods for establishing unambiguous mappings and/or matches among the different coding systems, terminologies, existing ontologies, and data models to facilitate their semantic interoperability. In fact, faced with the increasing need to allow cooperation between the various health actors and their related health information systems, it appears necessary to link and connect these terminological resources to make them interoperable. This allow the different actors to speak the same language while using different representation of the same things. Thus, the scope of mapping medical terminologies or coding systems is to provide a link between one terminology and another, and/or between terminologies and data models in order to:

- reuse data collected for a specific aim also for other purposes;
- keep the value of data in case of migration to newer database formats and schemas;
- avoid the repeated editing of the same data and consequently the risk of errors;
- facilitate the semantic interoperability between systems that are sources of data and systems that are targets for data reuse.

In the present Annex, an updated review of the literature related to medical terminology mapping methodologies and techniques is presented together with an overview of the existing mappings between the SNOMED CT and the other medical terminologies and between the most used medical coding systems. The mapping with information models is only briefly addressed, since it would require a different kind of review, although the basic principles are the same.

[3] http://www.iso.org/iso/catalogue_detail.htm?csnumber=40784
[4] http://www.hl7.org/implement/standards/product_brief.cfm?product_id=77

2 Mapping between medical terminologies

Mappings between medical terminologies or classification systems are defined for a given purpose, and represent the agreement reached between medical specialists that need to be expressed in the form of correspondence tables (see Table 1), and they must be refined for particular use cases and users in diverse settings [10]. Standardly, given two heterogeneous representations, a mapping can be viewed as a triple $<e, e', r>$, where e, e' are the entities (e.g. terms, classes, etc.) belonging to the different representations (e.g. coding systems, terminologies), and r is the relation asserted by the mapping.

Table 1 . Mapping example between two coding systems (ICD-10 and ICPC-2)

Concept in ICD-10	Concept in ICPC-2
B35 (Dermatophytosis)	S74 (Dermatophytosis)
N01 (Headache)	R51 (Headache) G44.3 (Chronic post-traumatic headache) G44.8 (Other specified headache syndromes)

The semantic correspondence between the terms belonging to the medical terminologies that need to be mapped, can be qualified in order to make a more efficient interpretation of the resulting mappings (e.g. specifying if the mapping is an exact match between the pairs, or if they are different terms expressing the same concepts, synonyms, etc.). Mappings can be performed manually or automatically. In the first case, domain experts are called to give consent and to manually identify a “one-to-one” or “one-to-n” correspondence between terms/concept of the first terminology/coding system and terms belonging to the second terminology/coding system to define explicit relationships among them. For the mapping between terminology and data models, the same levels of precision is applied.

Even if clinical mapping is created thanks to the reached consent of many domain experts and the most common practice in the past was doing it manually, in dynamic environments such as the web for obvious reasons the manual concept reconciliation is inefficient and infeasible, and so cannot provide a general solution for finding semantic mappings. The move from manual to semi-automatic matching of heterogeneous resources has been justified in the literature for addressing the issue of scalability (for matching between large resources), and by the need to accelerate the matching process. Researchers also argue for moving to fully-automatic, that is, unsupervised, schema matching in settings where a human expert is absent from the decision process. In particular, such situations characterize numerous emerging applications, such as agent communication and semantic web service composition.

As stated in the previous section, in the presence of a consistent number of existing medical terminologies and clinical data models, even providing clinical mappings between them, interoperability remains a significant problem, since the content, the taxonomy structure, the terminological coverage, the granularity, the cross-mapping, and the definitions and attributes vary between existing terminologies and models. Many established medical coding systems still lack a precise semantic support or framework, even if this issue has been addressed by the emergence of description logic encoded medical terminologies (as in the case of SNOMED CT, ICD-11, etc.). Furthermore, there is no clear formal reference that establishes the precise meaning of mappings between the coding systems, and the interpretation of mappings cannot be done coherently—

different groups of mappings can hold different implicit semantics [4]. This seems obvious since each terminology or classification system, and every data modeling approach is created for a specific purpose and mapping has to be based on a clinical consent. Consequently, a general standard for this is difficult to achieve.

3 State of the Art

3.1 Medical Terminology Mapping

In the last two decades researchers have spent lots of efforts on defining semi-automatic techniques for mapping heterogeneous medical terminologies and coding systems, recently trying also to approach their formal encoding to make them useful in different application scenarios.

