Emerging medical information standards as applicable to clinical research data

A study performed in the context of the project

'Exploring eyeGENE, an International Genotype / Phenotype Database, from a Bioinformatics Perspective'

July 16th, 2008

Ron Rudnicki, Werner Ceusters
New York State Center of Excellence in Bioinformatics and Life Sciences
701 Ellicott street
Buffalo NY, 14203

Table of contents

1 INTRODUCTION

The eyeGene project [1] is designed to open new and exciting frontiers for eye research. The immediate goal of the project is to facilitate the creation of integrated processes and shared information resources to further research on genotype-phenotype associations. An additional goal of the project is to integrate systems for medical research. This is in accordance with the goals stated in the NIH roadmap:

"*Clinical research needs to develop new partnerships among organized patient communities, community-based physicians, and academic researchers. . . . NIH will promote creation of better integrated networks of academic centers that work jointly on clinical trials and include community-based physicians who care for large groups of well-characterized patients. Implementing this vision will require new ways to organize how clinical research information is recorded, new standards for clinical research protocols, modern information technology . . .*" [2]

The data collected in eyeGene includes:

- Patient profile data – e.g. data representing birth date, contact information, race, gender
- Family history data – data on relatives sharing the patient's primary diagnosis
- Phenotypic data – a set of dynamic (i.e. user defined) data elements used to record clinical findings related to one or more of 21 diagnoses
- Genetic lab test results – data about lab procedure, gene, exons screened, DNA changes, protein changes, and genotype
- Files – e.g. images and consent forms

The issue confronting the eyeGene project at this point of its evolution and the one to be considered here is which of the many standards currently in use or in development are best equipped to further the goals of the project by allowing for the semantic interoperability needed between the specialized information of medical research (e.g. eyeGene phenotype data) and data from clinical trials, data repositories containing genomic and proteomic data, and electronic health records at a national and international level.

This document gives an overview of the state in the art in standardization focusing on the US, but without being exclusive.

2 EXECUTIVE SUMMARY

The current view on standardization in healthcare information technology is that it should support ***semantic interoperability***: 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, as defined by the end users of both systems. Although semantic interoperability in the health domain is already a widespread reality, it is so only at different degrees and at various scales around the world. Many standards are available, several successfully applied in limited settings, but even these do not meet all of the increasing needs of clinicians and researchers. The goals of translational medicine require standards that are capable of creating interoperability across domains of medicine and across national boundaries.

In addition, it is not only important to develop national and international standards and the related (local) implementation guides, but also to prepare the context to establish priorities and requirements on the domains to be standardized, to adopt a coherent set of standards by the authorities, to promote their adoption, and to evaluate their impact. This is very well understood by the European Commission which funded in 2006 the RIDE project to analyse past and current achievements and gaps in Health Information Technology standardization with the goal to make recommendations for the various Member States [3]. Having delivered one of the PIs for the RIDE project, we used the experience and knowledge gained to prepare this current report, thereby (1) focusing more in detail on the US market, and (2) reporting on new developments.

In section 3 we first describe the current challenges about the standards for semantic interoperability as they pertain to biomedical data. Next follows a discussion of the value of standards and different societal agents capable of creating and maintaining them. Lastly in this section, a brief road map of standardization activities for the future is presented.

In section 4 we survey the most relevant organizations in charge of the development of standards, both in the formal networks of ANSI and ISO with national and international standardization bodies, and independent initiatives (e.g. HL7, ASTM, DICOM). Furthermore, we describe a set of relevant initiatives to support and facilitate the deployment of standards, either by governments or spontaneously convened.

In the remaining sections we focus on the standardization activities about four specific areas:

- The structured content of patient summary (section 5),
- The structured content of clinical research data (section 6)
- Terminologies and coding schemes, with an extended analysis of SNOMED-CT (section 7), and
- Ontology language and tools (section 8).
- Gaps and overlaps in standards (section 9)

Finally, we provide a comprehensive list of 214 acronyms (section 10), that cannot be exhaustive, due to the nature of this field, but intends to cover most of the generic needs. In addition to an extensive bibliography *about* standards and the related activities we provide an explicit list of the most relevant web sites (section 11).

3 THE CHALLENGES ON THE STANDARDS FOR SEMANTIC INTEROPERABILITY

Let us consider an ideal scenario, where an ophthalmologist is asked expert advice on some diabetic patient. He receives from the GP in charge of that patient all relevant information about history and active problems and from the endocrinologist the pertinent information about current state of complications. Issues of confidentiality and security being solved, clinical information is directly exchanged between the computers of the above professionals through the internet.

The electronic record system of the ophthalmologist rearranges each unit of information under the proper section and subsection, e.g. reason for referral, history of major diseases, previous operations, current medications, state of diabetes, active and inactive eye problems. It also prepares links to potentially relevant guidelines, to a drug database, and to information for patients (see e.g. the infobuttons in [4]). Given the current state of standards, is this a realistic scenario and if not, what level of communication and interoperability is possible in today's health information environment?

3.1 Need for a complete, harmonized, generalized set of standards

Healthcare is an information intensive industry; equal to finance and banking in terms of its dependence on information. Yet current use of common standards in the health sector is primitive compared with other industries. Nonetheless, it is clear from the current economic climate (global and national), the increasing number of competing demands and a growing importance of information and communication technologies in healthcare, that the health sector will need to follow other industries in the way to the Information Society.

