

Wüsteria

Barry Smith^a, Werner Ceusters^b, Rita Temmerman^c

^a IFOMIS (Institute for Formal Ontology and Medical Information Science), Saarland University, Germany and
Department of Philosophy, University at Buffalo, NY, USA

^b ECOR (European Centre for Ontological Research), Saarland University, Germany

^c CVC (Centre for Terminology and Communication), Erasmushogeschool, Brussels

Abstract

The last two decades have seen considerable efforts directed towards making electronic health records interoperable through improvements in medical ontologies, terminologies and coding systems. Unfortunately, these efforts have been hampered by a number of influential ideas inherited from the work of Eugen Wüster, the father of terminology standardization and the founder of ISO TC 37. We here survey Wüster's ideas – which see terminology work as being focused on the classification of concepts in people's minds – and we argue that they serve still as the basis for a series of influential confusions. We argue further that an ontology based unambiguously, not on concepts, but on the classification of entities in reality can, by removing these confusions, make a vital contribution to ensuring the interoperability of coding systems and healthcare records in the future.

1. Introduction

The goal of an electronic health record (EHR) is to achieve faithful clinical data entry in such a way as to meet the requirements of communicability for both human and machine. [1] To this end much emphasis has been placed on clinical coding, with the rationale that it is codes that will make it possible to associate with the terms used by humans in expressing patient data the sorts of uniform syntax and precise meanings that can be interpreted by software. Code-based terminologies now exist in many different flavours. The Unified Medical Language System contains in its MetaThesaurus over 100 such systems, which are said to comprehend in all some 3 million medical “concepts” [2].

On the EHR front, too, progress is being made. CEN/TC251 has brought Europe-wide acceptance of the need for a comprehensive, communicable and secure EHR as a prerequisite for the delivery of high-quality healthcare, and this European vision has gained international acceptance, leading to the establishment of new standards at the ISO level [3]. As we shall show, however, the realization of this vision is stymied by the fact that the new standards inherit from the earlier work of ISO Technical Committee (TC) 37 a fundamental incoherence.

ISO TC 37 was founded in 1951, largely through the efforts of a certain Eugen Wüster (1898–1977), an Austrian businessman, saw-manufacturer, professor of woodworking machinery, and devotee of Esperanto, who ran the secretariat of TC 37 for the first decades of its existence. [4] Wüster was principal author of almost all the seminal documents on terminology standardization and thus responsible for very many of the ideas which, because of ISO's rules governing re-use of prior standards, have been propagated in ever widening

circles ever since. Given this astonishing influence, it is worth spending some time to convey the flavour of Wüster's thinking.

First, we need to note that four distinct views of concepts can be distinguished in the literature. On the psychological view concepts are *mental entities*, analogous to ideas or beliefs; on the linguistic view concepts are the (somehow regimented) *meanings of general terms*; on the epistemological view concepts are *units of knowledge* (as this term is used in phrases such as 'knowledge representation'); and finally, on what some might call the 'ontological' view, concepts are *abstractions of kinds, attributes or properties* (i.e. of general invariant patterns on the side of entities in the world). Sadly, elements of all four views are found mixed up together in almost all terminology-focused work in informatics today. [5]

Wüster himself is a proponent of the psychological view. Our knowledge of concepts, he tells us [6], is rooted in the experiences of the new-born infant, which finds itself "constantly amidst a panoply of diverse sensory impressions". The child begins thereupon to mentally sub-divide this sensory mosaic into individual objects (and Wüster stresses repeatedly in this connection that objects in reality are *constructed* by human beings, and that there is a high degree of arbitrariness and variability to such construction). The child can thereafter also *remember* objects, such memories constituting what Wüster calls "individual concepts". Examples are: "'Napoleon' or the concept of my fountain pen." [6]

If, as Wüster would have it, "a speaker wishes to draw the attention of an interlocutor to a particular individual object, which is visible to both parties or which he carries with him, he only has to point to it". Otherwise, however, "the only thing available is the individual concept of the object, provided that it is readily accessible in the heads of both persons." (Those engaged in communication about, say, Napoleon, are thus somehow required to gain access to the interiors of each other's heads.)

In the course of time, the child notices that some individual objects – e.g. apples, or bricks, or cans of paint – are "interchangeably alike" and are given the same name by older speakers of the language. "The child learns to blend the individual concepts of such objects in its thinking" and thus arrives at *general concepts*, which are, like individual concepts, "thought (= mental) objects. They exist only in the heads of people." As individual concepts can be grouped together into general concepts, so general concepts can be grouped together into concepts of higher degrees of abstraction, as when we move from the general concept *apple* to the superordinate concept *fruit*. The formation of concepts at these higher levels, too, Wüster thinks, is "highly dependent on human discretion."