Various studies have investigated the implementation of platforms to achieve interoperability between medical terminologies. The UMLS Metathesaurus is one example, even if, to be precise, it does not make semantically integrated terminology interoperable but rather provides rich health knowledge sources that can potentially be used towards mapping or connection identification. In fact, to disambiguate concepts and facilitate automated terminology merging efforts, a general trend is to make the definitions of medical concepts more explicit. In doing that, a large number of studies proposed algorithms and heuristics for discovering mapping between concepts belonging to different classifications and terminologies based on the use of the UMLS Metathesaurus as a knowledge source for extracting standardized definition for the concepts and their semantic mappings. Examples in this directions are described in [11, 8, 18] for the mapping between SNOMED RT and ICD-9-CM (International Classification of Diseases 9th release – Clinical Modifications); in [21] for the mapping between MedDra and SNOMED CT; and in [37] for the mapping between ICPC2-Plus, an interface terminology derived by the ICPC-2 classification[5], and SNOMED CT, the most comprehensive biomedical terminology in the world including more than 300,000 active concepts covering different healthcare domains. In this last study authors developed some linguistics bases for mapping the two terminologies and demonstrated that while considering the UMLS, which includes both ICPC-2 PLUS and SNOMED CT, mapping to SNOMED CT regards only the 46.5% of ICPC-2 PLUS terms, by performing lexical mapping and using post-coordination of remaining unmapped terms, authors reached a total percentage of 80.58% of ICPC-2 PLUS terms mapped to SNOMED CT. Some studies focused on the mapping of legacy interface terminologies to SNOMED CT, showing results not only in terms number of matches, quality of maps, use of post-coordination, but also in terms of observations about SNOMED CT, including inconsistencies, redundancies and omissions [35]. Finally, other researchers enriched the structure of ICD-10 using category and chapter mappings to SNOMED CT(e.g. in [23] the method was applied to a Swedish primary health care version of ICD-10).

One practical example of mapping between data required for continuity of care of stroke patients after hospitalization, showed a coverage of less than 50% between the data professionals required, and available Snomed CT concepts. This was a fully manual process, and it was restricted to the requirements of the information model in HL7 messages to get one precise matching code from Snomed CT for one data element, so only the 100% match solution was accepted [17].

Other approaches rely on the use of formal languages (e.g. First Order Logic – FOL) for encoding medical terminologies in order to extract semantic mappings. For instance, [6] provided a representation formalism based on Entity Relationship Diagrams (ERD) and First Order Logic (FOL), as part of an integration framework for representing the structure of five resources: ICD, SNOMED, NHS clinical terms, UMLS, and GALEN. GALEN as a project also informed various standards such as CLAML for the proper representation of classifications, and methods for

[5] ICPC Plus is developed and maintained by The Family Medicine Research Centre (FMRC) of the University of Sydney and contains over 7,000 terms that are commonly used in Australian general practice.

supporting cross-mappings. One example method is called *the dissection*. This is a grammar like analysis of the complete medical concepts from the perspective of a terminology model. An example of this application is explained in ISO 18104[6] where the categorial structure of underlying conceptual models of terminological constructs is analysed. For a well formed nursing diagnose, the minimum of a focus of human functioning in relation to a disease, and the professional judgement of the nurse must be present in an instance of a nursing diagnose. [13] The dissection of a nursing diagnostic statement into these two parts gives a more precise form of mapping. Additional terminological characteristics can also be added for comparison, such as subject of care, degree, potentiality and other components of the categorial structure. See table 5 in [13: p.160] as example:

Table 5. Dissection of NMDSN Concept "Patient or Family Have Fear".

NMDSN item	ICF concept	ICNP® concept
Focus: fear	Focus: range of emotion (as part of emotional functions)	Focus: fear
Judgment: fear (meaning present)	Judgment: impairment	Judgment: demonstrates fear or fear yes, or decrease/increase of fear
Subject of information: patient and/or family	Subject of information: individual client	Subject of information: individual and/or family
Degree: yes/no	Degree: no, mild, moderate, severe, complete	Degree: the set of options from the selected judgment category
Potentiality: actual	Potentiality: actual	Potentiality: actual

Decision: The focus for fear is similar for NMDSN and ICNP. ICF is too broad, although fear falls under code b1522. Nursing requires for all emotions a separate code. The wording for judgment differs, but the meaning is similar. Subject of information is different for ICF (individual only) from the others. Degree in NMDSN differs from the others. Potentiality is similar for all three. Thus, the concept "fear" cannot be cross-mapped for individual care. It is possible to map for aggregate purposes, but only if it pertains to individuals and not to groups.