Most healthcare information technology solutions so far are limited to a particular Predefined Operational Domain (POD), i.e. to a set of interoperable applications used by a particular sub-community of users, which may be defined by constraints on the geographical area (a jurisdiction), on the workflows, on the health problems, or on the healthcare settings involved. These solutions apply standards for healthcare informatics, plus some additional implicit or explicit agreements, needed to reach the appropriate level of semantic interoperability within a POD. However, these agreements are hardly scalable, i.e. it is difficult to merge the solutions from different PODs.

An additional problem is that for standards to be effective, they should be widely adopted and applied, and complemented by suitable local implementation guides. In addition, standards developed by diverse organizations must exhibit a common vision. Only then will standards permit society to make more effective use of resources and allow more effective communication among all parties to particular activities, transactions, or processes (and health information is a good example of this).

Timely and appropriate standards are specifically critical to the long-term viability of electronic health records, as they in principle allow products and services from different vendors to work seamlessly, facilitate competition between solution providers, and reduce uncertainty in the marketplace. Indeed, producers and users of health information systems can participate more effectively when those systems work together, and the standardization process can contribute mightily to achieving this. After all, consumers feel more confident when systems operate

seamlessly, efficiently, securely and effectively and work in a way that respects them as individuals. Adopting common standards is a pre-requisite for achieving this.

3.2 Standards-making processes

Standards can be thought of as ***agreements*** on how to implement technologies allowing, for example, buyers to choose compatible medical equipment and software from a variety of vendors (thus encouraging both innovation and price competition). An important criterion for a successful standard is the impact it achieves in its target environment (i.e. the extent to which any standard is adhered to in practice).

Many standards we take for granted today are in fact products and services that have come to be broadly used and implemented on a national or even global basis. For example, these kinds of ***de facto standards*** are driving the growth and use of applications of the Internet, and are moving faster than both traditional and non-traditional standards-setting organizations can keep pace with. Sometimes such proprietary standards emerge when a single vendor controls a large share of the market for a particular item (e.g. the Windows operating system for personal computers).

Mandated standards are those standards usually prescribed by Government, Federal or State legislation. However, most are not legal documents in themselves. Consensus standards, for example, are developed by committees with representatives from those with a stake in the outcome, who value and have arrived at a general agreement for a consistent approach to a particular process. The committees can include representatives of vendors, the medical community, government and other interested parties who choose to participate in the laborious processes that writing and agreeing on standards can involve.

Formal national standards bodies make the network of international organizations such as the International Standards Organisation (ISO) [5] and the European Standardization Committee (CEN) [6].

Other independent Standard Developing Organizations (SDOs) are successful in their own subdomain. Organizations extremely active in the Health Informatics field are DICOM [7], Health Level Seven (HL7) [8], the American Society for Testing and Materials (ASTM) [9], and the IEEE [10].

Those organizations will be described later in this document.

All standards committees can have problems reaching decisions as rapidly as new technologies are developed. Purchasers of medical equipment and software can more easily build extensible systems by buying items that store and exchange information according to one or more of these consensus standards, rather than proprietary standards. Such standards are sometimes described as open standards.

3.3 The scope of the standards on data content

In the realm of setting standards, high priority actions are promoting the availability and use of robust data content standards, including:

- data elements that are collected to describe different types of entities or to document different types of events, e.g., gender, date of birth, presenting complaint, blood pressure reading, provider identifier

- what constitutes a description of a particular entity or event (e.g., a birth certificate, a report of an injury, a definition of a reportable case, an autopsy report, an entry in a cancer registry) or, in other words, the set of data elements that constitute such a description
- message formats and the data elements to be included in various types of messages that are transmitted between locations or organizations (e.g., a laboratory test order or result, notifiable disease report)
- allowable data values for specific data elements – these can be entire classifications, code sets, or controlled vocabularies (e.g., ICD-10 for cause of death [11], ICD-O for cancer registry entries [12], LOINC (Logical Observations: Identifiers, Names, Codes) for laboratory observation names) [13], large sets of unique identifiers for particular entities or individuals, (e.g., provider identifiers), or very restricted sets of values (e.g., race/ethnicity codes)
- mappings between different value sets, e.g., between MedDRA (Medical Dictionary for Regulatory Activities) [14] and SNOMED (Systematized Nomenclature of Medicine) [15] or between SNOMED and ICD-9-CM [16];
- information models that define the context in which the standards will be used – the entities of interest and the relationships among them - thus in essence defining the universe of standards that are needed for a defined domain;
- survey questions and any coded responses to such questions, e.g., for functional status, behavioural risk factors, health care, research, and public health;
- guideline, protocol, and algorithm formats (e.g., Arden Syntax [17], Guidelines Interchange Format (GLIF) [18], Guideline Element Model (GEM)) [19] for detecting and preventing errors and reminding health professionals and patients of appropriate actions.

Standards for these broad categories of data content are applicable across the spectrum of health administration, clinical care, public health (including bioterrorism detection, emergency preparedness, and disaster response), and clinical and health services research. Individual standards may apply to all of these arenas or to a subset of them.

3.4 Making progress in the field of health information

Increasing interest in electronic health records has underscored the important role that standards of semantic interoperability play in the whole electronic health record endeavour to ensure compatibility and transferability of patient records from one setting to another. Making progress with electronic health records involves settling on a number of standards and, in particular, getting agreement to implement common standards or building blocks in the following areas:

- identification;
- privacy and data protection;
- security and authentication;
- messaging and communication; and

- coding and classification.