Terminology work is designed to provide clear delineations in this "realm" of concepts via definitions [7], and Wüster thinks that *terms* can be assigned to concepts only when such definitions have been formulated. (How else, after all, are we to gain access to the denizens of this strangely ethereal realm?)

2. Concepts and Characteristics

Wüster's account of concept learning and his insistence on the arbitrariness of concept-formation have long since been called into question by cognitive scientists. Even very small children manifest in surprisingly uniform ways an ability to apprehend objects in their surroundings as instances of natural kinds – in ways which go far beyond what they apprehend in sensory experience. There is now much evidence (e.g. in [8]) to the effect that, for objects in the biological realm, this ability rests on a shared innate capacity to apprehend the surrounding world in terms of underlying structures or powers. The latter are

invisible to the child, but adults may learn to recognize them as structures of a molecular sort.

Wüster's idea according to which, before we can assign a term to a concept, we must first "delineate" the concept, is also open to serious objections. To delineate means: to list the totality of "characteristics" which form a concept's *content* or *intension*. Unfortunately Wüster provides conflicting elucidations of what such "characteristics" might be [7], conceiving them sometimes as if they were themselves *concepts* (so that, like other concepts, they would exist in the heads of people), and at other times as *properties of objects* existing in the world. (This is in keeping with the general failure to discriminate clearly between objects and concepts which runs through all of Wüster's thinking.)

Some recent terminology work is clearer in this respect [9]. Unfortunately, however, even in more recent ISO documents [10] the problems still linger, since the relevant communities have still to find a coherent means by which concepts and their characteristics should somehow span the divide between concepts as creatures of the mind and as properties of objects in the world. This lingering incoherence, which spreads also independently of ISO's influence [11], explains why so many terminologies contain certain characteristic families of errors in coding and documentation which flow from the fact that those involved in their authoring and maintenance are unsure as to whether their task is the representation of *ideas in people's heads, meanings of words, consensus knowledge of experts in a discipline or types of entities in the world* [5]. Consider for example the definition of *disorder* that we find in SNOMED: "Disorders are concepts [!] in which there is an explicit or implicit pathological process causing a state of disease which tends to exist for a significant length of time under ordinary circumstances." [12, p. 23]. Taken together with SNOMED's definition of concepts as "unique units of thought" [12, p. 11], this would seem to imply that all disorders are imagined.

3. Wüsterian Medicine

Wüster's assumption to the effect that concepts are formed through the application of human discretion to perceived similarities may have led some to suppose that his ideas are well-suited to the area of medical terminology, which is after all subject to the constant coinage of novel terms. Unfortunately, however, there is one prominent feature of medical reality which makes Wüster's approach here inapposite. For in medicine we often have to deal with families of entities in reality in relation to which we are able to grasp few characteristics "identifiable in encounters of similars", and certainly too few to allow definitions of the corresponding concepts. (It is in part for this reason that some 85% of SNOMED-CT's concepts remain in its July 2003 version [12] undefined.)

Consider, for example, a tumour. This starts out as initially undetectable mutations in a small number of cells and then becomes transformed by degrees into a full-fledged object on the scale of coarse anatomy. For very many types of pathogenic process it seems at best simplistic to suppose that we could isolate in perception certain "essential properties" which could be identified in definitions as the "characteristics" of corresponding general concepts. That the detection, classification and diagnosis of such processes involves to such a high degree the application of statistical techniques is already a sign of the fact that we are dealing here with patterns in reality which go beyond the realm of concepts as this is conceived, in Wüsterian fashion, in terms of lists of necessary and sufficient conditions.

The reason for this miscalibration turns on the fact that the Wüsterian notion of concept *has nothing to do with medicine (or biology) at all*. Wüster and his early TC 37 colleagues were concerned primarily with standardisation in the domain of *commercial artefacts*, and especially of *manufactured products*. Wüster himself was the author of a multivolume work

entitled *The Machine Tool. An Interlingual Dictionary of Basic Concepts* (London 1976). Machine tools truly are such as to manifest characteristics identifiable in encounters of similars – because they have been manufactured as such. Vocabulary itself is treated by Wüster and his TC 37 followers “as if it could be standardised in the same way as types of paint and varnish or aircraft and space vehicles” [13, p. 12].

