Zanstra P, Rector A, Ceusters W, de Vries Robbé P. Coding systems and Classifications in Healthcare: the link to the record. International Journal of Medical Informatics 1998;48:103-109

Coding systems and Classifications in Healthcare: The link to the record

PE. Zanstra[1], AL. Rector[3], W. Ceusters[2], PF de Vries Robbé[1]

(1) *Department of Medical Informatics and Epidemiology, University of Nijmegen, PO Box 9101, 6500 HB NIJMEGEN, The Netherlands*
(2) *Office Line Engineering NV, Het Moorhof, Hazenakkerstraat 20, B-9520 Zonnegem, Belgium. Tel: +32 53 62 95 45.*
(3) *Medical Informatics Group, Department of Computer Science, University of Manchester, Oxford Rd, Manchester M13 9PL, UK*

Introduction

Currently we are entering the third phase of introducing information technology to the health care delivery system. The first phase was mainly concerned with automating the administrative process. The second phase was about delivering medical applications. The third phase is about integrating a diversity of medical and administrative systems in one coherent inter-operable environment.
As a result we see an emerging need for safe and sensible communication between these applications. Next to a good insight of what information is really necessary to exchange, also the language used is important. For instance, the slow progress in the implementation of knowledge based systems may to a great deal be attributed to the lack of semantical coupling with the patient record.
In these writings where possible we carefully avoid the adjective 'electronic' while dealing with the Healthcare Record. The fact that a record is stored by some electronic means is not distinguishing. We can make a perfect electronic image of the paper record, as we also can make a perfect 1:1 image of a paper-based classification. The point is we now no longer are constrained by the paper format. There is an opportunity to give the record a formal basis, leading to a number of advantages w.r.t. manipulating the stored data. It would therefor be more appropriate to talk about the Formalised Health Care Record (FHCR).
The urgent needs are reflected in the many projects around the Computerised Medical Record (GEHR, MARGOT etc.), projects for a unified medical language (UMLS[15], GALEN, CANON, Convergent Terms project, etc.). Because of the sheer size of medicine all these projects are necessarily limited in their scope. There is a need for a broad view on the further development of medical language in a computational linguistic perspective. The cost of researching and developing clinical systems is too great to be borne by a single national market. Whilst much re-use of system components is feasible there are significant costs associated with the 'localisation' of systems to the needs of a particular market. Perhaps the most important of these costs is the localisation to the linguistic needs of each national market. Medicine is a descriptive, language intensive activity, and the costs of developing, and perhaps more importantly maintaining, the linguistic resources needed to localise clinical systems are clearly high. This presents a genuine barrier to the development of systems for use in the global marketplace. And not only for systems, but more general a common ‘language’ is of prime importance to communication, documentation, education, evaluation, comparison etc. Any practical approach to the management and exploitation of linguistic resources in large scale clinical information systems must be based on common methods and internal representations for linguistic information. This information must be reusable across a wide range of systems and local variants of those systems, and the cost of maintaining that information must be separable from those of maintaining the rest of the system.
Significant regional differences in linguistic usage exist even within single languages, and even more so when minority languages are taken into consideration. To be truly successful, a programme of linguistic engineering in medicine must have a strategy for recognising and managing regional as well as national linguistic differences. These considerations further complicate and increase the expense of 'localisation' of products for the global market.

Market situation and prospects

There is an increasing awareness of the crucial role of a 'common medical language' for the further development of medical information systems. The time of 'black box thinking' producing statements as

'Natural Language Processing' will solve that are over. Also NLP needs strong and large domain models in order to be effective and reliable [3]. Introduction of knowledge based systems is severely hampered by the lack of common terminology [11,14]. A major conclusion from the 1995 SCAMC conference was that the lack of a common medical terminology is the major obstacle for further development of Medical Informatics applications. From a European point of view this is at first sight a bit surprising, as the USA is not facing the diversity of languages and cultures as found in Europe. A few months after SCAMC in Amsterdam the AMICE conference with the theme 'Strategic Alliances between patient documentation & Medical Informatics' was held. One of the intentions was to bring together the different disciplines. Unfortunately this did not really happen. That was not due to ill organisation of the conference, but rather because of the groups speaking different ‘languages’. We hardly observed participation of people concerned with the medical record in discussions about terminology. Though not as outspoken, a similar situation was observed at Medinfo 95 in Vancouver.
Traditionally classification systems have been developed for a number of quite different purposes:

- Statistics (ICD, etc.)
- Nursing (NANDA, etc.)
- Case Mix (DRG, etc.)
- Reimbursement (ICPM, CADM, etc.)
- Epidemiology (ICPC, etc.)
- Literature retrieval (UMLS)
- Quality of Care
- Protocols
- Medical Records (Read, SNOMED, Gabrielli)

Though from the above list it is clear that there are standards for some elements of the Medical Record specifically elements that characterise the patient, such as signs and symptoms are hardly standardised.

The business view

If to be used in automated systems, terminology development can certainly not be seen in isolation. First of all it requires insight in the business process of the health care organisation. How does information flow through an organisation, and between organisations. So we have to address the problem of what is generally put under the umbrella of EDI. Secondly for documentation we need storage structures. As stated in the introductory lines there is a lot of activity at present in the area of structuring the Electronic Healthcare Record. There clearly is a relation between terminology and the record. There is however grey area, where it is not clear where it belongs. Thirdly we need the terms and a delivery mechanism. Below we will address these items shortly, just to point out some views on the matter.

Observations with respect to the business process and EDI

Research and developments concerning electronic data interchange (EDI) in Health Care are important issues. Major questions are: Which type of technology is appropriate in applying EDI in Health care; Which type of interchange format should best be used; How to determine the contents of 'messages' that are to be interchanged. Expected improvements using EDI are (amongst others): Increasing efficiency and reliability in data-interchange; Availability of information to users in the time and place that is needed; Cost saving effects as a result of the elimination of manual tasks and reduced delays.
One of the articles of the Maastricht Treaty on European Union (1992) says: ' the Community shall identify projects of common interest and shall implement any measure that may prove necessary to ensure the interoperability of networks, in particular in the field of technical standardisation'. We have to ask ourselves what is meant by 'interoperability' and 'technical standardisation'. We will therefor elaborate on the subject of EDI. It must be realised that the observations below not only apply to EDI in the classical sense of messages exchanged between different physical locations over electronic networks, but that the same principles hold true for integrating two applications which may even reside in one single computer!
Solving problems concerning the application of information technology and interchange formats in Health Care appears to be a complicated and time-consuming matter. This might account for the pragmatic approach concerning the contents of messages that should be exchanged. Existing 'sender determined' message definitions have been implemented. For example, the (national) guidelines [1] for putting together a 'standard' referral letter for (Dutch) General Practice to a clinical specialty contain an extensive list of data-items that should or might be added to this letter. This list of items has been determined by General Practitioners, without hardly any discussion with clinical specialists concerning

their information needs. Indeed, little attention has been paid so far to the effectiveness problem and semantic problem. The effectiveness problem deals with the following question: How to assure that interchanged data contribute to a more effective performance of Health-Care processes. The semantic problem addresses the relation between data on the one hand and the meaning of the concepts they are referring to on the other.