The adoption of standards by authorities provides a mechanism for actually getting agreed standards in the field of health information implemented nation-wide. The focus should be on the integration of information where needed to achieve the best possible health outcomes. Four primary drivers may be identified, all of which require interface/interaction among heterogeneous systems and applications.

1. **Continuity of health records over a patient's lifetime**. The goal is to provide continuity of care and its documentation in health records for a patient throughout his/her lifetime, independent of the institutions or practices that are involved for particular episodes of health care. This includes the ability for patients/consumers to document personal health interventions and disease prevention actions. An important aspect of attaining continuity is the provision of mechanisms not only for secure data exchange and identification of persons, places and organizations, but also of each individual entity in reality about which data is collected [20], and all this in a way that allows integration and visualization of a patient's lifetime data at the point of care.
2. **Collaborative service provision**. The goal is the coordination of health-related tasks among independent organizations or individuals that take part in particular episodes of health care. Examples include processing of patient orders, specialty referral and consultation, communication and processing of prescriptions and post-operative home care follow up.
3. **Integration of knowledge for decision support**. The goal is to facilitate the context-appropriate, patient-specific application of high quality knowledge resources for decision support in the two activities just mentioned. Various kinds of support may be aimed at doctors, nurses, other care providers and patients themselves. Purposes are to reduce errors, promote best practices and foster cost-effectiveness. Here not only ontologies play an important role, but for sure standards about best practices in ontology development [21].
4. **Aggregation of health information over a population**. The goal is to foster aggregation of health information across patients, outcomes research, health services research, predictive modelling and public health surveillance and monitoring.

Hospitals, health systems, and medical groups are all making big investments in clinical applications that eventually build the electronic health record. A jurisdiction simply won't get the value from these investments unless it can combine, compare, and share the information across applications, across departments, and across sites and settings of care. Standards make this possible.

Standards – including standard vocabularies - are the building blocks of effective National and Regional Health Information Infrastructures because they promote interoperability of systems and the creation of comparable data at different sites. Standards are essential for efficient and effective transmission of data between individuals and their healthcare providers and among the myriad of organizations that must exchange data to perform essential public health and health care functions. Use of standards is a pre-requisite for the efficient integration, aggregation, and re-use of data collected at multiple sites, and across time. Standardized health data will enable

safer and better health care, more timely public health and surveillance, and more cost-effective clinical and health services research.

The organizations responsible for the maintenance and quality of specific health data standards should coordinate their activities to avoid duplicate or conflicting standards and provide a single point of entry for requests for changes and additions. We turn now to a discussion of those organizations and the status of their efforts in this regard.

4 THE WORLD OF STANDARDIZATION AND STANDARDIZATION POLICIES

The standards related to semantic interoperability should be considered in the wider context of standardization in general. In this section we describe the environment surrounding the production and usage of standards, with a particular focus on the US.

4.1 The goals of standardization

The overall objective of standardization is to facilitate the production, handling, or use of products or services in the framework of free trade and free market to the best possible satisfaction of both users and suppliers [22].

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, and to form the basis for further developments.

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

In order to be useful, standards need to:

- be easily available, well publicised and obtainable at the lowest possible cost. On economic grounds it must prove cheaper and quicker to rely on standards than to make new developments from scratch to cover (entirely) the same needs.
- reflect the state of the art at their time of publication, meaning that the area of knowledge they cover must be reasonably stable and correct and that the standards are able to accommodate also these systems that have advanced the furthest in implementing the state of the art.

'Reasonable' stability does not bar further progress from being made. The fact that standards do evolve raises the issue of the backward compatibility of resulting products. It is noteworthy that while standards are meant to introduce a certain degree of order in provision of products, their inevitable successive versions may bring some degree of confusion, which makes critical the need for 'versioning' within standards so as to assure compatibility between versions.

Standards may derive from various processes but in most cases 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 it 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).

- Public authorities on a national or regional level may also trigger the development of standards, and try to stimulate interested parties to find consensus. In some cases, especially related to health and safety of the citizens, public authorities may use standards as part of regulation where technical standards detail how to meet legal requirements e.g. for safety of a product.
- Informal standards may also appear spontaneously, often as the result of a success story, with various interested parties declaring their willingness to share the same characteristics for their

products. This involves a whole range of different situations from one market leader actually owning the specification and deciding on possible changes, to various more or less formal consortia which may adapt a rule set resembling that of formal standards bodies. The long term maintenance of such specifications is sometimes a problem. In the HIT area there are over 250 such informal bodies that publish standards and are more or less open.

Under no circumstances however, standards should be used to keep new and novel paradigms and products from the market. Recent evolutions in the HIT arena are suggestive for such aberrations to take place. This is in particular a risk when large and powerful companies or organisations have 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 itself [23, 24].

4.2 National formal standardization

Whatever initiative is at the origin of standards - from the suppliers, from the users (customers), or from public authorities, all with different agendas in mind - if they are to become part of officially acknowledged regulations, they need to be endorsed by some official body. At this point, they are granted the status of *de jure* standards.

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 [25]. 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 Standard 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 consistently adhere to a set of requirements or procedures known as the "*ANSI Essential Requirements: Due process requirements for American National Standards*" that govern the consensus development process [26].

The process to create these 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. Thousands of individuals, companies, government agencies and other organizations such as labor, industrial and consumer groups voluntarily contribute their knowledge, talents and efforts to standards 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)

[5], and, via the U.S. National Committee (USNC), the International Electrotechnical Commission (IEC) [27].