3.2 Ontology Mapping and Semantic Web technologies for mapping

During the last ten years ontologies and the use of Semantic web technologies [1] has been seen as a better solution to semantic interoperability because this allows describing the semantics of information sources and makes its contents explicit by providing a shared comprehension of a given domain of knowledge [31]. As mentioned above, different kind of ontologies have been created and are employed in healthcare; these include OpenGalen, SNOMED-CT, the Foundation Medical Anatomy (FMA), etc. and other ontological representation of existing clinical terminologies (e.g. NCI Thesaurus). Unfortunately, ontologies and their structure are not really familiar and natural to most healthcare providers and their use raises heterogeneity problems to a higher level [32]. Heterogeneity can be of different types: i) syntactic heterogeneity, that occurs when two ontologies are not expressed in the same ontology language; ii) terminological heterogeneity, when different medical terms represent the same concept in different ontologies (e.g. the use of heart and cardiac to represent the same concept); iii) semantic heterogeneity that occurs whenever two contexts do not share the same interpretation of information (e.g. homonyms and synonyms) for different problems such as the difference in coverage, in granularity and perspective; and, finally, iv) semiotic heterogeneity that is caused by the subjective interpretation of the used terms by humans [16].

[6] ISO 18104:2014 Health informatics -- Categorial structures for representation of nursing diagnoses and nursing actions in terminological systems. Available at: http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=59431

To solve this heterogeneity ontology matching procedures are provided, in order to find correspondences between semantically related entities of the ontologies, thus enabling the knowledge and data expressed in the matched ontologies to be interoperable.

When considering this domain, in the literature, two active areas of research are: *Ontology Mapping* (or Ontology Matching) to show how concepts of one ontology are semantically related to concepts of another ontology [24]; and *Ontology Integration*, that allows access to multiple heterogeneous ontologies and can be realized either by merging ontologies into a single one, or by keeping ontologies separate [9]. Uniform access to multiple heterogeneous information sources is provided by knowledge integration systems, that are global schemas allowing a unified view for querying the set of local schemas. Generally, to find matchings between two ontologies, similarity measures are used, because they measure the degree of similarity between the ontologies. Some example of similarity measures that can be used during ontology matching are given in [28] and [26]. These include terminological methods that compare the labels of the entities/ concepts, using a syntactic approach with the help of lexicons such as Wordnet, etc. Other methods compare the internal structure of the concepts that are matched (e.g. the attributes, the interval of values); the relationships between them and other components of the ontologies, while more sophisticated methods compare the semantics, the interpretation or the model of the entities of the ontologies. Other strategies adopted to address the challenge of ontology mapping include machine learning [7], rule based mapping and logic driven frameworks. An updated survey of ontology mapping techniques is given in [15].

Semantic Web technologies have been used both for creating mashups of biomedical data [5], and for terminology integration purposes. For example, in [2] the Resource Description Framework (RDF) language is exploited for the comparison of formal definitions in LOINC and SNOMED CT. Much work has been done for the mapping of different Biomedical Ontologies with concept overlap (e.g. mapping between the anatomical concepts of the Foundation Medical Anatomy (FMA) ontology and the corresponding concepts in SNOMED [39], ICD10, etc.), and for their integration by means of medical ontology repositories (e.g. BioPortal[7], the web-based biomedical ontology repository developed at the Stanford University that includes more than 430 ontologies collecting their mappings by using a community based method [25]. One of the oldest attempts on medical ontology integration was the ONIONS methodology, proposed in [12], applied to some relevant medical terminologies (e.g. in the UMLS project), that is based on two main tasks: 1) development of a well-tuned set of generic ontologies to support the integration of relevant domain ontologies in medicine; and 2) explicit tracing of concept mappings, constraints and choices in ontology building in order to allow extensions and/or updating.

