

Semantic Interoperability in Healthcare State of the Art in the US

A position paper with background materials
prepared for the project



March 3rd, 2010

Werner Ceusters & Barry Smith
New York State Center of Excellence in Bioinformatics and Life Sciences
Ontology Research Group
701 Ellicott street
Buffalo NY, 14203
USA

Table of contents

1	EXECUTIVE SUMMARY	3
2	PROBLEMS AND CHALLENGES.....	3
2.1	TOO STRONG INVOLVEMENT OF INDUSTRY	4
2.2	AN UNJUSTIFIED BELIEF IN THE VALUE OF CONCEPT-BASED APPROACHES	6
3	ISSUES IN REPRESENTATIONS.....	7
3.1	FIRST GENERATION SYSTEMS	7
3.2	SECOND GENERATION SYSTEMS	8
3.2.1.	<i>Unified Medical Language System (UMLS)</i>	8
3.2.2.	<i>SNOMED-CT</i>	9
3.3	THIRD GENERATION SYSTEMS: ONTOLOGIES	10
3.3.1.	<i>Open Biomedical Ontologies</i>	12
4	APPENDIX: STANDARDIZATION FROM A US PERSPECTIVE	13
4.1	US FORMAL STANDARDIZATION	13
4.2	US STANDARD DEVELOPMENT ORGANIZATIONS	14
4.2.1.	<i>Health Level Seven, HL7</i>	14
4.2.2.	<i>Digital Imaging Communications in Medicine, DICOM</i>	14
4.2.3.	<i>The Institute of Electrical and Electronics Engineers (IEEE)</i>	15
4.2.4.	<i>The American Society for Testing Materials (ASTM)</i>	15
4.2.5.	<i>OMG/CORBA</i>	16
4.3	US AGENCIES AND GOVERNMENTAL INITIATIVES WITH AN IMPACT ON SEMANTIC INTEROPERABILITY WORK IN THE eHEALTH DOMAIN	16
4.3.1.	<i>ONCHIT - Office of the National Coordinator for Health information Technology</i>	16
4.3.2.	<i>CHI – Consolidated Health Informatics (US e-government plan)</i>	16
4.3.3.	<i>American Health Information Community – (AHIC)</i>	17
4.3.4.	<i>Healthcare Information Technology Standards Panel (HITSP)</i>	18
4.3.5.	<i>NCRR - National Center for Research Resources</i>	18
4.3.6.	<i>USHIK - United States Health Information Knowledgebase</i>	18
4.3.7.	<i>CMS - Centers for Medicare & Medicaid Services</i>	19
4.3.8.	<i>Centers for Disease Control and Prevention – CDC</i>	20
4.3.9.	<i>Public Health Data Standards Consortium – PHDSC</i>	21
4.4	OTHER ORGANIZATIONS AND INITIATIVES.....	22
4.4.1.	<i>National Uniform Billing Committee - NUBC</i>	22
4.4.2.	<i>The Certification Commission for Healthcare Information Technology (CCHIT)</i>	23
4.4.3.	<i>IHE – Integrating the Healthcare Enterprise</i>	24
4.4.4.	<i>National Uniform Claim Committee - NUCC</i>	25
4.4.5.	<i>Clinical Data Interchange Standards Consortium (CDISC)</i>	25
4.4.6.	<i>Biomedical Research Integrated Domain Group (BRIDG)</i>	25
4.4.7.	<i>The Nationwide Health Information Network (NHIN)</i>	26
4.4.8.	<i>Cancer Biomedical Informatics Grid (caBIG)</i>	26
5	REFERENCES	27

1 EXECUTIVE SUMMARY

Semantic interoperability can be defined as the ability of two or more computer systems to exchange information in such a way that the meaning of that information can be automatically interpreted by the receiving system accurately enough to produce useful results to the end users of both systems.

Several activities are currently being performed by a variety of stakeholders to achieve semantic interoperability in healthcare.

Many of these activities are *not* beneficial, because they place too great a focus on business aspects and not enough on involvement of the right sorts of researchers, in particular those that are able to see how the data and information relate to the entities of concern on the side of the patient.

The lack of a central focus on the patient, and the associated focus on ‘concepts’, have spawned a variety of mutually incompatible terminologies exhibiting non-resolvable overlap.

The predominance of the healthcare IT industry in the writing and selection of semantic interoperability standards mitigates against the benefits that standards, when well designed, can bring about.

2 PROBLEMS AND CHALLENGES

There is no doubt that Information Technology (IT) standards have had a major positive impact on many facets of data capture, communication and analysis in data-intensive domains such as healthcare and the life sciences. Certainly, this has led to better research with results that are both more reliable and capable of being more effectively disseminated. These standards have allowed research groups working in different regions or specialisms to combine and reuse their resources more readily. And they are beginning to have positive effects in supporting continuity of care.

The question is whether the current e-Health standardization efforts and other approaches to achieving *semantic interoperability* are equally beneficial. Semantic interoperability can be defined as:

the ability of two or more computer systems to exchange information and have the meaning of that information automatically interpreted by the receiving system accurately enough to produce useful results to the end users of both systems.

Although semantic interoperability under this definition is already a reality in some corners of the health domain, it is so only at different degrees and at various scales, and typically because ‘being useful to end users’ refers to end users who are human beings and thus have the capacity to make sense of the data exchanged even when it is incomplete, mistaken, redundant, ambiguous, lacking adequate formalization, and so forth. Computers, in contrast, have no such capacity.

Some approaches to semantic interoperability are successfully applied in limited settings. However, many of the increasingly more sophisticated needs of clinicians and researchers are not, and will not be, met by these approaches, since the realization of the goals of modern translational medicine requires semantic interoperability which spans scientific domains and

national boundaries. With the vast amount of data becoming available and exchanged, the challenge now is to ensure that transmitted data are understood not only by the human beings on both ends of the IT communication channel, but also by computer systems and their associated software. There is a growing need for systems that are able to act and react automatically to changes in the data repositories to which they have access. Only thus will we have the opportunity to avoid data overload on the side of the end user.

Currently many health and life science databases, including ontologies, terminologies and electronic health records (EHRs), are organized in ways that only fulfill the needs of the original designers, but have little chance of bringing benefits to the research community at large. Thus, to give only one example, resources designed to support semantic interoperability in the experimental biology and clinical trial domains do not support interoperation with counterpart resources developed in the contexts of hospital care and general practice.

At the same time, however, some of the standards and mechanisms put in place to achieve the needed semantic interoperability are *not* beneficial, because they place too great a focus on business aspects and not enough on involvement of the right sorts of researchers, in particular those who are able to see the difference between data and information – and associated billing practices – on the one hand, and what the data and information are about – the biological and clinical phenomena – on the side of the patient.

We have identified the following barriers standing in the way of achieving truly beneficial semantic interoperability in Health IT systems in the US:

- too large a number of players (clinicians, patients, payers, industry, government ...) with competing agendas
- insufficient coordination based on a shared set of coherent principles
- overestimation of the value of terminologies and concept-based ontologies
- inadequacy of current systems, primarily electronic health record systems, to capture data adequately for example in ways that support continuity of care
- inconsistent and badly documented standards, some of them maintained by consultants and others who benefit from inadequate standards and from poor documentation
- shortage of trained personnel who can span the divide between IT and biological and clinical expertise
- too rapid turnover of trained personnel, so that promising systems are abandoned or poorly maintained
- vague notions of ‘meaningful use’ which have little to do with semantic interoperability
- the rush – and available government funding – to install existing EHR systems none of which has the right foundations to support full semantic interoperability.

2.1 Too strong involvement of industry

Standards development organizations such as ANSI [1-2] do not hide the fact that the principal focus of their work is commercial and private-sector needs. HITSP [3] is following a similar path, as evidenced by its healthcare interoperability specifications. [4] It is thus no surprise that

columnists such as Dana Blankenhorn question whether we really want industry writing the nation's health IT standards since 'CCHIT's "standards" are, in fact, mainly approvals of whatever industry is doing'. [5]

All of this having been said, we recognize that there is a great body of knowledge and capability on the part of many of those who have been spearheading efforts at writing standards in support of semantic interoperability in the health domain thus far. The goal should therefore lie in establishing procedures that will make it practically possible for a new generation of researchers to become involved, i.e. researchers who are aware of the possibilities which new technologies can play in realizing semantic interoperability in a coherent fashion. Current barriers to new entrants include: membership fees required by some established standards organizations, idiosyncrasy of existing standards, which were often developed in ad hoc ways and independently of, for example, the work of standards bodies such as W3C; poor documentation of existing standards, making it practically impossible for new entrants to gain access.

Industry can meet some of these hurdles, and can afford the needed payments and they can afford to volunteer the time of employees to continue in support of existing practices, and thereby to protect its own investment in established systems that are to a worrying degree no longer fit for purpose. Even industrial organizations, however, will often see no business case in making the needed investments; why should single firms pay the costs of retooling and retraining that would be needed for the required global approach? On the other hand, current efforts within the pharmaceutical industry, for example within the framework of the Pistoia Alliance (<http://www.pistoiaalliance.org/>), are exploring ways to overcome such hurdles by providing an open foundation of data standards, ontologies and web-services that would streamline the pharmaceutical drug discovery workflow.

Academic researchers and other members of the non-profit sector in the health arena have however in general still not stepped up to the plate to write the needed improved standards, primarily because of funding issues. Even worse, there is now considerable evidence that the research community, when submitting proposals to fund research on developing and testing the new standards that would be needed to reap the full benefits of today's technology, is being pressured by funding agencies to adopt the standards supported by the healthcare industry just because they are 'established', and have been designated as such by official agencies. In this way progress in the standards arena is being blocked because the 'best of breed' is being selected from a group of candidates created on the basis of outmoded technology.

In biological domains such as model organism research, in contrast, academic and non-profit sectors have made considerable progress. Above all, one lesson learned in these domains concerns the tremendous benefits to semantic interoperability that is brought, in the era of the Web, by the use of public domain resources. Whenever standards artifacts such as ontologies belong in the public domain they can be used and reused freely both as stand-alone resources and as incorporated inside other resources, for example inside annotation databases. In this way they serve to increase tremendously the possibilities for effective access to and use of these resources.

The predominance of industry in the writing or selection of semantic interoperability standards mitigates against these benefits for a number of reasons. To highlight just one extremely important issue: all legacy EHRs we know of exhibit the confusion between information and reality, as exemplified by the practice of allowing diagnoses to be entered into the EHR system without requiring that wherever possible it should be specified also *what these diagnoses are*

about, i.e. the underlying disorder. As a consequence, it is impossible for software agents analyzing EHR data to figure out the number of diseases a given patient has suffered from over a given period of time, the number of diseases that have been cured, and so on. It remains forever unrecorded whether different diagnoses for a particular patient correspond to distinct disorders, to conflicting opinions about one single disorder, or to consistent opinions about one disorder that has been evolving over time. And because standards for healthcare messaging such as the ISO standard Reference Information Model (RIM) of Health Level 7 (HL7) do not make these distinctions either, will mean that these problems will persist still further.