4.3 Global formal standardization

Basically, global standardization relies on the International Standards Organisation (ISO), the International Electro-technical Commission (IEC) [27], and the International Telecommunication Union (ITU) [28], all three established in Geneva, Switzerland.

An essential differentiating characteristic of global standards, as compared to national ones, is that they are legally less stringent with regard to national standardization. The agreements between ISO and its member NSBs do not imply that global standards override national ones. The decision as to whether to incorporate a global document into a national corpus of formal voluntary standards is left at the discretion of each NSB.

ISO is the world's largest developer of standards [5]. ISO is a network of the national standards institutes of 156 countries, on the basis of one member per country, with a Central Secretariat in Geneva, Switzerland, that coordinates the system. It is a non-governmental organization: its members are not, as is the case in the United Nations system, delegations of national governments. Nevertheless, ISO occupies a special position between the public and private sectors. This is because, on the one hand, many of its member institutes are part of the governmental structure of their countries, or are mandated by their government. On the other hand, other members have their roots uniquely in the private sector, having been set up by national partnerships of industry associations.

A major sector in ISO is health care technology, and there are about 20 technical committees (TCs) dealing directly or indirectly with health care topics.

ISO cooperates closely with the International Electrotechnical Commission, which is responsible for standardization of electrical equipment. ISO standards are numbered, and have a format that contains "ISO/IEC IS nnnnn:yyyy: Title" where "nnnnn" is the standard number, "yyyy" is the year published, and "Title" describes the subject. IEC will only be included if the standard results from work of JTC1 which deals with Information Technology [29].

4.3.1. ISO/TC215 "Health Informatics"

ISO/TC215 [30] is the Technical Committee of ISO for Health Informatics, and was set up in 1998, after CEN/TC251 [31]. Its work program is now bearing fruit and its first standards have been published. Its scope is defined as: "Standardization in the field of information for health, and Health Information and Communications Technology (ICT) to achieve compatibility and interoperability between independent systems. Also, to ensure compatibility of data for comparative statistical purposes (e.g. classifications), and to reduce duplication of effort and redundancies."

There are 23 *participating* countries with 23 *observer* countries. As of 2007, the total number of ISO standards published under the direct responsibility of ISO/TC215 is 48.

ISO/TC215 liaises with several organisations: CDISC [32], DICOM [7], ICN [33], IMIA [34], UN/ECE [35], W3C [36], etc.

The work of ISO/TC215 is distributed among the following Working Groups:

- TC 215/CAG 1 Executive Council, Harmonization and Operations
- TC 215/WG 1 Data structure
- TC 215/WG 2 Data interchange
- TC 215/WG 3 Semantic content
- TC 215/WG 4 Security
- TC 215/WG 5 Health cards
- TC 215/WG 6 Pharmacy and medicines business
- TC 215/WG 7 Devices
- TC 215/WG 8 Business requirements for Electronic Health Records
- TC 215/WG 9 SDO Harmonization

4.4 Standard Development Organizations

4.4.1. Health Level Seven, HL7

HL7 (Health Level Seven, by reference to the 7th layer of the OSI model) was founded in 1987 by several vendors of software for the healthcare industry, with the assistance of academics and major Health Maintenance Organizations [8]. Their goal was to develop consensual messages formats to facilitate better interoperability inside Hospital Information Systems (HIS).

In 1994, HL7 was accredited by ANSI, as a Standards Developing Organisation (SDO), meaning that HL7 approved specifications are channelled into the official standardization process as formal American National Standards.

Message specifications ('HL7 standard') Version 1.0 was approved in 1987, and was followed by version 2.0 in 1998. Subsequently, version 2 evolved regularly. 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, using the Reference Information Model (RIM) developed to help make messages more consistently implemented than they are for Version 2 [37].

A major focus of current interest in HL7 is the RIM. The large task of forming an object model of basic building blocks for all Health information is now by HL7 considered to be complete enough for productive use. Nevertheless, the RIM, and specifically its documentation, has been found to contain several fundamental flaws [23, 24]. Nevertheless, the RIM has been accepted as an ISO Standard.

Of particular note is an XML-based ***'Clinical Document Architecture'*** [38] set of specifications to exchange structured clinical documents, that was approved in 2000 (ANSI-HL7 CDA, Release 1). The release 2 of the CDA was aligned to the RIM.

Current contributors or 'Benefactors' to HL7 include vendors (Siemens, GE Medical Systems, HBOC-McKesson, IBM, Oracle, Microsoft, Philips), USA or non-USA agencies (USA Veterans Affairs), UK NHS, Centres for Disease Control and Prevention (USA CDC), Standards

Australia, AFNOR (France). Public-private partnerships have also been established with Infoway (Canada), NICTIZ (The Netherlands). Other 'benefactors' include, amongst others, USA healthcare providers or health insurance funds, such as Mayo Fdn, Duke, Kaiser Permanente.

HL7 has many International Affiliates, including: Argentina, Australia, Brazil, Canada, China, Croatia, Czech Republic, Denmark, Finland, France, Germany, Greece, India, Ireland, Italy, Japan, Korea, Lithuania, Mexico, New Zealand, Poland, Spain, South Africa, Switzerland, Taiwan, The Netherlands, the United Kingdom. HL7/USA is said to be under consideration.

