Medical Informatics Europe 2002
G Surján et al. (eds)
IOS Press 2002

Co-ordination between clinical coding systems and pragmatic clinical terminologies based on a core open system: the role of ISO/TC215/WG3 and CEN/TC251/WG2 standardisation?

Jean Marie RODRIGUES, Beatrice TROMBERT PAVIOT, Caroline MARTIN, Paul VERCHERIN, Olivier SAMUEL

Department of Public Health and Medical Informatics, Faculté de Médecine Université Jean Monnet, 25 rue Ambroise Paré, F 42 000 Saint Etienne France, Email rodrigues@univ-st-etienne.fr.

Abstract. The article addresses the need to co-ordinate efforts to develop clinical coding systems and pragmatic clinical terminologies like SNOMED CT. In the first part, a description is given of the current context of divergent and replicated efforts. Then is presented a "reference terminology representation" approach based on a formal terminology representation as an open source available in the public domain with diversity in the linguistic expressiveness of end users let to competing developers and researchers. The last part is devoted to the contribution of the standardisation process in healthcare terminology initiated by CEN/TC251 and supported now by the work of CEN/TC215/WG3 to this new approach which can be summarised as the practical realisation of an ontology.

1. Introduction

The late developments in healthcare informatics in Europe and elsewhere stressed as the first priority the electronic health care record and the exchange of clinical data. These developments would be much easier if broadly shared coding systems and pragmatic clinical terminologies were available in the different national languages. Unfortunately the historical trend of development of coding systems in healthcare and of pragmatic clinical terminologies are going on in many different countries and national languages. The present outcome is still a Tower of Babel that requires healthcare professionals to enter the same clinical information several times for different purposes. This prevents comparison of data held in archive "cemeteries" of stored clinical data (including case mix). It also prevents reuse of the knowledge so expensively produced in one place, in another place, particularly in countries less technologically advanced in healthcare information.

We have proposed in [1] a new approach called reference terminology representation approach based on an open source formal representation we explain in the second part and we review the contributions of standardisation efforts in Europe (CEN/TC251) and the world (ISO/TC215) to this new approach.

2. Context: divergent and replicated efforts

2.1. Specialised clinical coding systems

In many western countries there are huge, parallel, manpower efforts to develop clinical coding systems for procedures, lab tests, devices etc: e.g. Health Procedures (Australia, Canada, France, Germany, Nordic Countries, USA LOINC, CPT5 and ICD-10-PCS). Unfortunately, these efforts are purpose related (e.g. payment of fees, casemix, benchmarking, safety etc), country specific, sometimes clinical specialist college specific and language dependent. They cannot support the reuse of stored clinical data for another purpose by another country, another clinical specialist college or in another language. These data resources become what are called "data cemeteries", and consequently there is little return on investing in the maintenance of the clinical coding systems.

This has resulted, for specific coding systems (e.g. surgical procedures) for some countries (small western European countries, developing countries, central and eastern European countries) either in the decision not to produce them and to implement translated coding system or to restrict clinical information applications to particular countries or specialist colleges. Even when produced, the resources required for annual updating/translating programmes can be disproportionate to the size of a country or to a European/International college of specialists (e.g. chronic renal failure, endoscopy etc).

2.2. WHO International Classifications Family

For pathology and functional status (and very soon primary care) a *de facto* standard, the WHO International Classifications Family (ICD, ICF and later ICPC 2), is being maintained and extended by traditional manual means. This is a constant high-cost maintenance and revision cycle accumulating endless, and not well connected, scattered updates. However there is insufficient manpower available. As a consequence, this coding systems family even for the most achieved development (ICD-10-AM Australia) cannot support seamless use by real-world healthcare professionals, most of whom require a concurrent pragmatic clinical terminology coding system for day-to-day clinical management.

2.3. Pragmatic clinical terminologies

Specialised and large-scale clinical controlled vocabularies (often called clinical terminology) among which the most comprehensive and consistent can be expect to be SNOMED Clinical terms [2] show limits to support comparison, multipurpose use, and consistent and easy maintenance.

2.4. Unconnected third-generation tools

Recent developments of third-generation tools have been carried out within unconnected organisations, continents and legacies: UMLS (National Library of Medicine) and GALEN (University of Manchester) are public domain, while SNOMED RT is proprietary (American College of Pathologists).

3. The reference terminology representation approach

3.1. Definition

Following the ISO definition [3], a reference terminology representation (RTR) is the set of canonical concepts, their structure, relationships and, if present, their systematic and formal definitions. These features define the core of a controlled health terminology.

The difference between a reference terminology representation and a pragmatic reference terminology should be noted. A reference terminology representation is a logical system of concepts — a core model used by the artificial intelligence of a computer. A pragmatic reference terminology is a well-formed terminology aiming at coherence and consistency by using a controlled vocabulary understandable by healthcare professionals. To stress the difference between the flexibility of natural language versus exactness of formal representation see the example in table 1.

Table 1: Example from laparoscopic cholecystectomy (ISO/DTS 17117: 2000)

Intermediate dissection
MAIN excising
ACTS_ON gallbladder
BY_APPROACH TECHNIQUE inspecting
ACTS_ON peritoneal cavity
BY_MEANS_OF laparoscope
WITH_GUIDANCE_BY laparoscope
Generations in natural language
Laparoscopic cholecystectomy
Excision of gallbladder by the endoscopy method
Excision of gallbladder by the endoscopy approach
Excision of gallbladder, using an endoscope device.
Excision of gallbladder, using a laparoscope device.
Excision of gallbladder under control of a laparoscope and has approach inspection of the peritoneal cavity by means of a laparoscope.