Even worse: scientifically doubtful arguments are sometimes used to ‘justify’, rather than to correct, such problems. Thus one can read in [6] – a study of the NCI Thesaurus (NCIt) ‘quality assurance life cycle’ responding to an earlier paper critiquing the quality of the NCIt [7] – that ‘*While other faults were also accurately identified, such as the incorrect use of the “all” description logic qualifier, making corresponding changes to NCIt was not cost-effective*’ and that ‘*Many of the problems the review identified, if corrected, would not materially affect the ability of NCIt to meet the use cases that it must support*’. The problem is that these problems will materially affect the multitude of those external researchers who might otherwise be potential users of the NCIt.

Stead and Lin recently argued [8] that ‘*persistent problems involving medical errors and ineffective treatment continue to plague the industry*’ and that ‘*Many of these problems are the consequence of poor information and technology (IT) capabilities*’. We hope that industry and standards development organizations pay attention to their request for changes in this regard.

2.2 An unjustified belief in the value of concept-based approaches

Current attempts to achieve semantic interoperability all rely on agreements about the understanding of the so-called ‘concepts’ stored in terminology systems such as nomenclatures, vocabularies, thesauri, or ontologies. The idea is that, if all computer systems use the same terminology, or mutually compatible ones, then they can understand each other perfectly. The reality, however, is that the number of terminology systems with mutually incompatible definitions or non-resolvable overlap amongst concepts grows exponentially. This is true even where the terminology systems are said to be ‘ontologies’ or to have ‘ontology-like properties’. Ontologies, thereby, initially touted as a solution to the problem of semantic non-interoperability, in fact contribute ever more to this very problem. This is because these systems leave unspecified what *concepts* actually are, or to what, if anything, they might correspond; the various locally created sets of concepts are thereby left unmoored from any common benchmark in reality.

Thus, while ontologies have made considerable progress precisely in the biological realm, such progress is much less impressive in the clinical domain, because so many of the ontologies now being developed in the various clinical specialisms are being built by groups working independently of each other and with little or no resort to common high-quality resources.

Increasingly, one or other version of description logic (primarily OWL DL) is being used in their development. However, the use of a logical representation language is clearly not enough to ensure the high quality of an information resource [9], and even ontologies employing the same formal language are often not combinable into a single resource because of multiple incompatibilities between the ways different groups use this language and the ways they

populate it with non-formal terminology content when expressing biological or clinical information [10].

Ontologies such as the Gene Ontology which have been able to break out of the conceptual orientation, are increasingly being used in biological domains to provide effective corridors of semantic interoperability between distinct biological information resources [11]. The idea is:

first, that, if multiple bodies of relevant information can be annotated using common, non-redundant sets of ontology terms with definitions formulated using some common logical language, then the information they contain will thereby be more easily accessible and capable of being computationally integrated together;

second, that the sets of ontology terms should be declared to be of global application through the insistence upon one single ontology for each relevant domain; the tendency for ontologies to contribute to silo formation is hereby thwarted [12];

and *third* the separate, single ontologies created for each single domain are viewed not as representations of the ‘concepts’ inside the heads of different groups of scientists, but rather as representations of one and the same biological reality, as revealed through experimental evidence and through the scientific literature.

There are, of course, multiple reasons why this strategy is not meeting with the same successes in the field of human health as in the domains of biology. The medical domain is not only much more complex than strictly biological domains covered by, for example, the Gene Ontology; it is also a domain in which, for the reasons addressed in the above, experiments in open source provision of resources are much more difficult to carry out in an effective way.

3 ISSUES IN REPRESENTATIONS

3.1 First generation systems

There is a wide variety in the sorts of systems that are used to create abstractions of reality. Amongst the oldest such systems, both as concerns the technology used and the paradigm applied, are classifications, nomenclatures and thesauri.

A *classification* provides a set of classes used to arrange items into separate groups, mainly for statistical purposes. The classes are therefore typically mutually exclusive. Some classifications (e.g. ICD) also provide a taxonomy among classes, to facilitate the clustering of classes for the purpose of creating synthetic statistical tables. Such taxonomies typically do not reflect the way reality is structured, since they embody peculiar terminological conventions designed to ensure mutual exclusivity.

A *nomenclature* provides a list of expressions with the goal of capturing in a systematic and reproducible way a set of details. An example in use in the US is LOINC [13], the Laboratory Logical Observation Identifiers Names and Codes database. LOINC is published by the Regenstrief Institute and is a standard adopted for the representation of laboratory procedures and for the structured labeling of medications. The original release of the nomenclature was in the spring of 1995.

A *thesaurus* is a system of predefined descriptors, usually designed for indexing and retrieval purposes. An example is the NCI Thesaurus (NCIt) [14], which was created by the National

Cancer Institute's Center for Bioinformatics and Office of Cancer Communications starting in 1997 from a collection of local terminologies used for coding documents and from a clinical trials coding scheme. The main goals for creating and maintaining the NCI are:

- 1) to provide a science-based terminology for cancer that is up-to-date, comprehensive, and reflective of the best current understanding;
- 2) to make use of current terminology "best practices" to relate relevant concepts to one another in a formal structure, so that computers as well as humans can use the Thesaurus for a variety of purposes, including the support of automatic reasoning;
- 3) to speed the introduction of new concepts and new relationships in response to the emerging needs of basic researchers, clinical trials, information services and other users [15].

The NCI is certainly a useful tool for the internal purposes of the NCI, which must be given credit for trying to bridge the clinical and basic biology terminology realms in a single resource. It must be given credit also for its sophisticated technology, for keeping track of updates, for being one of the earliest to federate its ontology operationally with another ontology system (MGED Ontology), and for trying to harmonize with external ontology modeling practices. In realizing goal 2), however, the NCI acquires 'ontology-like' properties. Unfortunately, as shown in [16], the ontological features of the system do not work well together with its terminological parts. The system thereby suffers from a number of problems encountered in so many of the biomedical terminologies produced in recent years.

The NCI Thesaurus, like every other major ontology-oriented artifact in the biomedical domain, is a never-ending work in progress, whose content is dictated by the needs of its users and customers. If, however, it wants to establish itself as a useful and trustworthy terminological resource and to play the role of a reference ontology in other contexts, then a considerable effort will have to be made in order to clean up its hierarchies and to correct the definitions and ambiguous terms which they contain. We strongly suggest the use in this endeavor of a principles-based methodology that will allow the NCI to be tested not just for internal consistency but also for consistency with that part of reality which it is intended to represent.

3.2 Second generation systems

3.2.1. Unified Medical Language System (UMLS)

In 1986 the US National Library of Medicine [17], part of the National Institutes of Health within the US Department of Health and Human Services, initiated the UMLS research project [18-19] with the goal of overcoming the barriers to effective use of health information technology created by the different ways different information sources use language to refer to one and the same entity (for example 'atrial fibrillation', 'auricular fibrillation', 'af'). The UMLS offers as a solution to these barriers an extensive set of terminologies with semantic links between terms from different sources. The UMLS project delivers these capabilities in three knowledge sources: the Metathesaurus, the Semantic Network and the SPECIALIST lexicon, together with several tools including MetamorphoSys, lvg and MetaMap. Our focus here is on the Metathesaurus.

The Metathesaurus is a database built from more than 100 versions of various vocabularies used in patient care, billing, public health, cataloguing of biomedical literature and research. These are referred to as the "source vocabularies" of the Metathesaurus and include CPT, Gene Ontology, HL7 V3.0, ICD-9-CM, ICD-10-CM, LOINC, MeSH, Medline, RxNorm and SNOMED-CT.

The Semantic Network component of the UMLS consists of a set of 135 Semantic Types, which are used to categorize all of the concepts contained in the Metathesaurus, together with a set of 54 Semantic Relationships between these Types. Semantic Types are divided into Entities and Events. Entities are further subdivided into Physical Objects and Conceptual Entities. Semantic Relationships include physical relationships such as *part_of*, *contains*, and *connected_to*; functional relationships such as *treats*, *causes* and *manifestation_of*; spatial relationships such as *adjacent_to*, *surrounds*, and *traverses*; and conceptual relationships such as *evaluation_of*, *assesses_effect_of*, and *diagnoses*.

3.2.2. SNOMED-CT

SNOMED-CT® [20] was developed by the College of American Pathologists (CAP) [21], and grew out of the merger, expansion, and restructuring of the SNOMED RT® (Reference Terminology) [22] and the United Kingdom National Health Service Clinical Terms (also known as the Read Codes) [23]. CAP and the NHS Information Authority have been collaborating on the development of SNOMED-CT since April of 1999. Alpha testing started in 2001 [24].

July 1, 2003, the CAP signed a US\$32.4 million, five-year sole source contract with the National Library of Medicine to license English and Spanish language editions of SNOMED-CT [25]. The agreement provides free-of-charge access to SNOMED-CT core content and all version updates, starting in January 2004, to qualifying entities through the NLM's UMLS. Qualifying entities include US federal agencies, state and local government agencies, territories, the District of Columbia, and any public, for-profit and non-profit organization located, incorporated and operating in the US. In April 2007, SNOMED-CT was acquired by the International Healthcare Terminology Standards Development Organization (IHTSDO) based in Copenhagen.

SNOMED-CT is based on "*concepts*." Each concept represents a unit of thought or meaning and is labeled with a unique identifier. Each concept has one or more terms linked to it that express the concept by means of natural language strings. Each concept is interrelated to other SNOMED concepts with which it is logically connected. These relationships are used to provide a computer readable description, and sometimes a definition, of the concept. These connections allow SNOMED-CT to be searched, retrieved, reused or analyzed in a variety of ways. Hierarchical relationships define specific concepts as children of more general concepts. For example, "kidney disease" is defined as a kind of "disorder of the urinary system." In this way, hierarchical relationships provide links to related information about the concept. As of January 2008, SNOMED-CT contains 378,111 health care concepts organized into hierarchies, with approximately 1.36 million relationships between them, and more than 1,068,278 associated natural language terms. SNOMED-CT is officially available in English and Spanish language editions.

The main merits of SNOMED-CT for *clinical* documentation are its broad terminological coverage, which has been demonstrated repeatedly in the course of its development and in various application areas [26-32].