4.4.1.1 How the HL7 version 3 development process works

As an organization, HL7 addresses the large domain of health care initially by dividing itself into a number divisions comprised of technical committees and special interest groups. Individuals and companies must subscribe to the HL7 organization if they want to influence the final form of the HL7 and RIM specification, and they would choose to attend committees or SIGs that interest them.

The technical committees and the special interest groups try to ensure that the RIM meets the needs of the standard specifications they are producing. At each working group meeting, the current state of play is presented and new proposals for improvements are discussed.

For a couple of weeks after the main subscribers meeting, many document revisions are circulated by e-mail to the protagonists of each working group. The results of this consultation and argument process are then fed into the RIM harmonisation committee. This is a much smaller group of people than the full HL7 community, and they have a meeting of their own in between the main HL7 meetings. The RIM harmonisation committee takes all the suggestions and requests and issues a new RIM model draft, in UML, just before the next full HL7 meeting. Then the cycle starts again.

4.4.1.2 Versions 2.x

Version 2 was released in 1988. Subsequently, version 2 evolved regularly, with v2.1 approved in 1990, v2.2 in 1994, v2.2 in 1997, v2.3 in 1999, v2.4 in 2000, and v2.5 in 2003.

It still forms the basis for the major HIS implemented in most countries. In 2000, XML encoding of version 2 messages has been approved.

Specifications of Versions 2.x cover:

- Patient Administration - Admission, Discharge, Transfer, and Demographics;
- Order Entry - Orders for Clinical Services and Observations, Pharmacy;
- Dietary, and Supplies;
- Query - Rules applying to queries and to their responses;
- Financial Management - Patient Accounting and Charges;
- Observation Reporting;
- Appointment Scheduling and Resources;
- Primary Care Referral Messages;
- Medical Record Information Management.

- Clinical Context Object Workgroup (CCOW), in support of visual integration. Successive versions of CCOW specifications were also published, with v1.0 in 1999, v1.1 in 1999, v1.2 in 2000, and v1.3 in 2001.

4.4.1.3 Motivation for different versions of HL7

HL7 version 2 was released in 1988. Various revisions and improvements were agreed over the years and by 1994 HL7 v2 was widely adopted. When system vendors say that their system is already HL7 compliant, they may mean version 2.

However, HL7 version 2 failed to achieve any great degree of interoperability. This is widely attributed mainly to the fact that version 2 still allowed too many choices - too much 'optionality'. This allowed all the different system vendors to build a big Tower of Babel. A useful summary of the key limitations of HL7 version 2 is:

- Complex integration process
- Misunderstanding of specifications
- Different and implicit information models
- Difficulty in verifying conformance
- Difficulty in measuring progress
- Widespread optionality
- Lack of explicit support for new technologies (e.g. Object oriented technologies, Extensible Mark-up Language - XML, Web technologies)

One very important technical difference between version 2 and 3 is that all HL7 mini-standards developed in version 2, covering the various interactions between specific application types, in version 3 arise from a single common Reference Information Model (RIM). All application messages and documents comply with a predefined set of requirements known as the Message Development Framework (MDF). The aim is consistent definition of different information objects and their representation in messages. This should allow for easier implementation of the standard as a whole and for clearer conformance requirements.

4.4.1.4 Version 3

Version 3 represents a significant departure from "business as usual" for HL7. Offering lots of optionality and thus flexibility, the V2.x series of messages were widely implemented and very successful. These messages evolved over several years using a "bottom-up" approach that has addressed individual needs through an evolving ad-hoc methodology. There is neither a consistent view of the data that V2.x messages carry nor that data's relationship to other data. HL7's success is also largely attributable to its flexibility. It contains many optional data elements and data segments, making it adaptable to almost any site.

While providing great flexibility, its optionality also makes it impossible to have reliable conformance tests of any vendor's implementation and also forces implementers to spend more time analyzing and planning their interfaces to ensure that both parties are using the same optional features.

Version 3 addresses these and other issues by using a methodology based on a reference information (i.e., data) model. Using rigorous analytic and message building techniques and incorporating more trigger events and message formats with very little optionality, HL7's primary goal for Version 3 is to offer a standard that is definite and testable, and provide the ability to certify a vendor's conformance. However, there have thus far not been successful implementations of version 3 in operational systems.

4.4.2. Digital Imaging Communications in Medicine, DICOM

Digital medical image sources, and the use of computers to process them after their acquisition, were introduced in the seventies. In 1983 the American College of Radiology (ACR) [39] and the National Electrical Manufacturers Association (NEMA) [40] formed a joint committee in order to standardise a method for the transmission of medical images and their associated information. 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 towards a standard method of communicating digital image information. It also introduced the name DICOM [7] (Digital Imaging and Communications in Medicine).

DICOM is now an international standards organization creating and maintaining standards for 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.

DICOM aims at achieving compatibility and improving workflow efficiency between imaging systems and other health information systems. Every major diagnostic medical imaging vendor in the world has incorporated the standard into their product design and they participate in its enhancement.

Since its origin, DICOM has paid much attention to establishing working relationships with other related standard initiatives throughout the world:

- ASTM for its initial version;
- the Internet protocol TCP/IP in 1993;
- CEN in the nineties (this solid co-operation resulting in a number of jointly developed supplements);
- JIRA (the Japan Industries Association of Radiological Systems) [41] for the convergence of a Japanese interchange media format (IS&C) with DICOM;
- ANSI-HISB [42] in the USA, from which DICOM adopted a harmonised patient name structure.
- HL7 resulting in the creation of a joint DICOM-HL7 working group in 1999;
- ISO/TC215, with which a Type A liaison has been established in 1999, shortly after its creation. ISO/TC215 has not created a working group for bio-medical imaging standards, but is relying instead on DICOM.