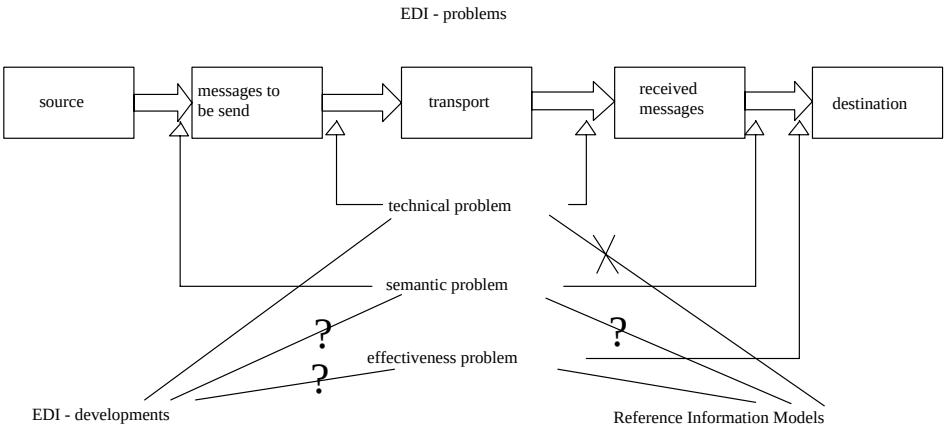

Figure 1: The EDI problem: an integrated view on the levels of meaning, business process, and technical interfacing. The world of EDI has paid little attention so-far to the semantic and effectiveness problems. These principles are not only valid for 'classical' EDI between sites, but also applicable for system integration

Modelling activities aimed at the specification of data-interchange format independent models are becoming of major importance in the field of EDI. Large so called 'Domain information models' (e.g. [5]), which are in fact data models, are (becoming) available. They contain the results of user requirements specifications by domain experts. These models consist of the description of activities and data, their attributes and the relationships between these entities that are relevant to the domain. They could serve as the main source for composing the contents of messages.

Relation between the Health Care Record and Terminology

Medical records use medical concepts to describe patients and their treatment. Medical concept systems supply the basic units of information for medical record systems. The record systems provide the structure for linking the concepts; the concept systems provide the content held by the medical record — whether those concept systems are in the form of traditional 'coding system', in some form of free text, or using the GALEN's CORE Model built in its concept representation language GRAIL.
Demands for more detailed computer-based medical records and demands to use computer-based medical records as a central store of information to be used for many different purposes have brought with them the requirement for more effective representation of medical concepts. Traditional systems have used coding and classification systems or free text. Coding systems are overly rigid and contained too little detail to support clinical care. Free text, on the other hand, is difficult to analyse and impossible to control sufficiently to make sound aggregation of data possible. GALEN's approach to concept modelling offers a third approach, allowing much more flexibility and detail than traditional coding systems but retaining the structure required for retrieval and controlled data entry.
In most situations, the data model for the medical record — or medical information system — will be separate from the model of the concepts used by that model. There are organisational, practical and technical issues behind this separation. Organisationally and historically, different groups are responsible

for the implementation of medical records and medical information systems and for the compilation and standardisation of terminology, and coding systems
The standard solution is to separate the models of medical concepts — 'the codes' in most existing systems — from the data model of the medical record which represents how those codes are held and combined.