Despite these positive assessments of the performance of SNOMED-CT as concerns coverage, there are also negative assessments along primarily three lines: term formation principles, SNOMED-CT as an ontology, and practical usefulness. There is a lot of literature on poor consistency of coding, too. [31, 33]

In [34], Ceusters *et al.* identified several sources of confusion and ambiguity on the basis of an analysis of the procedure axis of SNOMED International (1998). Bodenreider *et al.* used lexical techniques to study the (in)consistent use of modifiers such as “bilateral”/“unilateral” and “congenital”/“acquired” in SNOMED International [35]. While every occurrence of “bilateral X” or “congenital X” would call for a “unilateral X” and “acquired X” respectively, but this requirement was met in very few cases. Elkin *et al.* concluded that “*The current implementation of SNOMED-RT does not have the depth of semantics necessary to arrive at comparable data or to algorithmically map to classifications such as ICD-9-CM*” [36].

In [37] serious problems associated with using SNOMED-CT as an ontology instead of a terminology, i.e. for reasoning, were highlighted. SNOMED-CT organizes terms according to a minimalist model and (during the design phase) lets a description logic compute whether statements are consistent with the model. This does not guarantee however that statements are consistent with reality nor is it a safeguard against semantic inadequacy of the labels: often, users when accessing a term (e.g. via a browser) attach to it a meaning that is not consistent with the formal statements through which the term is defined [37-38].

Schulz recently provided evidence that SNOMED implicitly supports at least three different kinds of ontological commitments running in parallel but not clearly separated [39], viz. (i) to independently existing entities, (ii) to representational artifacts, and (iii) to clinical situations. His analysis shows how the truth-value of a sentence changes according to which of these perspectives is employed. He argues that a clear understanding of the kind of entities in reality denoted by SNOMED CT’s concepts is crucial for its proper use and maintenance.

3.3 Third generation systems: ontologies

Ontologies are currently a hot research topic in healthcare and life science, their main purpose being precisely, or so it is hoped, to assure semantic interoperability of systems. More than in other domains, it seems, there is then a divide between two groups of researchers, the first approaching this issue exclusively from an information science and software engineering perspective, the second taking a stance that is informed by philosophical considerations relating to the use and mention of terms and to the relations between terms and entities in reality.

The former group understands by ‘ontology’ the formal representation of some ‘conceptualization’ of an application domain [40]. An ontology thus consists of a first order vocabulary with a precise model-theoretic semantics and formal definitions for its terms, the later standing for concepts and their interrelationships in the corresponding application domain. An ontology in this sense is a contribution to knowledge representation, and as such draws on earlier frame- and semantic network-based approaches. The ontologist works with minimalist ‘models’, that are then used as templates to stand in for those parts of reality that fit the model (hence you can only see what the model allows you to see). The models are usually implemented by means of some form of description logic (DL).

The key characteristic features of description logics reside in the constructs for establishing relationships between *concepts* by means of *roles* [41]. Concepts are given a set-theoretic

interpretation: a concept is interpreted as a set of individuals, and roles are interpreted as sets of pairs of individuals. The domain of interpretation can be chosen arbitrarily, and it can be infinite. In this context, it is important to understand that '*Model-theoretic semantics does not pretend, and has no way to determine what certain words and statements "really" mean. (...) It offers no help in making the connection between the model (the abstract structure) and the real world*' [42] (pp. 30-31).

It is this lack of explicit reference that disturbs those researchers who take an analytical-philosophical stance, and for whom the term 'ontology' denotes rather a representation of reality. This 'realist' community argues that an ontology should correspond to reality itself, in a manner that maximizes descriptive adequacy while at the same time conforming to the constraints of formal rigor and computational usefulness. By 'ontology' they mean: a representation of some pre-existing domain of reality which (1) reflects the properties and relations of objects within its domain in such a way that there is a systematic correlation between reality and the representation itself, (2) is intelligible to a domain expert, and (3) is formalized in a way that allows it to support automatic information processing.

Corresponding to these two understandings of 'ontology', there is a dichotomy also in the way terminologies are conceived. For the first group a terminology is seen as a class of systems, either in the form of a printed text or in some digital form, that contain the terms which specialists in a specific domain are supposed to use when exchanging information. The purpose of a terminology from this perspective is then two-fold: it is to allow the unambiguous understanding of what is conveyed, and to stabilize as much as possible the terminology within a specific domain. The realist group then adds a third requirement, namely that there should be, for each term, some referent in reality – and more specifically that for each general term there should be instances in reality which can be located in space and time and for example observed in the lab or clinic. The formal systems in which ontologies are expressed, including the OWL Web Ontology Language, are structured in such a way that they can allow the ontology to be used for reasoning about such instances.

With respect to *patient data*, for example, the ontological approach enables explicit reference to be made not only to the types of entities referred to in an EHR, but also to the instances (particular cases) of these types existing in a given patient at a given time. The EHR itself, given current technologies, will in many cases refer only implicitly to these real instances. Thus it will refer to the existence of *some* fracture in a nasal bone, rather than to *this particular fracture* #1 caused in *this particular fall* (#2) and involving *this particular seizure* (#3). In this way it will be able to describe in a formally rigorous way the relationships between the entities involved [43], and use this information then for purposes of reasoning.

One principal advantage of the realist approach is that it is often the case that the same entities which are represented in clinical ontologies – for example cells, tumors, lesions – are often represented also in the field of biological ontologies, where the realist approach has thus far proved most successful. In this way some of the benefits, including tested best practices, of the realist approach can be transferred into the clinical domain; in this way, too, the path is cleared for the creation of the ontological and terminological resources adequate to the personalized medicine and translational bioscience research of the future.

A good biomedical ontology reflects the most general categories in the corresponding domain of reality, i.e. those categories in terms of which the biomedical data is organized. Unfortunately,

most ontologies in biomedicine are marked by a number of serious defects when assessed in light of their conformity to both terminological and ontological principles. [44-45] This means that much of the information formulated using such ontologies remains implicit to both human interpreters and software tools. Vital opportunities for enabling access to the information in such systems are thereby wasted. These defects manifest themselves in difficulties encountered when the underlying resources are used in biomedical research. Such defects are destined to raise increasingly serious obstacles to the automatic integration of biomedical information in the future, and thus they present an urgent challenge to research.

The major overarching challenge to be met by ontology is thus two-fold: (1) to bridge the gap between clinical research conclusions and the need to make personal decisions in healthcare and (2) to bridge the gap between data models evolved separately in the two discrete worlds of healthcare and bioinformatics.

3.3.1. Open Biomedical Ontologies

Open Biomedical Ontologies is an umbrella organization for well-structured controlled vocabularies for shared use across different biological and medical domains. It includes concept-based ontologies such as the Gene Ontology [46] and MGED [47].

Within the Open Biomedical Ontologies (OBO) framework [48], it has now been agreed upon that contributing ontologies are to be constructed in line with the OBO Relationships Ontology whose foundations are laid down in [49]. This standardization initiative is called the OBO Foundry.

The OBO Foundry is a collaborative experiment, involving a group of ontology developers who have agreed to the adoption of a growing set of principles specifying best practices in ontology development. These principles are designed to foster interoperability of ontologies within the broader OBO framework, and also to ensure a gradual improvement of quality and formal rigor in ontologies, in ways designed to meet the increasing needs of data and information integration in the biomedical domain.

Ontologies are admitted into the Foundry, and to its on-going process of peer review, only if their developers commit to the acceptance of a set of common principles [12], of which the most important for our purposes are:

- that terms and definitions should be built up compositionally out of component representations taken either from the same ontology or from other, more basic, feeder ontologies (methodology of cross-products);
- that ontologies should use upper-level categories drawn from Basic Formal Ontology [50] together with relations unambiguously defined according to the pattern set forth in the OBO Relation Ontology [51];
- that for each domain there should be convergence upon exactly one Foundry ontology (principle of modularity) [52].

4 APPENDIX: STANDARDIZATION FROM A US PERSPECTIVE

The overall objective of standardization is to facilitate the production, handling, and use of products or services within a framework of free trade and of the free market in such a way as to satisfy to the maximal degree possible both users and suppliers [53]. The operational goal of standardization is to provide sets of consistent specifications – called “standards” – to be shared by all parties manufacturing the same products, or providing the same services, standards which will then form the basis for further developments.

Standards should be rooted in the consolidated results of science, technology and experience, and aimed at the promotion of optimum community benefits.

Standards may derive from various processes, but in most cases they result from a voluntary process initiated by important actors in a domain to bring order and clarity and to establish a common base for market development. Typically, this process involves both suppliers of products and their customers. Standardization in many sectors has been dominated by suppliers, but increasingly the development of standards is under pressure from end users (the ‘consumers’), or even initiated by them. This is particularly the case nowadays for health information technology (HIT).

Under no circumstances, however, should standards be used to keep new and novel paradigms and products from the market. The recent evolution of the HIT arena suggests that there is a danger that standards will indeed be used in just this way. This is in particular a risk when large and powerful companies or organizations have an over-large impact on the development of standards: they tend to accept only those standards with which their products are compatible and as a consequence hamper not only the implementation of better and more advanced paradigms, but also the advance of science and thereby also the advance in our understanding of human health and disease [54-55].

4.1 US formal standardization

The National Standards Body (NSB) for the United States is the American National Standards Institute (ANSI). It is a private, non-profit organization founded in 1918 that administers and coordinates the U.S. voluntary standardization and conformity assessment system [56]. The Institute’s mission is to enhance both the global competitiveness of U.S. business and the U.S. quality of life by promoting and facilitating voluntary consensus standards and conformity assessment systems, and safeguarding their integrity.

It is worth noting that the standardization activities in the U.S. differ from the formal standardization process in Europe. In fact, ANSI itself does not develop national standards; instead, it delegates the production of standards to accredited Standards Developing Organisations (SDO). It is for this reason that standards developed by SDOs can become automatically formal standards in the US, but not in European countries.

In order to maintain ANSI accreditation, standards developers are required to adhere consistently to a set of requirements or procedures known as the “*ANSI Essential Requirements: Due process requirements for American National Standards*” that are laid down as governing the consensus development process [1].

The process of creating voluntary standards within each SDO is guided by ANSI's cardinal principles of consensus, due process and openness and depends heavily upon data gathering and compromises among a diverse range of stakeholders. ANSI ensures that access to the standards process, including an appeals mechanism, is made available to anyone directly or materially affected by a standard that is under development. The standards developed by a SDO according to the ANSI rules become '*ANSI standards*'.

ANSI is the sole U.S. representative and dues-paying member of the two major non-treaty international standards organizations, the International Organization for Standardization (ISO) [57], and, via the U.S. National Committee (USNC), the International Electrotechnical Commission (IEC) [58].