Semantic web technologies have been also leveraged for the creation of the multi-terminology platform HMTP (Health Multi-Terminology Portal) connected to the CISMeF Terminology database (the leading French health gateway to index Internet resources from the main institutions, using the MeSH thesaurus and the Dublic Core metadata element set) [19], further improved into the HETOP (Health Terminology/Ontology Portal) that includes the major health terminologies and ontologies available in French (or in English and translated in French), aiming to enable centralized access to them [14]. In order to integrate the terminologies included in HETOP authors followed an approach that can be synthesized in three main steps: i) design of a terminology generic model into which each terminology can be integrated; ii) design a process capable of integrating terminologies into the HMTP; iii) build and integrate inter-terminology mappings into the HMTP.

[7] http://bioportal.bioontology.org/

Mappings in particular were performed based on UMLS concepts and on NLP tools developed by the CISMeF team [20].

On the other hand, one of the most recent works in the field is the collaborative and formal editing of the International Classification of Diseases, 11th revision (ICD-11), coordinated by the Stanford Center for Biomedical Research (BMIR) at Stanford University. For this revision, a set of tools for collaborative ontology editing and publishing, in an integrated platform (composed of the tool I-CAT, based on Web Protégé, and the BioPortal repository that facilitates the integration between ICD concepts and other biomedical ontologies) for the creation of the entire ICD-11 ontology lifecycle [33, 34].

Other studies focus on the formalization and semantic verification of the existing mappings between medical coding systems in order to make explicit the formal references that precise the meaning of such mappings, typically defined on clinical bases. For example, in [4] authors proposed the formal representation of two coding systems which have received great widespread and preference within the European Union, ICPC-2 and ICD-10, providing the logical encoding of the existing ICPC-ICD conversion mappings in terms of OWL axioms and the verification of their coherence using logical reasoning; as well as the outline of other semantic techniques for automated analysis of implications of future mapping changes between the two coding systems. This study showed that the formal analysis of mappings and their clear logical encoding is an important prerequisite for the correct integration of different systems.

4 Existing Mapping between International Coding systems and between SNOMED CT and other terminologies

4.1 Mapping between international coding systems used in primary care and for reimbursement.

4.1.1 ICPC and ICD Mappings

A technical and official conversion exists between the International Classification of Primary Care (ICPC) and the International Classification of Diseases 10th release (ICD-10). As described in [38], this mapping has been carried out to allow primary care physicians to implement ICD-10 as a reference nomenclature within the classification structure of ICPC and leading to a substantial increase of the diagnostic potential of ICPC. This conversion was revised with the release of ICPC-2 and its electronic edition ICPC-2-E [27]. The mapping between these two coding systems are of three types: 1 to 1 mapping (a set of three-digit ICD-10 classes are compatible on a one-to-one basis with a three-digit rubric in the first or seventh component of ICPC); M to N mapping (a set of ICD-10 three-digit rubrics had to be broken open into four-digit-rubrics for at least one compatible conversion to one or more ICPC rubrics); 1 : N mapping (one ICPC rubric may be mapped to n ICD-10 (three or four-digit) classes and one ICD-10 (three or four-digit) rubric may be mapped to m ICPC rubrics.

In 2005 a bilingual Thesaurus (English and Dutch), namely the ICPC2-ICD10 Thesaurus[8], including the mappings between ICD-10 and ICPC-2 was developed at the Academic Medical Center at the University of Amsterdam and is now maintained by the Transition Project Foundation. The function of this thesaurus is to be used as a diagnostic terminology for semi-automatic double coding in EHRs which uses ICPC-2. In fact it maps the diagnoses found in ICPC-2, which only includes the more common diagnoses, to the wide range of diagnoses found in ICD-10. Its last official update was released in 2005, and in the same year it was included in the UMLS Metathesaurus. Furthermore, it also includes mappings between ICD-9-CM and ICD-10 provided by the World Health Organization (WHO), which, opportunely reviewed by physicians, can allow for a transcoding between ICD-9-CM and ICPC-2.

4.1.2 SNOMED CT and other terminologies/coding systems mappings

SNOMED CT is a terminology that provides explicit cross maps to other international health-related classifications and coding systems in use around the world, e.g. diagnosis classifications such as ICD-9-CM, ICD-O3, and ICD-10, as well as the OPCS-4 classification of interventions. Additional cross-maps are also under development or consideration. The existing cross-maps have been created to facilitate the reuse of SNOMED CT-encoded data for other purposes, such as reimbursement or statistical reporting.