3.2. The recommended approach

It is clear that the expressiveness of natural language is highly variable and cannot be normalised even within just the English language. This is even truer when you consider the different natural languages. For these reasons, we advocate in [1] both a convergent and diversity approach. The rationale is that we believe it is possible to achieve only convergence for a health reference terminology representation (concept model) and not for a pragmatic reference terminology translated in different natural languages: this convergence shall be towards an open-source *de facto* standard. Only convergence is possible because it is necessary to allow diversity in natural language expression of diverse clinical and linguistic groups. The terminology representation must be in the public domain but any developer must be allowed to produce proprietary software from the open source.

3.3. The core formal representation organisation

We propose to co-ordinate the approach following the organisational model implemented in Galen In Use [4][5] for the development of CCAM, the new French coding system for surgical procedures [6]. The knowledge acquisition proceeds in four steps:

- from the linguistic expression to an intermediate representation by a modelling centre that must be a clinical terminology facility devoted to the domain but not an artificial intelligence research centre;
- from the intermediate representation to the reference terminology representation within the representation manager centre, which is the facility with artificial intelligence researchers;
- from the reference terminology representation to a controlled linguistic expression using natural-language generator processing software available in one research centre;
- from the controlled linguistic expression a cross validation with the initial linguistic expression by the clinical terminology facility to assure the quality of the reference terminology representation.

4. Health care information standardisation

Since 1990 there is a standardisation process in healthcare information initially in Europe (CEN for Comité Européen de Normalisation) and lately since 1998 in ISO which supports quality assessment of, and guidelines to structure, controlled vocabularies, nomenclatures and classifications. The strategy of the process has been defined in CEN by the work on model for representation of semantics represented by the acronym MOSE [7]. It is based on the definition of the categorial structure of a system of concepts which is a "reduced system of concepts made of semantic categories, which can point out the most significant regularities that can be exploited in the system of concepts analysed" [7]. It is an incomplete reference terminology representation following the ISO definition [3]. There are 5 information types required:

a) a list of the relevant sets of concepts
b) a list of the semantic categories
c) a list of relevant semantic links
d) a list of associated semantic categories
e) the combinatorial rules that allow generation or verification of well-formed concepts.

A categorial structure may be used as the basis to prepare classifications or compositional and co-ordinated systems or nomenclatures with systematic names. The terms "ontology", "epistemology" and "typology" are in use in different communities with various meanings. These terms appear to overlap sometimes with the term "categorial structure", as defined by the European Pre-standard.

This methodology has been applied within CEN to surgical procedures coding systems [8] and within CEN and ISO to nursing diagnostics and nursing actions [9] [10]. Works are ongoing for representation of conditions in Classifications, Coding systems and Clinical terminologies [11], clinical laboratory [12] and continuity of care [13].

It is the base of the core open system organisation recommended in [1].

5. Conclusion

Widespread electronic recording and transmission of clinical data look like the main challenges for health informatics of the beginning of the century. There are still terminology works to do to insure unambiguous comprehension, accuracy and safety. Without such achievements most of the opportunities given by technical advancement in IT technology would stay helpless.

The main issues for the coming years in clinical terminology will be knowledge acquisition by cross validation and open availability.

There is a crucial need to co-ordinate between clinical coding systems centres both at national and WHO collaborative centres levels, researchers, pragmatic clinical terminologies proprietary developers, CEN and ISO standardisation experts.

Who is going to be the starter partner?

References

[1] Rodrigues J.M,Trombert Paviot B and Martin C. *Is there a need for co-ordinated international efforts for clinical coding? Br J Healthcare Comput Info Manage* 2001; 18(10):16-19.

[2] Iggulden P,Price C. *SNOMED Clinical Terms – Looking forward to the next phase Br J Healthcare Comput Info Manage* 2001; 18(10):20-22.

[3] ISO/TC215/DTS 17117: 2000. *Controlled health vocabularies — vocabulary structure and high-level indicators.*

[4] Rector AL, Baud R, Ceusters W, Claassen W, Rodrigues J-M, Rogers J, Rossi Mori A, van der Haring E, Solomon WD, Zanstra P. *A comprehensive approach to developing and integrating multilingual classifications: GALEN's classification workbench.* Theater-style demonstration. Presented at the AMIA 1998 Annual Symposium (AMIA'98). Lake Buena Vista, Florida, USA.

[5] Trombert-Paviot B, Rodrigues J-M, Rogers, Baud R, van der Haring E, Rassinoux A-M, Abrial V, Clavel L, Idir H. *GALEN: a third-generation terminology tool to support a multipurpose national coding system for surgical procedures.* International Journal of Medical Informatics 2000; 58–9: 71–85.

[6] http://www.cnamts.fr/

[7] CEN/TC251/ENV 12264:1996. *Medical informatics - Categorical structures of systems of concepts - Model for representation of semantics*

[8] CEN/TC251/prEN1828:2001. *Categorial structure for classifications and coding systems of surgical procedures*

[9] CEN /TC 251/ENV14032 2000, *Health Informatics – System Of Concepts To Support Nursing*

[10] ISO/TC215/099/WD. *Integration of a Reference Terminology Model for Nursing.* London 2001

[11] CEN/TC251/WI 2001. *Categorial structure for representation of conditions in Classifications, Coding systems and Clinical terminologies*

[12] CEN /TC 251/ENV1614 1995, *Medical Informatics – Structure for nomenclature,classification and coding of properties in clinical laboratory sciences*

[13] CEN /TC 251/ENV13940, *Health Informatics – System Of Concepts To Support Continuity of care*