4.2 US Standard Development Organizations

4.2.1. Health Level Seven, HL7

HL7 (Health Level Seven) was founded in 1987 by several vendors of software for the healthcare industry, with the assistance of academics and major Health Maintenance Organizations. Their goal was to develop consensual messages formats to facilitate better interoperability within and between Hospital Information Systems (HIS).

In 1994, HL7 was accredited by ANSI as an SDO, meaning that HL7 approved specifications are automatically channeled into the official, global standardization process as formal American National Standards.

Version 1.0 of the "HL7 Standard" message specifications was approved in 1987, and was followed by version 2.0 in 1998. Subsequently, version 2 has itself evolved through a succession of modified releases. It still forms the basis for the many HIS systems implemented in the US and many European countries.

Version 3 message specifications use a formal Message Development Framework methodology, employing what is called the Reference Information Model (RIM), which developed to help make messages more consistently implemented than they are for Version 2 [59]. The RIM is now a major focus of current interest in HL7. The large task of forming an object model of basic building blocks for all health information is now considered by HL7 to be complete and mature enough to be recommended for productive use, even though the RIM, and specifically its documentation, have been found to contain several fundamental flaws [54-55]. Nevertheless, the RIM has been accepted as an ISO Standard, without there being thus far successful implementations of the version 3 HL7 standard that is built on its basis in operational systems.

4.2.2. Digital Imaging Communications in Medicine, DICOM

In 1983 the American College of Radiology (ACR) [60] and the National Electrical Manufacturers Association (NEMA) [61] formed a joint committee in order to standardize a method for the transmission of medical images and associated metadata. In 1985 this committee published the ACR-NEMA Standards Publication No. 300-1985. Version 2.0 was published in 1988. In 1993 version 3.0 marked a major step toward a standard method of communicating digital image information. It also introduced the name DICOM [62] (Digital Imaging and Communications in Medicine).

DICOM is now an international standards organization creating and maintaining standards for the communication of biomedical diagnostic and therapeutic information in disciplines using digital images and associated data. It has liaison A status with ISO/TC215. Its secretariat is administered by the NEMA Diagnostic Imaging and Therapy Systems Division along with 9 professional societies that assume working group secretariats.

Relevant in the context of semantic interoperability are DICOMs WG-08 on ‘Structured Reporting’ and WG-20 on ‘Integration of Imaging and Information Systems’.

The current priorities for DICOM [63] are issues relating to security, performance, new modality technology, structured and coded documents for specific clinical domains, and workflow management.

4.2.3. The Institute of Electrical and Electronics Engineers (IEEE)

The IEEE resulted from the merging in 1963 of the AIEE (American Institute of Electrical Engineers) and the IRE (Institute of Radio Engineers), and through these predecessors thus dates back to 1884 [64]. AIEE addressed wire communications, light and power systems, while IRE, itself resulting from the merger of two largely local organizations (the Society of Wireless and Telegraph Engineers and the Wireless Institute), addressed wireless communications.

Relevant in the context of semantic interoperability in health care are the standards:

- IEEE 11073, ***Standard for Medical Device Communications***: a family of documents that defines the entire seven layer communications requirements for the Medical Information Bus (MIB). This is a robust, reliable communication service designed for Intensive Care Unit, Operating Room, and Emergency Room bedside devices;
- IEEE 1157, ***Standard for Health Data Interchange***: a family of documents that define the communications models for medical data interchange between diverse systems. This effort has been called “MEDIX”. The common data model currently being worked on by members of the ANSI Healthcare Informatics Standards Board (HISB) is part of this effort.

4.2.4. The American Society for Testing Materials (ASTM)

ASTM International [65], formerly the American Society for Testing Materials, is another body which develops standards under ANSI. It was founded in 1898 and today forms a global forum for the development and publication of voluntary consensus standards for materials, products, systems, and services. Individuals (over 30,000 from 100 nations), rather than corporate entities, are members. These members include producers, users and consumers as well as representatives of government and academia.

ASTM/E31 is the technical committee responsible for Healthcare Informatics. It has published several standards that in turn inspired a variety of international standards.

Most recently, ASTM has balloted, and passed, a standard for the Continuity of Care Record (CCR). This is a family of XML-format message types with the original use of supporting electronic patient care referrals transmitted between healthcare providers. The CCR is now seen as having archival value within an Electric Health Records repository. In July 2004, ASTM agreed to harmonize CCR with the Clinical Document Architecture (CDA) developed

independently by HL7. The result is an implementation guide to the so-called Clinical Care Document (CCD).

4.2.5. *OMG/CORBA*

The Object Management Group (OMG) is developing several services for health care [66] as part of its CORBA initiative (Common Object Request Broker Architecture), which aims to produce an open, vendor-independent specification for an architecture and infrastructure that will allow software components written in multiple computer languages to work together. Relevant CORBA specifications are the Terminology Query Service (TQS) and Clinical Observations Access Service (COAS).

4.3 US agencies and governmental initiatives with an impact on semantic interoperability work in the eHealth domain

4.3.1. *ONCHIT - Office of the National Coordinator for Health Information Technology*

On April 27, 2004 President Bush committed the US to pursuing the goals of reducing medical errors, lowering medical costs, and providing better information for consumers and physicians through a commitment to Health Information Technology. This Executive Order [67] directed Health and Human Services (HHS) Secretary Mike Leavitt to establishing the position of the Office of the National Coordinator for Health Information Technology (ONCHIT) [68].

ONCHIT aims to provide leadership for the development and nationwide implementation of an interoperable health information technology infrastructure to improve the quality and efficiency of health care and the ability of consumers to manage their care and safety.

June 3, 2008, ONCHIT released its plan for 2008-2012 [69]. The Plan has two goals: Patient-Focused Health Care, and Population Health, with four objectives under each goal. The theme of interoperability recurs across the goals, but applies in different ways to individual healthcare and to population health: for the former the goal is to *enable the movement of electronic health information to where and when it is needed to support individual health and care needs*; for the latter, to *enable the mobility of health information to support population-oriented uses*.

4.3.2. *CHI – Consolidated Health Informatics (US e-government plan)*

Through its Consolidated Health Informatics (CHI) initiative [70-71], the US government is establishing a portfolio of existing clinical vocabularies and messaging standards with the goal of enabling federal agencies to build interoperable health data systems. It is hoped that the standards will enable all federal agencies to “speak the same language” and share information without the high costs of translation or data re-entry. Federal agencies could then pursue projects addressing their individual business needs while at the same time serving larger goals such as sharing electronic medical records and electronic patient identification. CHI standards work in conjunction with the Health Insurance Portability and Accountability Act (HIPAA) transaction records and code sets and HIPAA security and privacy provisions.

About 20 departments/agencies including HHS, VA, DOD, SSA, GSA, and NIST are active in the CHI governance process, one effect of which is that federal agencies are incorporating the

adopted standards into their individual agency health data enterprise architectures, either by building new systems or by modifying systems which already exist.

4.3.3. American Health Information Community – (AHIC)

On September 13, 2005, HHS Secretary Leavitt announced the membership of the American Health Information Community (AHIC) [72], which was formed to help realize President Bush's call for most Americans to have electronic health records within ten years. AHIC is a federally chartered commission that provides input and recommendations to HHS on how to make health records digital and interoperable while ensuring that the privacy and security of these records are protected in a straightforward market-led way.

The development of recommendations by AHIC proceeds by identifying key work areas considered to have potential for breakthroughs in the furthering of standards that will lead to interoperability of health information. Once a work area is identified a corresponding work group is formed with the task of framing a use case to provide detailed guidance on the functions needed to advance critical efforts for the accelerated adoption of health information technology.

To date, seven AHIC workgroups have been created of which the following deal with semantic interoperability issues:

- Population Health and Clinical Care Connections: to make recommendations to AHIC so that essential ambulatory care and emergency department visit, utilization, and lab result data from electronically enabled health care delivery and public health systems can be transmitted in standardized and anonymized format to authorized public health agencies within 24 hours.
- Electronic health records: to make recommendations to AHIC so that standardized, widely available and secure solutions for accessing current and historical laboratory results and interpretations are deployed for clinical care by authorized parties.
- Quality: to make recommendations to AHIC that specify how certified health information technology should capture, aggregate and report data for a core set of ambulatory and inpatient quality measures.
- Personalized Healthcare: to make recommendations to AHIC for a process to foster a broad, community-based approach to establishing a common pathway based on common data standards to facilitate the incorporation of interoperable, clinically useful genetic/genomic information and analytical tools into electronic health records to support clinical decision-making for the clinician and consumer, and to make recommendations to the AHIC to consider means to establish standards for reporting and incorporation of common medical genetic/genomic tests and family health history data into electronic health records, and provide incentives for adoption across the country including by federal government agencies.

AHIC works through the development of use case descriptions, which provide a narrative and graphical description (a storyboard with figures and diagrams) of the behaviors of persons or things (actors), and/or a sequence of actions, in a targeted area of interest (domain) [73].

4.3.4. Healthcare Information Technology Standards Panel (HITSP)

The mission of HITSP, the Healthcare Information Technology Standards Panel [3], is to serve as a cooperative partnership between the public and private sectors for the purpose of achieving a widely accepted and useful set of standards specifically to enable and support widespread interoperability among healthcare software applications as they will interact in a local, regional and national health information network. HITSP is comprised of a wide range of stakeholders. It assists in the development of the Nationwide Health Information Network (NHIN) by addressing issues such as privacy and security within a shared healthcare information system.

The Panel is sponsored by ANSI in cooperation with strategic partners such as the Healthcare Information and Management Systems Society (HIMSS) [74], the Advanced Technology Institute (ATI) and Booz Allen Hamilton. Funding for the Panel is being provided via the ONCHIT-1 contract award from the U.S. Department of Health and Human Services.

The standardization process is based on four iterative functions: use case development, gap analysis, process implementation guidelines development and testing.

The identification of standards by HITSP is recorded in interoperability specifications that specify how and what standards should be used for given use cases.

4.3.5. NCR - National Center for Research Resources

The NCR [75] was formed on February 15, 1990 when the then Secretary of HSS, Dr. Louis W. Sullivan, approved the merger of the HSS Division of Research Resources and the NIH Division of Research Services. The mission of the NCR is to support laboratory scientists and clinical researchers with the environments and tools they need to make biomedical discoveries, translate them to animal based studies and ultimately apply them to patient-oriented research.

The NCR consists of four divisions. The Division of Biomedical Technology Research and Research Resources supports research, training and access to state-of-the-art technologies in both instrumentation and software. The Division of Clinical Research Resources seeks to enhance translational medicine. The Division of Comparative Medicine supports research in the development of new biologic models. The Division of Research Infrastructure provides competitive funding to modernize and construct research laboratories.