DICOM has 26 Working Groups:

- WG-01 Cardiac and Vascular Information
- WG-02 Projection Radiography and Angiography
- WG-03 Nuclear Medicine
- WG-04 Compression
- WG-05 Exchange Media
- WG-06 Base Standard
- WG-07 Radiotherapy
- WG-08 Structured Reporting
- WG-09 Ophthalmology
- WG-10 Strategic Advisory
- WG-11 Display Function Standard
- WG-12 Ultrasound
- WG-13 Visible Light
- WG-14 Security
- WG-15 Digital Mammography and CAD
- WG-16 Magnetic Resonance
- WG-17 3D
- WG-18 Clinical Trials and Education
- WG-19 Dermatologic Standards
- WG-20 Integration of Imaging and Information Systems
- WG-21 Computed Tomography
- WG-22 Dentistry
- WG-23 Application Hosting
- WG-24 Surgery
- WG-25 Veterinary Medicine
- WG-26 Pathology

The current priorities for DICOM [43] are issues relating to security, performance, new modality technology, structured and coded documents for specific clinical domains, and workflow management.

4.4.3. 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), but through its predecessors dates back to 1884 [10]. AIEE addressed wire communications, light and power systems, while IRE, itself resulting from the merging of two largely local organizations (the Society of Wireless and Telegraph Engineers and the Wireless Institute), addressed wireless communications.

IEEE has undertaken standardization activities in the United States via its subsidiary, the Institute of Electrical and Electronics Engineers Standards Association (IEEE-SA) [44], which develops industry standards in a broad-range of industries, including Biomedical and Healthcare.

Collaboration exists between the IEEE, CEN/TC251 and ISO/TC215. Working with ISO/TC215, and in accordance with the ISO/IEEE "Pilot Project", international representatives can participate in ballots via 'international co-ordination'. The votes are not binding (i.e. they are not counted in the final tally that determines the result of the ballot). A large suite of standards has been developed and published jointly by IEEE, CEN and ISO

The main efforts of IEEE in the area of health care standards are:

- 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 being worked on by most HISB members is part of this effort.

The collaboration between IEEE and ISO/TC215 results in standards that are partitioned into layers that may be combined as necessary to provide the communications appropriate for a given device. These standards are generally broken into three key areas:

1. device data/semantics (ISO/IEEE 11073-1xxxx series);
2. general communication services (ISO/IEEE 11073-2xxxx series);
3. transports (ISO/IEEE 11073-3xxxx series).

Standards from these three primary areas may be combined as necessary to create a full 7-layer communications stack that provides plug-and-play interoperability.

4.4.4. The American Society for Testing Materials (ASTM)

ASTM International [9], formerly the American Society for Testing Materials, develops standards under ANSI. It was founded in 1898 and forms today 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 entities, are members. ASTM individuals are producers, users, consumers, and representatives of government and academia.

ASTM/E31 is the technical committee responsible for Healthcare Informatics. It has published several standards that 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 messages with the original use of supporting electronic patient care referrals among healthcare providers. It is now seen as having archival value within and Electric Health Records repository. In July 2004, ASTM agreed to harmonise their CCR with the HL7 work on the Clinical Document Architecture (CDA), to produce an implementation guide on the so-called Clinical Care Document (CCD).

4.4.5. OMG/CORBA

As part of its open, vendor-independent specification for an architecture and infrastructure CORBA (Common Object Request Broker Architecture), the Object Management Group (OMG) is developing several services for health care [45]. Several CORBA specifications are available:

- Healthcare DTF Roadmap V1.0b;
- Person Identification Service (PIDS);
- Terminology Query Service (TQS);
- Clinical Observations Access Service (COAS);
- Resource Access Decision (RAD);
- Clinical Image Access Service (CIAS).

Others are under development:

- Healthcare DTF Roadmap V2.0;
- Healthcare DTF Toolkit 2.0 release;
- Summary List Management Service (SLIMS);
- Health Information Locator Service (HILS);
- Health Data Interpretation Facility (HDIF);
- Medical Transcription Management (MTM);
- Order Entry/Tracking Service (OE/TS);
- Party Management Facility;
- Pharmacy Interaction Service (PIF);
- . Remittance Management Service;
- Charge Capture Management Service;
- Claims Management Service;
- Person Demographics Service;
- Healthcare Relationship Service;

- . Health Benefit Plan Management Service;
- . Eligibility Service;
- Care Authorization Management Service;
- Enrolment Management Service.

4.5 Standards supporting initiatives

4.5.1. ONCHIT - Office of the National Coordinator for Health information Technology

On April 27, 2004 the President committed to reducing medical errors, lowering medical costs and providing better information for consumers and physicians through a commitment to the Health Information Technology. This EO [46] directed Secretary Mike Leavitt of Health and Human Services (HHS) to establish the position of the Office of the National Coordinator for Health information Technology (ONCHIT) [47].

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.