Certainly, there are also non-artefactual objects in the Wüsterian universe. As ISO/IEC JTC1 SC36 N0579 (for example) puts it:

an object is defined as anything perceived or conceived. Some objects, concrete objects such as a machine, a diamond, or a river, shall be considered material; other objects are to be considered immaterial or abstract, such as each manifestation of financial planning, gravity, flowability, or a conversion ratio; still others are to be considered purely imagined, for example, a unicorn, a philosopher’s stone or a literary character. [14]

Unfortunately such elucidations are so vague as to leave the putative user of the corresponding standards in the dark. Are *processes* objects? Are they concrete or abstract? Are *characteristics* objects? Are *concepts* objects? Are dispositions, functions, limbs, body cavities, blood flow, apoptosis, or types of pus, objects? Are they concrete or abstract? Material or immaterial? Real or imagined? The ISO literature still leaves us with no coherent means to provide answers to such questions, and this in spite of the fact that the task of creating a principled framework in which such answers could be given is of increasing importance to the future of medical coding and of the EHR. In the document just cited, however, in which real objects such as rivers are placed on the same level as imagined objects such as unicorns, ISO makes it clear what it thinks of the importance of this task for the future of terminology research:

In the course of producing a terminology, *philosophical discussions on whether an object actually exists in reality ... are to be avoided*. Objects are assumed to exist and attention is to be focused on how one deals with objects for the purposes of communication. [14, emphasis added]

As we have argued at length elsewhere, however [5], it is precisely such philosophical discussions which are required if we are to undo the sore effects of Wüster’s influence.

4. How Medical Terms Are Introduced

The typical scenario for the introduction of new terms into medical language is as follows. A new disease or virus is encountered in reality, and the communities involved recognize that they need some way to refer to the newly discovered kind as they encounter its successive instances. Agreement is then reached in these and those languages that these and those terms should be used henceforth to refer to *instances of this kind of entity*.

In terminology circles, however, the demand is now raised to add in addition some third thing: the corresponding *concept*. Because concepts themselves are ethereal in nature, they require the support of something else – namely *definitions* – to enable terminology users and associated software applications to gain access to them. At the same time the definitions thus created serve henceforth to restrict the sorts of entities that can be admitted as falling under the corresponding concepts.

In areas like manufacturing or commerce the purpose of standardization is precisely to bring about a situation in which entities in reality (such as machine parts, or contracts) are indeed *required* to conform to certain agreed-upon standards. Such a requirement is however alien to the world of medicine, where it is the entities in reality which must serve in every case as benchmark. Even in medicine, however, terminologists have been encouraged to focus on concepts and definitions rather than on the corresponding entities in reality.

We can now understand more precisely why so many of the medical ‘concepts’ in terminologies like SNOMED-CT remain undefined. The reason turns on the way in which

medical terms are introduced into our language. Such terms reflect entities in reality for which we characteristically have access to only a small fraction of the relevant biological or clinical features. Almost all disorder terms are introduced, not because we already have a clear definition reflecting known characteristics, but because we have a *pool of cases*.

This means that many medical terms are introduced before their users have any ‘conceptual’ understanding of what they *mean*. These users are however able to grasp what they *designate in reality*: they can see the relevant entities before them in the lab or clinic.

5. An Ontological Basis for Coding Systems and the EHR

There are many who hold that it will suffice to establish communication standards for the EHR if we can only establish a way to refer unambiguously to “concepts” as units of knowledge agreed upon by domain experts and defined in formal ways. As we hope to have shown in the foregoing, this detour through “concepts” – at least as realized in the domain of biomedicine – represents rather an alien accretion of what we can only call *International Standard Bad Philosophy*. It is time, we believe, to pursue new means of conceiving the relation of terms to medical reality in which the detour through concepts is abandoned and in which we draw instead on the best theories and tools which contemporary philosophy has to offer – and this means above all the right sort of ontology, an ontology that is able explicitly and unambiguously to relate to the universal types or kinds in reality as well as to the individual tokens (such as you and me) which are their instances.

The principal task of medical terminology systems is to represent such universal types or kinds, and the principal task of the EHR is to represent the corresponding instances. Our proposal, then, is to develop an ontology in which these two kinds of representations are tied together from the start, without the detour through the realm of concepts.

Note that we are not hereby claiming that to establish the ontology of the world of biomedical universals and instances will be a simple task. There is, as is clear, no single unified perspective on which all reasonable persons must agree if they would only open their eyes. Hence the popularity of T. S. Kuhn’s ideas on conflicting paradigms, and hence the influence of Wüster’s own ideas on what he sees as the human-induced arbitrariness involved in the “construction” of both objects and concepts. Against both Kuhn and Wüster, however, we see these matters precisely in terms of the existence of a plurality of different perspectives *on one and the same world* – perspectives corresponding, for example, to the different life science disciplines and to different biomedical terminologies. It is because of the immense complexity of this one world that it is accessible to us only in terms of a wide variety of such different perspectives.