One initiative within the Clinical Research Division is the CTSA program, which is building a consortium of institutions “designed to speed the process by which biomedical discoveries are translated into effective medical care for patients.” This goal is being realized through the granting of Clinical and Translational Science Awards [76], which are designed to enable institutions to develop the resources for integrating clinical care and research science across multiple disciplines and academic departments, schools, clinical and research institutes, and hospitals, CTSA's are expected to transform the conduct of translational medicine in the United States. A major hurdle in the way of accomplishing this goal is the integration of data from patient care systems with data from clinical research systems.

4.3.6. USHIK - United States Health Information Knowledgebase

The United States Health Information Knowledgebase (USHIK) [77] is a project funded by the Agency for Healthcare Research and Quality (AHRQ) [78] with management support from the

Centers for Medicare & Medicaid Services (CMS). The USHIK is a metadata repository of health information data elements, including their definitions, permitted values and source information models. The intent of the knowledgebase is to provide a means for healthcare organizations to synchronize their local information systems around healthcare standards as benchmarks. The methodology used to format the knowledgebase is said to be “based upon” the ISO/IEC 11179 “Specification and Standardization of Data Elements” standard.

The USHIK web interface allows for the browsing of information models and data elements. Comparisons between data elements are provided in the form of a matrix listing the meta-data for a set of elements selected by the user. Search capabilities include filtering results by registration authority, data element type, submitting organization, responsible organizations, registration and administration status, and text searches on component name, definition, permissible value and value meaning.

While useful for manual search and comparison of data elements, the lack of a tool set makes the repository of limited use for developers facing the need to synchronize large information models.

4.3.7. CMS - Centers for Medicare & Medicaid Services

The CMS Centers for Medicare & Medicaid Services [79] is the HSS agency responsible for administering the Medicare program and working with state governments to administer the Medicaid and State Children’s Health Insurance Program. With a budget of \$650 billion and 90 million beneficiaries the CMS plays a prominent role in the overall direction of the US healthcare system. In terms of standardization, two programs governed by the CMS are of significance: the Medicaid Information Technology Architecture, and the International Classification of Diseases Clinical Modifications and Procedure Classification System developed and maintained in conjunction with the CDC’s National Center for Health Statistics (NCHS) (see further under 4.3.7.2 below).

4.3.7.1 MITA - Medicaid Information Technology Architecture

The MITA Medicaid Information Technology Architecture project has the goal of transforming the business and information technology of the Medicaid enterprise. Its goal is a set of guidelines on which a national architecture of information systems can be built that will improve both the quality and efficiency of health care. Critical to the success of this project is the adoption of data standards, and the MITA initiative will coordinate the identification and use of common data standards for the Medicaid enterprise.

In March of 2006, CMS released the Medicaid IT Architecture Framework 2.0 [80]. While no data standards had been selected at that time, a methodology for adopting standards was defined as well as listings of standards that would either be required by the Health Insurance Portability and Accountability Act (HIPAA) or were seen as emerging potential candidates for standards.

4.3.7.2 International Classification of Diseases 9th and 10th Editions

The CMS and the NCHS [81] oversee the maintenance and production of the ICD-9-CM (volumes 1, 2, and 3) and the ICD-10-CM diagnostic codes and ICD-10-PCS treatment procedure codes. The NCHS is part of the Coordinating Center for Health Information and Services (CCHIS), one of the six coordinating centers of the Centers for Disease Control and Prevention (CDC).

The International Classification of Diseases coding system is published by the World Health Organization (WHO). ICD-9 was released by WHO in 1977. The ICD-9-CM is the official vocabulary used for billing and reimbursement purposes in the United States in conjunction with the UB-92 reimbursement form for hospitals and the HCFA-1500 form for physicians (see 4.4.1 below). One or more diagnostic codes together with codes for the related treatment procedures are submitted to payers, who then match the submission to determine the amount of reimbursement.

4.3.8. Centers for Disease Control and Prevention – CDC

In addition to the NCHS contribution to the development and maintenance of the US enhancements of the ICD coding systems discussed above, the CDC is also advancing the use of standardized health information through its Public Health Information Network (PHIN) initiative [82].

The PHIN is a national initiative to improve the capacity of public health to use and exchange information electronically by promoting the use of standards and defining technical requirements. The CDC specifies the role of the PHIN as:

- supporting the exchange of critical health information between all levels of public health and healthcare,
- developing and promulgating requirements, standards, specifications, and an overall architecture in a collaborative, transparent, and dynamic way,
- monitoring the capability of state and local health departments to exchange information,
- advancing supportive policy,
- providing technical assistance to allow state and local health departments to implement PHIN requirements,
- facilitating communication and information sharing within the PHIN community,
- providing public health agencies with appropriate and timely information to support informed decision making, and
- harmonizing with other federal initiatives.

In collaboration with state and local health departments CDC has created a set of applications that include:

- PHIN Messaging Services, which are definitions of message specifications and mapping guides that support specific public health business needs.
- PHIN XForms Question Framework, which defines and distributes standardized forms for public health practices based on a library of reusable, standard encoded questions.
- PHIN Vocabulary Services, which includes a Web-based enterprise vocabulary system (PHIN VADS) for accessing, searching, and distributing the vocabularies used within the PHIN.

The focus of the Messaging Services Team [83] is to create standardized messages for the domains of public health case reporting, biosurveillance, and laboratory processing for public

health use. At present, draft versions are available. As of June 2006, the messaging exchange standard was revised from HL7 Version 3 messages to HL7 Version 2.5 messages. The stated reason for the revision was to allow for the exchange of messages among a wider user base.

The components of the PHIN XForms Question Framework [84] are:

1. A question repository built from examination of public health forms used by states and local health departments for selected Nationally Notifiable Conditions. The repository includes value sets for questions which are bound to standard vocabularies.
2. Data Models – The information model includes metadata about the forms such as questions with answer value sets, question sets, and form segments, the data collection forms built from those components, and the default bindings to a generic Public Health Information Model.
3. An XForms framework utilizing a model-view-controller (MVC) pattern as the technology used to bind the question set vocabulary to the public health forms, collect and validate form data, and submit forms for processing.

Future Plans for the PHIN XForms Question Framework are stated to include the creation of a graphical user interface to author XForms based on a Question Repository, ontology-driven question search capabilities, a library of reusable, version-controlled forms and a definition of a Public Health Document Architecture.

The purpose of the PHIN VADS [85] is to provide standard vocabularies relevant to public health to the CDC and its partners. There are currently 267 value sets and approximately 700,000 concepts in the PHIN VADS. The selection of vocabularies is based upon the recommendations of CHI. Files can be downloaded in a variety of standard formats including tab-delimited, Excel, or XML.

4.3.9. Public Health Data Standards Consortium – PHDSC

The PHDSC Public Health Data Standards Consortium [86] is a non-profit membership-based organization of federal, state and local health agencies; national and local professional associations; academia, public and private sector organizations; international members, and individuals. Currently the PHDSC is comprised of 36 member organizations.

The mission of the PHDSC is to represent the public health community to the standards development organizations and to promote the use of data and systems standards by the public health community. This mission is accomplished by the PHDSC working in collaboration with SDOs to implement existing standards, modify standards to the needs of public health and research and, if needed, to develop new standards. Examples of PHDSC collaborations include membership in HITSP and participation in the standards development process of HL7, the Accredited Standards Committee (ASC) X12, the National Unified Billing Committee (NUBC) [87] and the National Uniform Claim Committee (NUCC) [88].

The organizational structure of PHDSC is headed by a 35 person board of directors which oversees the operations of 5 program areas: Data Standards, Privacy, Security & Data Sharing, Professional Education, Nationwide Health Information Network, and Communication and Outreach.

Of primary relevance to the present discussion are the efforts of the Data Standards Committee which coordinates data standards activities for the PHDSC through the following three Sub-Committees:

- Sub-Committee on Health Care Services Data Reporting (HCSDR) Guide
- Sub-Committee on Payer Typology, which developed and maintains a payer typology to allow for consistent reporting of payer data to public health agencies for health care services and research.
- Sub-Committee on External Cause of Injury Codes (ECIC), which is working on developing an educational strategy concerning the importance of external causes of injury codes when transmitting chief complaint data from emergency rooms to state and local health agencies.

The mission of the HCSDR Guide Committee [12] is to create and maintain an implementation guide for reporting health care service data. The result of the sub-committee's efforts is the ANSI X12N 837 Health Care Service Data Reporting Guide, which provides a standardized format and data content for reporting health care service data compatible with the 837 Health Claim transaction set standards identified by HIPAA). In addition, the guide includes data elements, for example pertaining to race and ethnicity, or patient county codes, that are not now needed for the payment of a claim and so are missing from the industry claims standard. The Guide includes these additional data elements as they are critical to quality, utilization, and public health studies.

The mission of the Payer Typology Committee [89] is to create a payer type standard to allow consistent reporting of payer data to public health agencies for health care services and research. This committee was formed in response to the current lack of a standard classification of the sources of payment data and is an acknowledgement of the fact that having such a standard is critical for examining the effects of payment policies.

The committee has created the Source of Payment Typology, which is said to have an organizational structure similar to that of the ICD classification system. This typology identifies general payer categories that and subsume related subcategories that are more specific. The users of this typology are permitted to add more specific categories as needed for their unique payment systems.

The External Cause of Injury Code (ECIC) Committee [90] has the mission of promoting the collection and reporting of standardized external cause of injury codes by health care providers. This mission is aligned with the national objective of measuring progress on injury and violence prevention and control, an objective whose realization is currently hindered by the lack of standardized external cause of injury codes.

4.4 Other organizations and initiatives

4.4.1. National Uniform Billing Committee - NUBC

The NUBC National Uniform Billing Committee [91] was formed in 1975 by the American Hospital Association (AHA) with equal representation from provider organizations (e.g. AHA and its state affiliates, the Healthcare Financial Management Association and the Federation of

American Health Systems) and payer organizations (e.g. the Health Care Financing Administration (HCFA), Medicaid, CHAMPUS, the Blue Cross and Blue Shield Association (BCBSA) and the Health Insurance Association of America (HIAA)). The Group Health Association of America/American Managed Care and Review Association (GHAA/AMCRA) has more recently become a member also.

NUBC was formed with the objective of developing a single billing form and standard data set that could be used for handling health care claims by institutional providers and payers throughout the U.S. UB-82, the first such form, was produced in 1982. When the NUBC established the UB-82 data set design and specifications, it also imposed an eight-year moratorium on changes to the structure of the data set design. After the expiration of this moratorium the UB-92 was created, which incorporated much of the form and content of the UB-82 but included changes designed to further reduce the need for attachments. Currently, more than 98% of hospital claims are submitted electronically to the Medicare program using the UB-92.