ONCHIT is headed by the National Coordinator for Health Information Technology who serves as the Secretary's principal advisor on the development, application, and use of health information technology; coordinates the Department of Health and Human Services' (HHS) health information technology programs; ensures that HHS health information technology policy and programs are coordinated with those of other relevant executive branch agencies; and to the extent permitted by law, develops, maintains, and directs the implementation of a strategic plan to guide the nationwide implementation of interoperable health information technology in both the public and private health care sectors that will reduce medical errors, improve quality, and produce greater value for health care expenditures, and coordinates outreach and consultation by the relevant executive branch agencies with the public and private sectors. He further provides comments and advice at the request of OMB regarding specific Federal health information technology programs.

4.5.1.1 ONCHIT's strategic plan for 2008-2012

June 3, 2008, ONCHIT released its plan for 2008-2012 [48].

The Plan has two goals: Patient-focused Health Care and Population Health, with four objectives under each goal. The themes of privacy and security, interoperability, IT adoption, and collaborative governance recur across the goals, but they apply in very different ways to health care and population health.

Goal 1) Patient-focused health care: Enable the transformation to higher quality, more cost-efficient, patient-focused health care through electronic health information access and use by care providers, and by patients and their designees.

> Objective 1.1 – Privacy and Security: Facilitate electronic exchange, access, and use of electronic health information while protecting the privacy and security of patients' health information

Objective 1.2 – Interoperability: Enable the movement of electronic health information to where and when it is needed to support individual health and care needs

Objective 1.3 – Adoption: Promote nationwide deployment of electronic health records and personal health records that put information to use in support of health and care

Objective 1.4 – Collaborative Governance: Establish mechanisms for multi-stakeholder priority-setting and decision-making to guide development of the nation's health IT infrastructure

Goal 2) Population health: Enable the appropriate, authorized, and timely access and use of electronic health information to benefit public health, biomedical research, quality improvement, and emergency preparedness.

Objective 2.1 – Privacy and Security: Advance privacy and security policies, principles, procedures, and protections for information access and use in population health

Objective 2.2 – Interoperability: Enable the mobility of health information to support population-oriented uses

Objective 2.3 – Adoption: Promote nationwide adoption of technologies and technical functions that will improve population and individual health

Objective 2.4 – Collaborative Governance: Establish coordinated organizational processes supporting information use for population health

The Plan further articulates 43 strategies that describe the work needed to achieve each objective. Each strategy is associated with a milestone against which progress can be assessed, and a set of illustrative actions to implement each strategy. As a group, the strategies are characterized by:

- Commitment to the engagement of multiple stakeholders across the public and private sectors;
- Concern for reliability, confidentiality, privacy, and security when exchanging, storing, and using electronic health information; and
- Focus on the consumer of health care as a critical participant in achieving the two overarching goals of the Strategic Plan, as described above.

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

The Consolidated Health Informatics (CHI) initiative [49, 50] is establishing a portfolio of existing clinical vocabularies and messaging standards believed to enable federal agencies to build interoperable health data systems. This commonality is hoped to enable all federal agencies to "speak the same language" and share information without the high cost of translation or data re-entry. Federal agencies could then pursue projects meeting their individual business needs aimed at initiatives 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 department/agencies including HHS, VA, DOD, SSA, GSA, and NIST are active in the CHI governance process. Through the CHI governance process, all federal agencies are

incorporating the adopted standards into their individual agency health data enterprise architecture used to build all new systems or modify existing ones.

On March 21, 2003, the Departments of Health and Human Services, Defense, and Veterans Affairs announced the first set of uniform standards for the electronic exchange of clinical health information to be adopted across the federal government. The standards all federal agencies are adopting are:

- **Health Level 7** (HL7) messaging standards to ensure that each federal agency can share information that will improve coordinated care for patients such as entries of orders, scheduling appointments and tests and better coordination of the admittance, discharge and transfer of patients.
- **National Council on Prescription Drug Programs** (NCDCP) standards for ordering drugs from retail pharmacies to standardize information between health care providers and the pharmacies. These standards already have been adopted under the Health Insurance Portability and Accountability Act (HIPAA) of 1996, and ensure that the parts of the three federal departments that aren't covered by HIPAA will also use the same standards.
- The **Institute of Electrical and Electronics Engineers** 1073 (IEEE1073) series of standards that allow for health care providers to plug medical devices into information and computer systems that allow health care providers to monitor information from an ICU or through tele-health services on Indian reservations, and in other circumstances.
- **DICOM** standards that enable images and associated diagnostic information to be retrieved and transferred from various manufacturers' devices as well as medical staff workstations.
- **Laboratory Logical Observation Identifier Name Codes** (LOINC) to standardize the electronic exchange of clinical laboratory results.

On May 6, 2004, the Departments of Health and Human Services, Defense, and Veterans Affairs announced the possible adoption of 15 additional standards agreed to by the CHI initiative to allow for electronic exchange of clinical information across the federal government. The 15 new standards build on the existing set of five standards adopted by HHS in March 2003. The new standards agreed to by federal agencies are being used as agencies develop and implement new information technology systems. The specific new standards are:

- Health Level 7 (HL7) vocabulary standards for demographic information, units of measure, immunizations, and clinical encounters, and HL7's Clinical Document Architecture standard for text based reports. (Five standards)
- The College of American Pathologists Systematized Nomenclature of Medicine Clinical Terms (SNOMED CT) for laboratory result contents, non-laboratory interventions and procedures, anatomy, diagnosis and problems, and nursing. HHS is making SNOMED-CT available for use in the U.S. at no charge to users. (Five standards)
- Laboratory Logical Observation Identifier Name Codes (LOINC) to standardize the electronic exchange of laboratory test orders and drug label section headers. (One standard.)