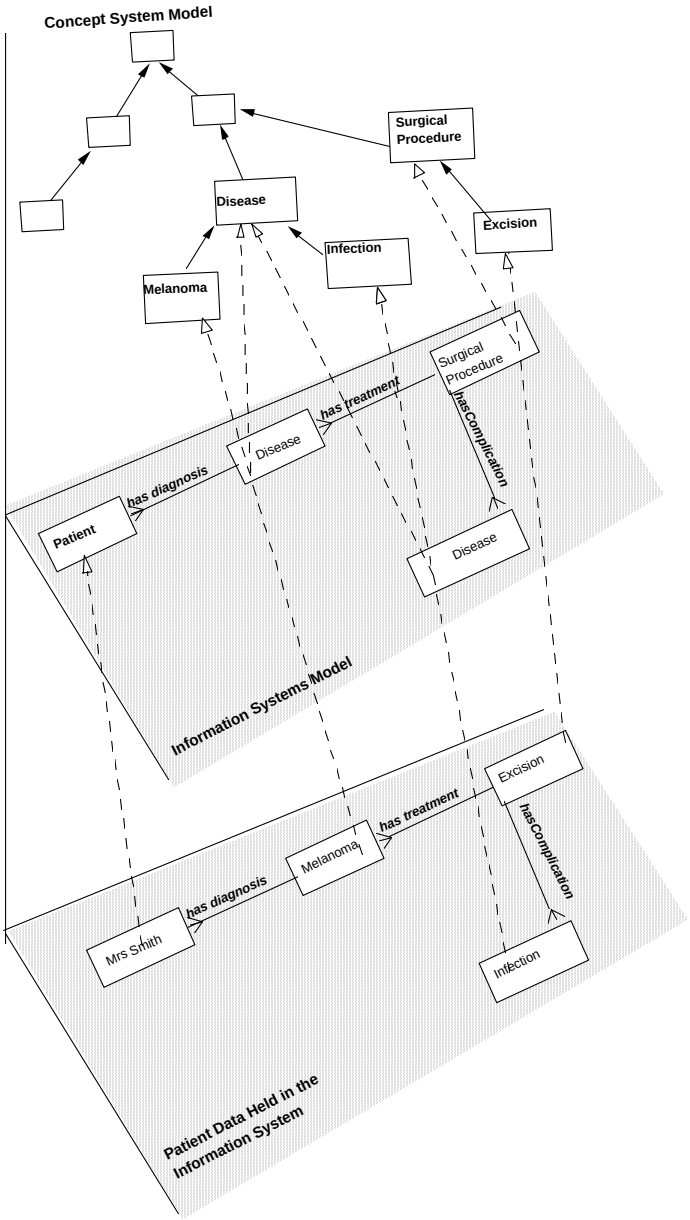

Figure 3 The relationships between from top to bottom: the system of concepts (classification), the data model of the medical record, and finally the instantiation of the record itself.

GALEN anticipates that most applications will continue to maintain the separation between the 'codes' or concept model and the data model — the difference will be that increasingly 'the codes' will be referenced back to the GALEN CORE Model allowing for a much richer and more consistent structure and much greater interconvertability than is now possible. It is a matter of discussion whether the concept system should have a notion of the meaning of links like hasDiagnosis, hasTreatment, hasComplication.

How do we move forward?

The European Federation of Classification Centres (EFCC) and PROREC have started, with funding from the European Healthcare Telematics Programme, the ToMeLo project. The main objective of

ToMeLo is to bridge the gap between developers of healthcare terminologies and healthcare information systems. As such, ToMeLo is an important step in realising the overall objective of the Health Telematics Programme: improving continuity of care through the integration of a diversity of medical and administrative systems in one coherent interoperable and multilingual environment.
There is a need for a broad view on the further development of medical language not only from the point of view of grouping (ICD, DRG's etc.), but above all from the perspective of patient documentation and systems integration. ToMeLo is therefor amongst others organising a series of workshops and and a conference on the subject. One of the objectives is a long term lasting co-operation between actors. At several places we now see this happen. The European Union is about to fund the Synex project, which is to integrate systems for documentation and links to knowledge sources in connection with terminology servers. The OECD has expressed interest in the GALEN approach to a set of compatible European classifications for medical procedures because they believe procedures are a better basis for health statistics than the recorded diagnoses, which is their present source. The Swedish SPRITerm server is linking up with the GALEN Common Reference Model, to provide proper semantically founded record structures.
ToMeLo is bringing the worlds of healthcare systems developers and medical terminology builders closer to each other. This is realised by setting up a number of workshops in which influencial representatives of both worlds will be confronted with each others and own problems. Currently, both worlds are mostly not aware of the intrinsic difficulties they are dealing with. ToMeLo is making the issues overt, as has been shown at the first workshop. The issues discussed in the workshop were: a)'Headings'/'Context', b) Multilingual systems and Localisation, c) Multi-professional systems, d) Linking to Knowledge Sources, e) Mediation, f) Evolutionary pathways.
Under the issue 'Headings' are: *Modalities* -Family history, past medical history; *Information target* - fetus, sample; *Health care role* - diagnosis, problem, episode, plan; Pragmatic - Physical examination, investigations; *Provider* - Nurse, doctor, physio; *Truth status* - true, false, uncertain; *Complexes* - full blood count, BP, With Mediation we mean how information including 'Headings' is transferred between systems. 'Headings' and 'Mediation' are tightly interrelated issues.
The major workshop outcome was that the issue of 'Headings'/'Context' has been placed clearly on the agenda as the most urgent topic for further work. The issue is not new, others have worked on it as well. Important is the recognition that the issue can not be resolved by any single party, but that is has to be collaborative work. These recommendations have been fed forward to the second EU-CEN workshop on the Medical Record, which was organised in May 1997 in conjunction with ToMeLo in Porto Carras,Greece.