The data elements included on the form are those the NUBC deems as being necessary for claims processing. Each element is then assigned a designated space on the form and each such space is assigned a unique numeric identifier. Other elements that are occasionally needed are incorporated into general fields that utilize assigned codes, codes and dates, and codes and amounts. The Code Sets created and maintained by NUBC [92] for these purposes are:

- Admission Source and Type Codes representing the priority and the source of an admission.
- Discharge Status / Patient Disposition Codes indicating the patient status as of the ending service date.
- Condition Codes used to identify conditions relating to a bill that may affect payer processing, such as whether a patient is homeless.
- Occurrence Codes with associated dates, used to refer to significant events relating to this bill that may affect payer processing, such as an auto accident.
- Occurrence Span Codes with associated dates used to identify an event that relates to the payment of a claim, such as Skill Nursing Facility level of care dates.
- Revenue Codes identifying a specific accommodation, ancillary service or billing calculation, such as emergency room charges.
- Value Codes which relate amounts or values to identified data elements necessary to process a given claim as qualified by the payer organization

4.4.2. The Certification Commission for Healthcare Information Technology (CCHIT)

CCHIT is an independent, voluntary, private-sector initiative for the certification of electronic health records and their networks [93]. The initiative's mission is to accelerate the adoption of health information technology by creating an efficient, credible and sustainable certification program. CCHIT was formed in 2004 by the American Health Information Management Association (AHIMA) [94], the Healthcare Information and Management Systems Society (HIMSS) [74] and the National Alliance for Health Information Technology (Alliance) [95]. In

the following year, additional funding was supplied by the American Academy of Family Physicians (AAFP) [96], the American Academy of Pediatrics (AAP) [97], the American College of Physicians (ACP) [98], the California Healthcare Foundation (CHCF) [99], the Hospital Corporation of America, McKesson, Sutter Health, United Health Foundation, and WellPoint Inc.

In 2005, CCHIT was contracted by the HHS to develop the certification criteria and validation process for Electronic Health Records (EHRs). CCHIT is governed by a Board of Commissioners which oversees the work of its professional staff and voluntary workgroups. The workgroups focus on creating the products of the commission – criteria covering health information technology product functionality, interoperability, and security.

The 2007 Ambulatory Interoperability Criteria [100] show their a close alignment with the HITSP specifications which were based in turn on the CHI recommendations.

Recently, CCHIT lost its monopoly position when other accreditation bodies for electronic healthcare records were accepted.

4.4.3. IHE — Integrating the Healthcare Enterprise

Started by HIMSS and RSNA, the Integrating the Healthcare Enterprise (IHE) initiative is a spontaneous undertaking organized to improve the integration of systems [101]. It aims at providing a process for a co-ordinated adoption of standards: clinicians and IT staff define needs; vendors develop solutions in the form of technical frameworks which can advance coordination. In 2004, 50 vendors were involved in the USA, 34 in Asia, and 58 in Europe. Professional societies (ECR, BIR, DRG, SIRM, HIMSS/RSNA, etc.) supervise documentation, testing, demonstration, and promotion. Partnerships now also exist with the American College of Cardiology (ACC), American College of Clinical Engineering (ACCE), HL7, and DICOM, and several individual members take part as well.

The needs for the IHE initiative comes from the recognition that standards are necessary but not sufficient for seamless implementations: they are not ‘plug and play’, as each interface requires site specific analysis and configuration in result of which the standards may be costly to implement and to maintain. IHE delivers integration profiles built on existing standards. IHE makes it clear that it is itself not a standards development organization. It uses existing standards (so far DICOM, HL7, Internet, Oasis, etc.) to address specific clinical needs. Its activity is to be regarded as complementary to that of the SDOs.

An IHE Integration Profile organizes a set of frameworks for coordinated, standards-based transactions among the corresponding functional components of health organizations in order to address a specific clinical or infrastructure need. IHE develops such solutions for IT systems integration in a stepwise and pragmatic manner, focusing on the most common integration challenges. It has developed close to 30 Integration Profiles focused on Radiology, Laboratory, IT Infrastructure (MPI, Security, etc.) and Cardiology and Medication.

IHE has established several chapters in Europe, including France, Italy, Germany, UK, Spain, Netherlands, Denmark, and Norway.

4.4.4. National Uniform Claim Committee - NUCC

The NUCC National Uniform Claim Committee [102] is chaired by the American Medical Association and consists of 12 voting members, including HCFA, the Alliance for Managed Care, ANSI ASC X12N, BCBSA, AAHP, HIAA, Medical Group Management Association, National Association of Insurance Commissioners, National Association of Equipment Services, National Association of State Medicaid Directors, and NUBC. With a mission similar to that of the NUBC, the NUCC develops the claims form for the non-institutional health care community. Its product, the HCFA 1500, is the major vehicle for collecting the Uniform Ambulatory Care Data Set (UACDS). The goal of the NUCC is for the uniform claim to be equivalent across products, contracts and government programs.

4.4.5. Clinical Data Interchange Standards Consortium (CDISC)

The CDISC Clinical Data Interchange Standards Consortium is a not-for-profit organization founded in 1997. Its mission is to develop global, platform-independent data standards that enable information system interoperability to improve medical research and related areas of healthcare [103]. The CDISC organization is led by a governing body, board of directors and industry advisory board. The CDISC working groups are staffed by volunteers from all segments of the biotechnology and pharmaceutical industries as well as government and academic organizations. There are now seven working groups within the CDISC organization [104]: the Submission Data Standards (SDS) team, the Analysis Dataset Model (ADaM) team, the Operational Data Model (ODM) team, the Laboratory (LAB) team, the SEND team, the Protocol Representation group and the Terminology team.

Some highlights of the acceptance and use of CDISC standards include the SDTM, which was selected by the FDA in 2004 as the recommended standard for submitting clinical trial data for regulatory submissions. In the same year, a survey showed a nearly 50% utilization rate by North American pharmaceutical companies of at least one CDISC standard.

4.4.6. Biomedical Research Integrated Domain Group (BRIDG)

The BRIDG project [105] is a collaborative effort of the Clinical Data Interchange Standards Consortium (CDISC) [103], the HL7 Regulated Clinical Research Information Management Technical Committee (RCRIM TC) [106], the National Cancer Institute (NCI) [14], and the US Food and Drug Administration (FDA) [107]. It was formed in 2004 as one result of the cancer Biomedical Informatics Grid (caBIG) initiative to develop a structured protocol representation that could be used to exchange clinical trial protocol information. The project goal is to provide a platform for interoperability amongst existing standards and to develop new standards in the domain of clinical research.

The BRIDG project is divided into two areas. The BRIDG Advisory Board sets the harmonization priorities, coordinates the development efforts of its constituencies and determines the strategic direction of the project. The Technical Harmonization Committee provides the management, support and interrelation of the BRIDG model.

The BRIDG model is an instance of a Domain Analysis Model (DAM). As such, it depicts a shared representation of what is called the “dynamic and static semantics” of a particular domain-of-interest. The BRIDG model has been adopted by HL7 as the domain analysis model

to be utilized by the Regulated Clinical Research Information Management Technical Committee (RCRIM TC). CDISC has committed to harmonizing its existing standards with the BRIDG model and as noted above has set schedules for doing so. The National Cancer Institute is using the BRIDG model to support application development within the caBIG program as part of the clinical trial management workspace. The FDA, through the RCRIM technical committee, is developing four HL7 messages based on the BRIDG model to support electronic submission of Study Design, Study Participation, Subject Data and Adverse Event reporting.

The BRIDG model, unfortunately, suffers from a number of inconsistencies [108].

4.4.7. The Nationwide Health Information Network (NHIN)

The mission of the NHIN Nationwide Health Information Network is to provide a secure interoperable health information infrastructure that will connect providers, consumers, and others involved in supporting health and healthcare across the U.S. [109]. Its goal is (1) to enable health information to follow the consumer in such way that it will be available for clinical decision making and (2) to support appropriate use of healthcare information beyond direct patient care so as to improve the nation's health.

In November 2005, the Office of the National Coordinator for Health IT (ONCHIT) awarded four contracts totaling \$18.6 million to Accenture, Computer Sciences Corporation (CSC), IBM and Northrop Grumman to develop prototype architectures for the NHIN and to interconnect three communities as a demonstration of utility [110].

A common characteristic among the architectures developed are that they provide technology neutral interfaces between the systems of their stakeholder organizations. The stakeholders include care delivery organizations using EHRs, consumer organizations that operate personal health records (PHRs), health information exchanges (HIEs) that enable the movement of health related data within participant groups, and organizations that make secondary use of data such as that required for public health, research and quality assessment. An overriding architectural principle of the NHIN is to create a "network of networks" that will provide the interconnections between existing stakeholder networks in such a way that they can support additional information exchange beyond their own bounds.

The architectures developed by the contracted groups will inform the selection of standards to be developed while also making use of standards that are in place. The process of choosing among standards will then be performed by HITSP.

We mention here specifically the architecture developed by CSC as it utilized the components of the National Multi-Protocol Ensemble for Self-Scaling Systems for Health (NMESH) project, a promising effort to connect and provide access to patient data from EHRs, (PHRs), and research data.

4.4.8. Cancer Biomedical Informatics Grid (caBIG)

The caBIG Cancer Biomedical Informatics Grid is sponsored by the National Cancer Institute (NCI) and is administered by the National Cancer Institute Center for Bioinformatics (NCICB). "The mission of caBIG is to provide infrastructure for creating, communicating and sharing bioinformatics tools, data and research results, using shared data standards and shared data models." This mission is intended to support translational and personalized medicine within the

domain of cancer research and cancer care. Of interest here is the cancer Common Ontologic Representation Environment (caCORE), a caBIG infrastructure component that provides a mechanism designed to create interoperable biomedical information systems. caCORE [111-112] is composed of three major components: the Enterprise Vocabulary Services (EVS), the cancer Data Standards Repository (caDSR), and the cancer Bioinformatics Infrastructure Objects (caBIO).

The EVS is the controlled vocabulary server of caCORE and as such it attempts to address the semantic dimension of interoperability by providing external applications with runtime access to nomenclatures, thesauri, and ontologies such as:

- NCI Thesaurus
- Gene Ontology
- National Drug File Reference Terminology
- LOINC
- Microarray Gene Expression Data (MGED) Ontology
- MedDRA
- SNOMED.