SNOMED CT and LOINC.

LOINC has been used throughout the world for the coding of laboratory medicine test orders and result reports. Due to the use overlap of SNOMED CT and LOINC concepts and their joined use in healthcare information systems, a formal cooperative agreement between the Regenstrief Institute and the IHTSDO (parent organizations) was signed in July 2013. The scope was to align

[8] http://www.rivm.nl/who-fic/thesauruseng.htm

laboratory LOINC and SNOMED CT to minimize duplications and work better together. By means of the mapping between LOINC and SNOMED CT, people will essentially use them as one terminology, being able to query various laboratory tests and parts of tests in LOINC in a novel way based on SNOMED CT hierarchies and definitions. Benefits also include better identification and tracking of disease outbreaks through enhanced maintenance of significant test sets as well as valuable decision support for physicians when diagnosing and treating patients. In 2014, the alpha prototype was released, consisting of 117 LOINC terms mapped to SNOMED CT concepts. The final output of the project is expected to include over 45,000 laboratory LOINC terms mapped to SNOMED[9].

SNOMED CT and WHO classifications

IHTSDO and the World Health Organization (WHO) have been collaborating for years to map their resources thanks to two mapping projects, the “SNOMED CT to ICD-10” and the “SNOMED CT to ICD-9” Cross-map Projects. The purpose of the cross-maps is to support semi-automated generation of ICD-10 and ICD-9-CM codes from clinical data encoded in SNOMED CT for reimbursement and statistical purposes. As can be seen on the IHTSDO website, the SNOMED CT to ICD-10 and SNOMED CT to ICD-9-CM maps are annually released on 31 January and 31 July as derivatives of the SNOMED CT updates.

Furthermore, another project, coordinated by the US National Library of Medicine (US NLM), aims at mapping SNOMED CT to ICD-10-CM (US version). Even though the cited projects belong to different organizations, there has been a lot of synergism between them. In fact, the ICD-10-CM map project heavily re-used the mapping methodology, tools and map data developed in the SNOMED CT to ICD-10 cross-map project [36]. For this mapping the March 2014 release of the US Edition of SNOMED CT and the 2014 version of ICD-10-CM have been used. The Map is published primarily as four RefSets[10] in the SNOMED CT Release Format 2 (RF2) and can be used by users that are licensed to use both SNOMED CT and ICD-10-CM [22].

US NLM also provided the “ICD-9-CM Diagnostic codes to SNOMED CT Map” within the UMLS Metathesaurus, that is different from the IHTSDO and WHO “SNOMED CT to ICD-9-CM cross-map project”, and has the goal of facilitating the translation of legacy data and the transition to prospective use of SNOMED CT for patient problem lists. The International release of this mapping was published in 2011, and updated to 2014. As for the diagnostic codes, NLM provided an ICD-9-CM Procedural codes to SNOMED CT map, whose international release goes back to 2012 and is updated to 2013.

Finally, SNOMED CT also works within the foundation layer of ICD-11. In particular, the IHTSDO, as part of the agreement with WHO, is undertaking a fundamental task to ensure alignment between SNOMED CT content and that proposed in the ICD-11 Common Ontology. This alignment has the aim to ensure to clinicians the global use of SNOMED CT in the EHR for clinical information and the link to ICD-11 for information related to mortality and morbidity, billing, public health, epidemiology etc. To this aim specialists have been recruited to identify gaps in SNOMED CT in relation to ICD-11, including checking of clinical definitions and confirmation of specific classification terms[11] [IHTSDO clinical review SNOMED/ICD11 alignment, April 2015].

[9] http://www.ihtsdo.org/snomed-ct/mapping-to-other-terminologies/loinc

[10] A RefSet is a set of references to SNOMED CT components which may associate additional properties with members of the set. It may also indicate associations between members of the set or between members of the set and content of another nomenclature, classification or knowledge structure [22]

[11] http://www.ihtsdo.org/news-articles/ihtsdo-seeks-clinical-experts-for-snomed-ct-content-review-projects

Researchers, together with IHTSDO and WHO parties, have studied and proposed different approaches in order to align SNOMED CT and ICD-11 and validate their cross-utilisation, as shown in [29] and [30] where the ICD11 SNOMED CT harmonization process is presented, leveraging on a common ontology. Here they stated that, if SNOMED CT is considered as an ontology, the subset of SNOMED CT necessary to represent the ontological part of ICD-11 foundation component is the common ontology between SNOMED CT and ICD-11, that is, therefore, not an “add-on” to the mapping process, but its conceptual core [29: p.345].