On our view, however, some terminologies are to be preferred to others because they project onto the world beyond in a way which enjoys a higher level of correctness or adequacy to the universals or kinds in reality. On the view of Wüster and his followers, in contrast, there is no independent benchmark in relation to which concept-systems could be established as correct, and thus also no independent fulcrum in terms of which concept-systems could be integrated together in robust fashion. On our view such integration can be attained precisely because perspectives are projected onto this common independent reality, which embraces entities at all levels of granularity, from the molecule to populations [15].

Our approach does not, be it noted, ignore the psychological and linguistic dimensions of the application of medical terms. Indeed, it takes great pains to ensure that its categories apply to the world itself in all salient dimensions, including beliefs and observations, utterances and terms. It is thus in a position to make it crystal clear, in relation to all the clinical data registered in EHRs, whether entities in the associated coding systems refer to *diseases*, or to *statements made about diseases*, or to *acts on the part of physicians*, or to

documents in which such acts are recorded, or to observations of such acts, or to statements about such observations. In this respect, too, it is opposed to the established approaches to the construction of coding systems for use in the EHR in recent years.

6. Conclusion

The application of a sound realist ontology to the domain of healthcare can make coding systems both logically more coherent and also more closely compatible with our commonsensical intuitions about the medically salient objects and processes in reality. It can thus not only help in detecting errors in existing coding systems but also, by allowing the formulation of intuitive principles for the creation and maintenance of such systems, help in avoiding similar errors in the future [16].

To achieve the requisite coding systems and the associated EHR architecture will of course require a huge effort, since the relevant standards need to be overhauled from the ground up by experts who are cognizant of the need for clarity and familiar with the methods of sound ontology. Even before that stage is reached, however, there is the problem of making all constituent parties – including patients, healthcare providers, system developers and decision makers – aware of how deep-seated the existing problems are.

Acknowledgements: Work on this paper was carried out under the auspices of the Alexander von Humboldt Foundation, the EU Network of Excellence in Medical Informatics and Semantic Data Mining, and the Project “Forms of Life” sponsored by the Volkswagen Foundation. Thanks for helpful comments are due also to Gerhard Budin and Gunnar Klein, who however bear no responsibility for the positions here adopted.

References

- [1] Redondo JR, Ceusters W, González JM, Iakovidis I. European electronic healthcare records towards the future. *Health in the New Communications Age*, 671-675, IOS Press 1995.
- [2] National Library of Medicine; UMLS Fact Sheet, updated 7 May 2004.
- [3] ISO 18308: Health informatics – Requirements for an electronic health record architecture. 2002 (<http://www.iso.ch/iso/en/CatalogueDetailPage.CatalogueDetail?CSNUMBER=33397>).
- [4] Oeser E, Galinski C. *Eugen Wüster (1898–1977). Leben und Werk*, Vienna: Infoterm, 1998.
- [5] Smith B. Beyond concepts, or: Ontology as reality representation, *Formal Ontology and Information Systems* (FOIS), Amsterdam: IOS Press, 2004;:73–84.
- [6] Wüster E. The wording of the world presented graphically and terminologically (selected and translated by JC Sager), *Terminology*, 2003;9(2):269-297.
- [7] Wüster E. *Einführung in die Allgemeine Terminologielehre und Terminologische Lexikographie*, Vienna/New York: Springer, 1979.
- [8] Gelman SA, Wellman HM. Insides and essences: Early understandings of the non-obvious. *Cognition*, 1991;38:213-244.
- [9] Wright SL, Budin G (eds.): *Handbook of terminology management*. Amsterdam: Benjamins 1997.
- [10] ISO-1087:1990 and ISO-1087-1:2000 *Vocabulary of terminology* (versions of 1990 and 2000).
- [11] See for example: <http://www.w3.org/2004/02/skos/core/spec/2005-05-04>
- [12] College of American Pathologists. *SNOMED Clinical Terms® User Guide*. January 2003 Release.
- [13] Temmerman R. *Towards new ways of terminology description*. Amsterdam: Benjamins, 2000.
- [14] ISO/IEC JTC1 SC36 N0579:1999. Text for FDIS 704. Terminology work: Principles and methods.
- [15] Smith B, et al. Relations in biomedical ontologies, *Genome Biology*, 2005; 6(5): R46.
- [16] Ceusters W, Smith B, Kumar A, Dhaen C. Mistakes in medical ontologies. *Ontologies in Medicine. Proceedings of the Workshop on Medical Ontologies*, Amsterdam: IOS Press, 2004:145-63.

Address for correspondence: Barry Smith, IFOMIS, Saarland University, Postfach 151150, D-66041 Saarbrücken. Email: phismith@buffalo.edu. Internet: <http://ifomis.org>.