The syntactic component of interoperability is addressed by the caDSR, a metadata repository and registry whose role to provide the link between data elements and the terms from the standardized vocabularies in the EVS. caCORE data elements are structured as defined in the ISO/IEC 11179 model, which means that they consist of two parts: a *Data Element Concept* – the conceptual definition of the data element – and a *Value Domain* – a specification of accepted values for the data element which can be provided by either a list of permitted values or by a definition including the data type (string, integer, date, etc.) and unit of measure. Data elements are unique pairings of these two parts.

The caBIG project and the caCORE infrastructure is a promising technology in the advancement of interoperability in HIT. However, notably missing from caBIG's attempt at enabling interoperability is the use of sound ontological principles (see section 8.2.3.1) in the creation of data elements. What is built does indeed conform to the ISO/IEC 11179 specifications, but these specifications alone are not sufficient to create data elements with precise and clear meanings. This is, for instance, exemplified by the poor design of and the many mistakes still present in the NCI Thesaurus [16].

5 REFERENCES

1. American National Standards Institute, *ANSI Essential Requirements: Due process requirements for American National Standards*. 2008: New York.
2. ANSI Healthcare Information Standards Board. *ANSI-HISB page*. Available from: <http://ansi.org/hisb/>.
3. Healthcare Information Technology Standards Panel. *Welcome to www.HITSP.org*. 2008; Available from: <http://www.hitsp.org/>.
4. HITSP Technical Committees, *HITSP Interoperability Specification Overview*. 2007.

5. Blankenhorn, D. (2009) *Who will control the coming health IT standards?* ZDNet Healthcare.
6. Coronado, S.d., et al., *The NCI Thesaurus quality assurance life cycle*. Journal of Biomedical Informatics, 2009. **42**(3): p. 530-539.
7. Ceusters, W., B. Smith, and L. Goldberg, *A terminological and ontological analysis of the NCI Thesaurus*. Methods of Information in Medicine, 2005. **44**: p. 498-507.
8. Committee on Engaging the Computer Science Research Community in Health Care Informatics; National Research Council, *Computational Technology for Effective Health Care: Immediate Steps and Strategic Directions*, ed. W.W. Stead and H.S. Lin. 2009.
9. Schorlemmer, M. and Y. Kalfoglou, *Institutionalising Ontology-Based Semantic Integration*. Journal of Applied Ontology, 2008. **3**(3): p. 131-150.
10. Brinkley, J.F., et al., *A framework for using reference ontologies as a foundation for the semantic web*, in *Proceedings of the AMIA Fall Symposium*. 2006. p. 95-100.
11. Bodenreider, O., *Biomedical Ontologies in Action: Role in Knowledge Management, Data Integration and Decision Support*, in *Yearbook of Medical Informatics: access to health information*, International Medical Informatics Association, Editor. 2008, J. Schattauer: Stuttgart.
12. Smith, B., et al., *The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration*. Nature Biotechnology, 2007. **25**: p. 1251-1255.
13. McDonald, C., Huff, S., Suico, J., Hill, G., Leavelle, D., Aller, R., Forrey, A., Mercer, K., DeMoor, G., Hook, J., Williams, W., Case, J., Maloney, P., *LOINC, a Universal Standard for Identifying Laboratory Observations: A 5-Year Update*. Clinical Chemistry, 2003. **49**(4): p. 624-644.
14. National Cancer Institute. *NCI Homepage*. Available from: <http://www.cancer.gov/>.
15. de Coronado, S., Haber, M., Sioutos, N., Tuttle, M., Wright, L. *NCI Thesaurus: Using Science-Based Terminology to Integrate Cancer Research Results*. in *Medinfo*. 2004: IOS Press.
16. Ceusters, W., Smith, B. , *A Terminological and Ontological Analysis of the NCI Thesaurus*. . Methods of Information in Medicine 2005. **44**: p. 498-507.
17. National Library of Medicine. *NLM Homepage*. Available from: <http://www.nlm.nih.gov/>.
18. Humphreys, B., Lindberg, D., Schoolman, H., Barnett, G., *The Unified Medical Language System: An Informatics Research Collaboration*. J Am Med Inform Assoc, 1998. **5**(1): p. 1-11.
19. Bodenreider, O., *The Unified Medical Language System (UMLS): integrating biomedical terminology*. Nucleic Acids Research, 2004. **32**: p. 267-270.
20. Parmenides of Elea. *On Nature*. ca 475 BC; Available from: <http://www.elea.org/Parmenides/>.
21. Hay, D. and K.A. Healy, *Defining Business Rules ~ What Are They Really?* 2000, The Business Rule Group.
22. Spackman, K., Campbell, K., Cote R. , *SNOMED RT: a reference terminology for health care*. . Proc AMIA Annu Fall Symp. , 1997: p. 640-4.
23. ACORD, *Data Dictionary for Global Insurance Industry*. 2005.
24. Haley Systems Inc. *Conceptual Role Model (CRM) Definitions - Structured*. 2007; Available from: <http://www.haley.com/0141049681215489/products/ACORD-CRM-definitions.html>.

25. Noy, N.F. and D.L. McGuinness, *Ontology Development 101: A Guide to Creating Your First Ontology*. 2001, Stanford Knowledge Systems Laboratory.
26. Humphreys, B., McCray, A., Cheh, M. , *Evaluating the coverage of controlled health data terminologies: report on the results of the NLM/AHCPR large scale vocabulary test*. . J Am Med Inform Assoc. , 1997 **4**(6): p. 484-500.
27. Strang, N., Cucherat, M., Boissel, J. , *Which coding system for therapeutic information in evidence-based medicine?* Comput Methods Programs Biomed, 2002. **68**(1): p. 73-85
28. Wasserman, H., Wang, J. , *An applied evaluation of SNOMED CT as a clinical vocabulary for the computerized diagnosis and problem list*. . AMIA Annu Symp Proc. , 2003: p. 699-703.
29. Brown, S., Bauer, B., Wahner-Roedler, D., Elkin, P. , *Coverage of oncology drug indication concepts and compositional semantics by SNOMED-CT*. . AMIA Annu Symp Proc. , 2003: p. 115-9.
30. Penz, J., Brown, S., Carter, J., Elkin, P., Nguyen, V., Sims, S., Lincoln, M. , *Evaluation of SNOMED coverage of Veterans Health Administration terms*. Medinfo. . Medinfo, 2004. **11**((Pt 1)): p. 540-4.
31. Chiang, M., Casper, J., Cimino, J., Starren, J., *Representation of ophthalmology concepts by electronic systems Adequacy of controlled medical terminologies*. . Ophthalmology, 2005. **112**(2): p. 175-183.
32. Richesson, R., Andrews, J., Krischer, J., *Use of SNOMED CT to Represent Clinical Research Data: A Semantic Characterization of Data Items on Case Report Forms in Vasculitis Research*. J Am Med Inform Assoc, 2006. **13**: p. 536-546.
33. J.E. Andrews, R.L. Richesson, and J. Krischer, *Variation of SNOMED CT coding of clinical research concepts among coding experts*. Journal of the American Medical Informatics Association, 2007. **14**(4): p. 497-506.
34. Ceusters, W., Steurs, F., Zanstra, P., Van Der Haring, E., Rogers, J., *From a Time Standard for Medical Informatics to a Controlled Language for Health*. International Journal of Medical Informatics, 1998. **48**(1-3): p. 85-101.
35. Bodenreider, O., Burgun, A., Rindfleisch, TC. , *Assessing the consistency of a biomedical terminology through lexical knowledge*. . Int J Med Inf. , 2002 **67**(1-3): p. 85-95.
36. Elkin, P., Harris, M., Ogren, P., Buntrock, I., Brown, S., Solbrig, H., Chute, C. . *Semantic augmentation of Description Logic based terminologies. Addendum to Proceedings of IMIA-WG6*. in *Medical Concept and Language Representation*. 1999. Phoenix.
37. Ceusters, W., Smith, B., Kumar A., Dhaen C. *Ontology-Based Error Detection in SNOMED-CT®*. in *Proceedings of Medinfo*. 2004.
38. Ceusters, W., Smith, B., *Ontology and Medical Terminology: Why Description Logics are not enough*. Proceedings of the Conference Towards an Electronic Patient Record (TEPR 2003), 2003.
39. Schulz, S. and R. Cornet, *SNOMED CT's Ontological Commitment*, in *ICBO: International Conference on Biomedical Ontology*, B. Smith, Editor. 2009, National Center for Ontological Research: Buffalo NY. p. 55-58.
40. Guarino, N. *Formal Ontologies and Information Systems*. in *FOIS98*. 1998: IOS Press.
41. Baader, F., McGuinness, D., Nardi, D. , ed. *The Description Logic Handbook*. . 2003, Cambridge University Press.
42. Farrugia, J. *Model-Theoretic Semantics for the Web*. in *12th International Conference on the WWW 2003*: ACM.

43. Ceusters, W., *Towards A Realism-Based Metric for Quality Assurance in Ontology Matching*, in *Formal Ontology in Information Systems*, B. Bennett and C. Fellbaum, Editors. 2006, IOS Press: Amsterdam. p. 321-332.
44. Kumar, A. and B. Smith, *The Unified Medical Language System and the Gene Ontology*, in *KI2003: Advances in Artificial Intelligence (Lecture Notes in Artificial Intelligence 2821)*. 2003. p. 135-148.
45. Ceusters, W., Smith, B., Kumar A., Dhaen C., *Mistakes in Medical Ontologies: Where Do They Come From and How Can They Be Detected?*, in *Ontologies in Medicine. Proceedings of the Workshop on Medical Ontologies, Rome October 2003*, D.M. Pisanelli, Editor. 2004, IOS Press.
46. Huff, S.M., et al., *An event model of medical information representation*. Journal of the American Medical Informatics Association, 1995. **2**: p. 116-134.
47. Coyle, J.F., A. Rossi-Mori, and S.M. Huff, *Standards for detailed clinical models as the basis for medical data exchange and decision support*. International Journal of Medical Informatics, 2003 **69**(2-3): p. 157-74.
48. Rector, A.L., et al., *A framework for modelling the electronic medical record*. Methods of Information in Medicine, 1993 **32**(2): p. 109-119.
49. Smith, B., Ceusters, W., Klagges, B., Kohler, J., Kumar, A., Lomax, J., Mungall, C.J., Neuhaus, F., Rector, A., Rosse, C., *Relations in Biomedical Ontologies*. Genome Biology, 2005. **6**(5): p. R46.
50. IFOMIS. *Basic Formal Ontology*. 2009; Available from: <http://www.ifomis.uni-saarland.de/bfo/>.
51. Smith, B., et al., *Relations in biomedical ontologies*. Genome Biology, 2005. **6**(5): p. R46.
52. Smith, B., *Ontology (Science)*, in *Formal Ontology in Information Systems - Proceedings of the Fifth International Conference (FOIS 2008)*, C. Eschenbach and M. Grüninger, Editors. 2008, IOS Press: Amsterdam. p. 21-35.
53. e-Health Standardization Focus Group. in *Current and future standardization issues in the e-Health domain: Achieving interoperability*. . 2004. Bruxelles: CEN/ISSS.
54. Smith, B., Ceusters, W., *HL7 RIM: An Incoherent Standard*. Studies in Health Technology and Informatics, 2006. **124**: p. 133-138.
55. Pacitti, D., *The Nature of the Negative: Towards an Understanding of Negation and Negativity*. 1991, Pisa: Giardini.
56. Simon, J., J.M. Fielding, and B. Smith, *Using Philosophy to Improve the Coherence and Interoperability of Applications Ontologies*, in *First Workshop on Philosophy and Informatics; DFKI, Cologne, Germany.*, B. Büchel, B. Klein, and T. Roth-Berghofer, Editors. 2004. p. 65 - 72.
57. Poli, R., *Descriptive, Formal and Formalized Ontologies*, in *Husserl's Logical Investigations Reconsidered*, D. Fisette, Editor. 2003, Kluwer: Dordrecht. p. 193 - 210.
58. International Electrotechnical Commission. *International Electrotechnical Commission*. 2008; Available from: <http://www.iec.ch/>.
59. Health Level 7. *Health Level Seven Version 3.0*. 2007; Available from: <http://www.hl7.org/>.
60. American College of Radiology. *ACR Homepage*. Available from: <http://www.acr.org>.
61. National Electrical Manufacturers Association. *NEMA Homepage*. Available from: <http://www.nema.org>.