SNOMED CT and ICPC-2

In 2009, the IHTSDO and the WONCA started a collaboration agreement that has led, in the last three years, to the “General/family practice SNOMED CT RefSet and ICPC-2 mapping project”, with the attempt to create mappings between SNOMED CT and ICPC and increase the potential of SNOMED CT for the subdomain of primary care. The main contributions of this project, with a full release in 2014, are: the construction of a specific SNOMED CT RefSet for General Practice and Family Medicine and the mapping from this RefSet to ICPC-2 [3] The mapping process followed in this project is similar to the one used to map SNOMED CT to ICD-10, and it involved 40 General Practitioners and Family Doctors from 8 countries.

SNOMED CT and nursing terminologies

Finally, progress has been made also in the integration of SNOMED CT to terminologies oriented towards nurses. In fact, in January 2014, the IHTSDO and the International Council of Nursing (ICN) announced the release of an equivalency table between pre-coordinated diagnostic concepts in the International Classification for Nursing Practice (ICNP), that is already mapped to NANDA (North American Nursing Diagnosis Association) taxonomy, one of the most used terminologies for the nursing subdomain, and clinical findings concepts in SNOMED CT[12]. This project is expected to bring two contributions: 1) the addition of ICNP nursing problems that do not exist in SNOMED CT; and 2) the creation of the nursing problem RefSet.

The example presented earlier for the categorial structure [13] mapped ICNP data to ICF data and the Nursing Minimum Data Set for the Netherlands (NMDSN). A limited set of about 24 nursing diagnostic statements were cross-mapped using the dissection and inter-rater agreement (Kappa calculation) methods. It illustrated that depending on the kind of use, for instance clinical, continuity of care, multidisciplinary care, or aggregate levels as quality measures or policy making, the accuracy of the mapping has a different impact. Since aggregation already encompasses selection and elimination of details, it is well possible that a mapping from fine grained to course is acceptable for statistical representations, where for the details of clinical care, that could potentially lead to errors due to lack of precision. It is therefore important to have the actual purpose in mind when applying methods to determine consistency and correctness of mappings.

[12] https://3mhealthinformation.wordpress.com/2014/04/07/icnpsnomed-ct-nursing-terminology-collaboration/

5 Conclusions

In this annex the problem of the mapping between international medical terminologies, coding systems, ontologies, and to some extend between clinical data models, has been treated. The reviewed studies have shown that, with the growing complexity of medical information and the increasing need to completely and correctly exchange clinical data among the different healthcare information systems, the huge amount of existing medical terminologies and standards not only need for a more formal representation for an explicit specification of the meaning of the concepts, but above all for the definition of unambiguous mappings between them in order to create precise correspondences between the concepts expressed in one terminology/ontology and the concept found in the other terminology/ontology and/or data model. In fact, the use of ontologies is to make explicit the meaning of terms used in healthcare information systems in order to resolve semantic heterogeneity. However, as seen in the present annex, the diversity of ontologies used in the healthcare domain increases semantic heterogeneity to a higher level and this can be resolved only by finding matches or correspondences between heterogeneous ontologies.

Thus, the definition of mappings and the integration of heterogeneous medical terminologies or ontologies is fundamental to guarantee Semantic Interoperability between healthcare information systems. This is possible in different scenarios: during exchange of clinical data encoded with different standards from one system to another; during data registration, for automatically coding or double coding clinical data and documents in electronic health records; and finally, when searching for healthcare information (e.g., facilitating automated mapping of queried medical terms). For all mappings applied, the actual goal of data use, and hence the level of granularity, must be taken into account to guide decisions on the accuracy of mappings.