Key Recommendations to EU/CEN community:

- 'Headings'/ 'Context' ripe for development;
- Needs cooperation between terminology and architecture groups;
- Need for a central Reference Resource of Concepts; Multilingual lexicons; Language independent development tools
- Communication architectures should address mediation & conversion

Acknowledgements

This work has been funded by the Health Care Telematics programme projects Tomelo (HC3108) and GALEN-In-USE (HC1018) of the European Union.

Literature:

[1] Nederlands Huisartsen Genootschap (Dutch college of General Practitioners). NHG Standaarden. Richtlijnen voor het handelen en de praktijkvoering van de huisarts. (DCGP Standards. Guidelines for actiontaking and practising of General Practitioners). Utrecht: NHG 1989.
[2] Linnarsson R. The Swedish Medical Terminology Database. An Electronic Presentation of Healthcare Terminologies: Some Terminology Browsers. MEDINFO 92. (Scientific Demonstration).

[3] Baud RH, Rassinoux A-M, Scherrer JR. Natural Language Processing and Medical Records, in Lun KC, Degoulet P, Piemme TE. Rienhof O (eds) Proceedings of MEDINFO 92, North-Holland Publishers, pp 1362-1367.

[4] Antti-Poika M, GALEN, in Proceedings of AIM-CEN European Workshop on the Medical Record, Brussels, March 31-April 2, 1993, vol II, pp199-205.

[5] Spitzer P, Chu R, Lau T, Mize E. HL7 Data modelling proposal. (A data/ information model for a generalized health care provider, based on the contents of the HL7 Version 2.1. specification). July, 1990.

[6] NHS Information Management Centre. The Common Basic Specification in 1992. Draft version 4.2. Birmingham 1992.

[7] Molteno B. The National Health Service data model: a decade of development. Health Informatics Europe, 1, 1993, 25-7.

[8] Claassen AMW, Zanstra PE, van der Haring EJ, Pomper P, de Vries Robbé PF. Using Network Formailisms in the Medical Domain: The Case for Genericity. in P Barahona, M Veloso and J Bryant (eds) , Proceedings of Medical Informatics in Europe MIE 94, pp57-62.

[9] Goble CA, Rector AL, Nowlan AL, Glowinski AG. Conceptual, Semantic And Information Models for Medicine. Published in the Proceedings of the 4th European-Japanese Seminar on Information Modelling and Knowledge Bases. 31st May - 3rd June 1994, Stockholm.

[10] Nowlan W A . Clinical Workstations: Identifying Clinical Requirements and Understanding Clinical Information. In International Journal of Bio-Medical Computing. 34 (1994) pp85-94.

[11] Rector AL, GALEN and PEN&PAD, in Proceedings of AIM-CEN European Workshop on the Medical Record, Brussels, March 31-April 2, 1993, vol II, pp304-308.

[12] Rossi Mori A. MOSE: Model for the Representation of Semantics in Medicine, in Proceedings of AIM-CEN European Workshop on the Medical Record, Brussels, March 31-April 2, 1993, vol I, pp57-67.

[13] Rossi Mori A, Gangemi A and Galati M, The Coding Cage, iin A Reichart, BA Sadan, S Bengtsson, J Bryant, U Piccolo (eds), Proceedings of Medical Informatics in Europe MIE 93, Freund Publishing House Ltd, 1993, pp 466-472.

[14] Zanstra, PE and Beckers, WPA, Computerized Assessment with Primary Care guidelines in the Netherlands, In C. Gordon and JP Christensen (eds) Health Telematics for Clinical Guidelines and Protocols, IOS 1995.

[15] Humphreys BL, Hole WT, McCray AT, Fitzmaurice, JM. Planned NLM/AHCPR Large Scale Vocabulary Test: Using UMLS Technology to determine the extent to which controlled vocabularies cover terminology needed for health care and public health. J. Am Med Inform Assoc.1996;4:281-287