62. Smith, B. and D.M. Mark, *Do mountains exist? Towards an ontology of landforms*. Environment and Planning B: Planning and Design, 2003. **30**(3): p. 411-427.
63. Goldberg, D., *Plato versus Aristotle: categorical and dimensional models for common mental disorders*. Compr Psychiatry, 2000 **41**(2 Suppl 1): p. 8-13.
64. Institute of Electrical and Electronics Engineers. *Institute of Electrical and Electronics Engineers*. 2008; Available from: <http://www.ieee.org/portal/site>.
65. American Society for Testing and Materials International. *ASTM Homepage*. 2007; Available from: <http://www.astm.org>.
66. The Object Management Group. *OMG Homepage*. Available from: <http://www.omg.org/>.
67. Office of the Federal Register, *Executive Order 13335—Incentives for the Use of Health Information Technology and Establishing the Position of the National Health Information Technology Coordinator*. Federal Register, 2004. **69**(84): p. 24059-24061.
68. Office of the National Coordinator. *Office of the National Coordinator: mission*. 2008; Available from: <http://www.hhs.gov/healthit/onc/mission/>.
69. European Committee for Standardization, *EN 12388:2005. Health informatics - Time standards for healthcare specific problems*. 2005.
70. *Veterinary Terminology Services*. 2006; Available from: <http://terminology.vetmed.vt.edu/default.htm>.
71. Hume, D., *A treatise of human nature*, ed. D.F. Norton and M.J. Norton. 2000, Oxford/New York: Oxford University Press.
72. Kant, I., *The critique of pure reason*. 2003, Project Gutenberg.
73. Office of the National Coordinator for Health Information Technology (ONC). *Harmonized Use Case for Electronic Health Records (Laboratory Result Reporting)*, . Available from: http://www.ansi.org/standards_activities/standards_boards_panels/hisb/hitsp.aspx?menuid=3.
74. Healthcare Information and Management Systems Society. *Healthcare Information and Management Systems Society* 2008; Available from: <http://www.himss.org/ASP/index.asp>.
75. National Center for Research Resources. *National Center for Research Resources*. 2008; Available from: <http://www.ncrr.nih.gov/>.
76. Banzato, C.E.M., *Classification in Psychiatry: The Move Towards ICD-11 and DSM-V*. Curr Opin Psychiatry, 2004. **17**(6): p. 497-501.
77. Skodol, A.E., et al., *Dimensional Representations of DSM-IV Personality Disorders: Relationships to Functional Impairment*. Am J Psychiatry, 2005. **162**: p. 1919-1925.
78. Agency for Healthcare Research and Quality. *AHRQ Homepage*. Available from: <http://www.ahrq.gov/>.
79. Centers for Medicare & Medicaid Services. *CMS Homepage*. Available from: <http://www.cms.hhs.gov/>.
80. Shedler, J. and D. Westen, *Refining DSM-IV personality disorder diagnosis: integrating science and practice*. Am. J. Psychiatry, 2004. **161**: p. 1350-1365.
81. National Center for Health Statistics. *NCHS Homepage*. Available from: <http://www.cdc.gov/nchs/>.
82. Meehl, P.E., *Clarifications about taxometric method*. Applied & Preventive Psychology, 1999. **8**: p. 165-174.

83. Magidson, J. and J.K. Vermunt, *Latent Class Models*, in *The Sage Handbook of Quantitative Methodology for the Social Sciences*, D. Kaplan, Editor. 2004, Sage Publications: Thousand Oaks. p. 175-198.
84. Krueger, R.F., et al., *Externalizing Psychopathology in Adulthood: A Dimensional-Spectrum Conceptualization and Its Implications for DSM–V*. *Journal of Abnormal Psychology*, 2005. **114**(4): p. 537–550.
85. Shumway, M. and T. Sentell, *Mental Health Measurement for Research and Practice: A Survey of Publications in Leading Journals*. *Abstr Acad Health Serv Res Health Policy Meet*, 2002. **19**: p. 16.
86. Rosse, C., et al., *A strategy for improving and integrating biomedical ontologies*, in *Biomedical and Health Informatics: From Foundations to Applications to Policy; Proceedings of the 2005 AMIA Annual Symposium*, C.P. Friedman, J. Ash, and P. Tarczy-Hornoch, Editors. 2005, American Medical Informatics Association: Washington DC. p. 639-43.
87. National Uniform Billing Committee. *NUBC Homepage*. Available from: <http://www.nubc.org>.
88. National Uniform Claim Committee. *NUCC Homepage*. Available from: <http://www.nucc.org>.
89. Ceusters, W., P. Elkin, and B. Smith, *Negative Findings in Electronic Health Records and Biomedical Ontologies: A Realist Approach*. *International Journal of Medical Informatics*, 2007. **76**: p. 326-333.
90. Jaffee, S. and T. Price, *Gene–environment correlations: a review of the evidence and implications for prevention of mental illness*. *Molecular Psychiatry*, 2007. **12**: p. 432–442.
91. Smith, B., et al., *On carcinomas and other pathological entities*. *Comparative and Functional Genomics*, 2005. **6**(7-8): p. 379 - 387.
92. Sadler, J., *Epistemic Value Commitments in the Debate over Categorical vs. Dimensional Personality Diagnosis*. *Philosophy, Psychiatry, & Psychology*, 1996. **3**(3): p. 203-222.
93. Zachar, P., *Psychiatric disorders are not natural kinds*. *Philosophy, Psychiatry and Psychology*, 2000. **7**: p. 167-194.
94. American Health Information Management Association. *American Health Information Management Association 2008*; Available from: <http://www.ahima.org/>.
95. National Alliance for Health Information Technology. *NAHIT Homepage*. Available from: <http://www.nahit.org>.
96. American Academy of Family Physicians. *AAFP Homepage*. :[Available from: <http://aafp.org>.
97. American Academy of Pediatrics. *AAP Homepage*. Available from: <http://www.aap.org>.
98. American College of Physicians. *ACP Homepage*. Available from: <http://www.acponline.org>.
99. California Healthcare Foundation. *CHCF Homepage*. Available from: <http://www.chcf.org>.
100. Haslam, N., *Kinds of Kinds: A Conceptual Taxonomy of Psychiatric Categories*. *Philosophy, Psychiatry, & Psychology*, 2002. **9**: p. 203-218.
101. Integrating the Healthcare Enterprise. *Welcome to Integrating the Healthcare Enterprise 2008*; Available from: <http://www.ihe.net/>.

102. Shapiro, S.C. and The SNePS Implementation Group, *SNePS 2.6.1 User's Manual*. 2004, Department of Computer Science and Engineering, University at Buffalo, The State University of New York: Buffalo, NY.
103. Clinical Data Interchange Standards Consortium. *CDISC Homepage*. Available from: <http://www.cdisc.org/index.html>.
104. Rudnicki, R., et al., *What Particulars are Referred to in EHR Data? A Case Study in Integrating Referent Tracking into an Electronic Health Record Application*, in *American Medical Informatics Association 2007 Annual Symposium Proceedings, Biomedical and Health Informatics: From Foundations to Applications to Policy*, Teich JM, Suermondt J, and H. C, Editors. 2007: Chicago, IL. p. 630-634.
105. Fridsma, D., Evans, J., Hastak, S., Mead, C., *The BRIDG Project: A Technical Report*. J Am Med Inform Assoc, 2008. **15**: p. 130-137.
106. Health Level 7. *RCRIM Homepage*. Available from: <http://www.hl7.org/Special/committees/rcrim/index.cfm>.
107. United States Food and Drug Administration. *FDA Homepage*. Available from: <http://www.fda.gov/>.
108. Baud, R., et al., *Reconciliation of Ontology and Terminology to cope with Linguistics*, in *Proceedings of MEDINFO 2007, Brisbane, Australia, August 2007*, K. Kuhn, J. Warren, and T. Leong, Editors. 2007, Ios Press: Amsterdam. p. 796-801.
109. U.S. Department of Health & Human Services. *Nationwide Health Information Network (NHIN): Background*. 2008; Available from: <http://www.hhs.gov/healthit/healthnetwork/background/>.
110. Rishel, W., Riehl, V., Blanton, C. (2007) *Summary of the NHIN Prototype Architecture Contracts*. Gartner Report.
111. Saltz, J., Oster, S., Hastings, S., Langella, S., Kurc, T., Sanchez, W., Kher, T., Manisundaram, A., Shanbhag, K., Covitz, P., *caGrid: design and implementation of the core architecture of the cancer biomedical informatics grid*. *Bioinformatics*, 2006. **22**(15): p. 1910-1916.
112. Komatsoulis, G., Warzel, D., Hartel, F., Shanbag, K., Chiukuri, R., Fragoso, G., de Coronado, S., Reeves, D., Hadfield, J., Ludet, C., Covitz, P., *caCORE version 3: Implementation of a model driven, service-oriented architecture for semantic interoperability*. *Journal Biomedical Informatics*, 2008. **41**: p. 106-123.