References

1. Berners-Lee T., Hendler J., and Lassila O.. The Semantic Web. Scientific American, 284(5), 34—43, 2001.
2. Bodenreider O.. Issues in Mapping LOINC Laboratory Tests to SNOMED CT. In AMIA Annual Symposium, AMIA2008, pp. 51—55, 2008.
3. Booth M., Miller G., O'Halloran J. General/family practice SNOMED CT RefSet and ICPC-2 mapping project: current status and next steps. Technical Report.
4. Cardillo E., Eccher C., Serafini L., Tamilin A.. Logical Analysis of Mappings Between Medical Classification Systems. In Artificial Intelligence: Methodology, Systems, and Applications, 13th International Conference, AIMSA 2008, Springer, vol.5253, 2008, pp. 311-321, (Proceedings of Artificial Intelligence: Methodology, Systems, and Applications, 13th International Conference, AIMSA 2008, Varna, Bulgaria,2008).
5. Cheung K. H., Kashyap V., Luciano J., Chen H., Wang Y., and Stephens S.. Semantic mashup of biomedical data. Journal of Biomedical Informatics, 41(5):683–686, 2008.
6. De Keizer N.F., and Abu-Hanna A.. Understanding Terminological System II: Experience with Conceptual and Formal Representation of Structure. Methods Inf Med, 39(1):22–29, Mar 2000.
7. Doan A., Domingos P., Halevy A.. Learning to Match Schemas of Data Sources: A Multistrategy Approach. In Machine Learning, Kluwer Academic Publishers (The Netherland), vol. 50, pp. 279–301, 2003.
8. Elkin P.L., and Brown S.H.. Automated enhancement of decryption logic-defined terminologies to facilitate mapping to ICD9-CM. Journal of Biomedical Informatics, 35:5–6, 2002
9. Euzenat J. and Shvaiko P.. Ontology Matching. Springer, Berlin / Heidelberg, 2007.
10. Foley M. M.. Key issues shaping clinical terminology and classification. Journal of AHIMA, 77(7), July/August 2006. extended online edition.
11. Fung K.W. and Bodenreider O.. Utilizing the UMLS for Semantic Mapping between Terminologies. In Proceedings of AMIA Annual Symposium 2005, pages 266–270, 2005.
12. Gangemi A., Pisanelli D. M., and Steve G.. Ontology Integration: Experience with Medical Terminologies, pages 163–178. IOS Press, 1998.
13. Goossen W. Cross-mapping between three terminologies with the international standard nursing reference terminology model. Int J Nurs Terminol Classif. 2006 Oct-Dec;17(4):153-64.
14. Grosjean, J; Merabti, T; Griffon, N; Dahamna, B; Soualmia, LF & Darmoni, SJ. Multi-terminology cross-lingual model to create the Health Terminology/Ontology Portal. In Proceedings of AMIA Annual Symposium (AMIA2012), Chicago, 2012
15. Hooi Y. K, Hassan M. F, Shariff A. M.. A Survey on Ontology Mapping Techniques. In Advances in Computer Science and its Applications, Lecture Notes in Electrical Engineering, Springer, Volume 279, pp. 829-836, 2014.
16. Iroju O., Soriyan A., Gambo I.. Ontology Matching: An Ultimate Solution for Semantic Interoperability in Healthcare. In International Journal of Computer Applications (0975 – 8887), Volume 51– No.21, August 2012.
17. van der Kooij J, Goossen WT, Goossen-Baremans AT, de Jong-Fintelman M, van Beek L. Using SNOMED CT codes for coding information in electronic health records for stroke patients. Stud Health Technol Inform. 2006;124:815-23.

18. Lomax J., and McCray A.T.. Mapping the Gene Ontology into the Unified Medical Language System. Comparative and functional genomics, 5(4):354–361, 2004.
19. Merabti, T; Grosjean, J; Abdoune, H; Joubert, M & Darmoni SJ. Automatic methods for mapping Biomedical terminologies in a Health Multi-Terminology Portal. In Proceedings of EGC 2011: atelier Extraction de Connaissances et Santé, 2011.
20. Merabti T, Massari P, Joubert M, Sadou E, Lecroq T, Abdoune H, Rodrigues JM, Darmoni SJ. An Automated Approach to map a French terminology to UMLS. In Proceedings of MEDINFO 2010.
21. Mougin F, Dupuch M., and Grabar N.. Improving the Mapping between MedDRA and SNOMED CT. In Artificial Intelligence in Medicine, Lecture Notes in Computer Science, Springer, Volume 6747, pp. 220-22, 2011.
22. National Library of Medicine – NLM. SNOMED CT to ICD-10-CM Map, Release Notes, v.4, 2014.
23. Nyström M, Vikström A., Nilsson G. H., Åhlfeldt H., and Örman H.. Enriching a primary health care version of ICD-10 using SNOMED CT mapping. In Journal of Biomedical Semantics 2010, 1:7 doi:10.1186/2041-1480-1-7
24. Noy N. F.. Ontology Mapping, pages 573–590. 2009.
25. Noy N. F., Musen M., Shah N., Dai B., Dorf M., Griffith N., Jonquet C., Montegut M., Rubin D., and Youn C.. Bioportal: A web repository for biomedical ontologies and data resources. In Proceedings of ISWC2008, Karlsruhe, Germany, 2008.
26. Ofoghi B., Lopez-Campos G. H., Martin Sanchez F. J., and Verspoor K.. Mapping Biomedical Vocabularies: A Semi-Automated Term Matching Approach. In Studies in Health Technology and Informatics, Volume 202: Integrating Information Technology and Management for Quality of Care, pp. 16 – 19. 2014. doi=10.3233/978-1-61499-423-7-16.
27. Okkes I.M., Jamoullea M., Lamberts H., and Bentzen N.. ICPC-2-E: the electronic version of ICPC-2. Differences from the printed version and the consequences. Family Practice, 17:101–107, 2000.
28. Rahm, E., and Bernstein, P. 2001. A Survey of Approaches to Automatic Schema Matching. VLDB Journal, pp. 334 – 350.
29. Rodrigues JM, Schulz S, Rector A, Spackman K, Üstün B, Chute CG, Della Mea V, Millar J, Persson KB. Sharing ontology between ICD 11 and SNOMED CT will enable seamless re-use and semantic interoperability. In Stud Health Technol Inform. 2013;192:343-6.
30. Rodrigues JM, Schulz S, Rector A, Spackman K, Millar J, Campbell J, Üstün B, Chute CG, Solbrig H, Della Mea V, Persson KB. ICD-11 and SNOMED CT Common Ontology: Circulatory System. In e-Health – For Continuity of Care. Studies in Health Technologies and Informatics, vol. 205, pp. 1043-1047.
31. Semantic Health Report 2009. Semantic Interoperability for Better Health and Safer Healthcare. European Commission, Information Society and Media.
32. Stefan S., Holger S., Martin B., and Barry S. Strengths and Limitations Of Formal Ontologies In The Biomedical Domain. Electronic Journal of Communication Information and Innovation in Health. pp. 31-45, 2009.
33. Tudorache T., Nyulas C. I., Noy N. F., Redmond T., and Musen M. A.. iCAT: A Collaborative Authoring Tool for ICD-11. In Proceedings of the ISWC 2011 Workshop Ontologies Come of Age in the Semantic Web (OCAS), CEUR-WS Vol-809, Bonn, Germany, October 24th, 2011.

34. Tudorache T., Nyulas C. I., Noy N. F., Musen M. A.. Using Semantic Web in ICD-11: Three Years Down the Road. In The Semantic Web – Proceeding of ISWC 2013, Lecture Notes in Computer Science, Springer, Volume 8219, pp. 195-211, 2013.

35. Wade G. and Rosenbloom S.T.. Experiences mapping a legacy interface terminology to SNOMED CT. In BMC Medical Informatics and Decision Making 2008, 8(Suppl 1):S3 doi:10.1186/1472-6947-8-S1-S3.

36. Wah Fung K., Xu J.. Synergism between the Mapping Projects from SNOMED CT to ICD-10 and ICD-10-CM. In Proceedings of AMIA Annu Symp 2012; pp. 218–227, 2012.

37. Wang Y., Patrick J., Miller G., and O'Halloran J.. Linguistic Mapping of Terminologies to SNOMED CT. In Proceedings of SMCS 2006, 2006.

38. Wood M., Lamberts H., Meijer J.S., and Hofmans-Okkes I.M.. The Conversion Between ICPC and ICD-10. Requirements for a Family of Classification Systems in the Next Decade. Family Practice, 9:340–348, 1992.

39. Zhang S., and Bodenreider O.. Experience in aligning anatomical ontologies. International Journal on Semantic Web and Information Systems, 2007, 3(2):1-26.