- The Health Insurance Portability and Accountability Act (HIPAA) transactions and code sets for electronic exchange of health related information to perform billing or administrative functions. These are the same standards now required under HIPAA for health plans, health care clearinghouses and those health care providers who engage in certain electronic transactions. (One standard.)
- A set of federal terminologies related to medications, including the Food and Drug Administration's names and codes for ingredients, manufactured dosage forms, drug products and medication packages, the National Library of Medicine's RxNORM for describing clinical drugs, and the Veterans Administration's National Drug File Reference Terminology (NDF-RT) for specific drug classifications. (One standard.)
- The Human Gene Nomenclature (HUGN) for exchanging information regarding the role of genes in biomedical research in the federal health sector. (One standard.)
- The Environmental Protection Agency's Substance Registry System for non-medicinal chemicals of importance to health care. (One standard.)

As noted by Richesson and Krischer [51], the 2004 standards identification team postponed recommendations on many important data areas such as the physical exam, medical history, and adverse events.

4.5.3. American Health Information Community – (AHIC)

On September 13, 2005, Health and Human Services (HHS) Secretary Mike Leavitt announced the membership for the American Health Information Community [52]. AHIC was formed to help advance efforts to reach President Bush's call for most Americans to have electronic health records within ten years. It is a federally-chartered commission and provides input and recommendations to HHS on how to make health records digital and interoperable, and assure that the privacy and security of those records are protected, in a smooth, market-led way.

The development of recommendations by AHIC proceeds by identifying key work areas considered to have potential for breakthroughs in the advancement of standards that will lead to interoperability of health information. Once a work area is identified a corresponding workgroup 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.

4.5.3.1 AHIC's organisational structure

To date, seven AHIC workgroups have been created.

4.5.3.1.1 Population Health and Clinical Care Connections

- To make recommendations to AHIC to implement the informational tools and business operations to support real-time nationwide public health event monitoring and rapid response management across public health and care delivery communities and other authorized government agencies.
- 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.

4.5.3.1.2 Consumer empowerment

- To make recommendations to AHIC to gain wide spread adoption of a personal health record that is easy-to-use, portable, longitudinal, affordable, and consumer-centred.
- To make recommendations to AHIC so that a pre-populated, consumer-directed and secure electronic registration summary is available to targeted populations. Make additional recommendations to AHIC so that a widely available pre-populated medication history linked to the registration summary is deployed.

4.5.3.1.3 Chronic care

- To make recommendations to AHIC to deploy widely available, secure technologies solutions for remote monitoring and assessment of patients and for communication between clinicians regarding patients.
- To make recommendations to AHIC so that widespread use of secure messaging, as appropriate, is fostered as a means of communication between clinicians and patients about care delivery.

4.5.3.1.4 Electronic health records

- To make recommendations to AHIC on ways to achieve widespread adoption of certified EHRs, minimizing gaps in adoption among providers.
- To make recommendations to AHIC so that standardized, widely available and secure solutions for accessing current and historical laboratory results and interpretations is deployed for clinical care by authorized parties.

4.5.3.1.5 Confidentiality, Privacy & Security

- To make recommendations to AHIC regarding the protection of personal health information in order to secure trust, and support appropriate interoperable electronic health information exchange.
- To make actionable confidentiality, privacy, and security recommendations to AHIC on specific policies that best balance the needs between appropriate information protection and access to support, and accelerate the implementation of the consumer empowerment, chronic care, and electronic health record related breakthroughs.

4.5.3.1.6 Quality

- To make recommendations to AHIC so that health IT can provide the data needed for the development of quality measures that are useful to patients and others in the health care industry, automate the measurement and reporting of a comprehensive current and future set of quality measures, and accelerate the use of clinical decision support that can

improve performance on those quality measures. Also, make recommendations for how performance measures should align with the capabilities and limitations of health IT.

- 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.

4.5.3.1.7 Personalized Healthcare

- To make recommendations to AHIC for a process to foster a broad, community-based approach to establish 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.
- 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 federal government agencies.

4.5.3.2 AHIC's working principles: development of use case descriptions

A Use Case provides 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). The use cases serve to provide detailed guidance on the functions needed to advance critical efforts for the accelerated adoption of health information technology as follows [53]:

- From AHIC's perspective, the harmonized use cases yield valuable insights into the Community's continuing efforts to identify and remove barriers to adoption of health information technology.
- For the nationwide health information network consortia, the harmonized use cases provide a foundation for the identification of critical architecture elements and establish the expectations for their prototype architectures.
- For the Health Information Technology Standards Panel, the harmonized use cases scope its efforts to develop named standards and implementation level guidance necessary for interoperable solutions.
- For the Certification Commission for Health Information Technology, the harmonized use cases provide insight into criteria for the certification of electronic health records and other aspects of the health IT landscape.

The existing use cases (listed by year) are:

2006 Use Cases

- Harmonized Biosurveillance (Visit, Utilization and Lab Result Data): Transmit essential ambulatory care and emergency department visit, utilization, and lab result data from electronically enabled health care delivery and public health systems in standardized and anonymized format to authorized public health agencies within 24 hours.

- Harmonized Consumer Empowerment (Registration & Medication History): Deploy to targeted populations a pre-populated, consumer-directed and secure electronic registration summary. Deploy a widely available pre-populated medication history linked to the registration summary.
- Harmonized Electronic Health Record (Laboratory Result Reporting): Deploy standardized, widely available, secure solutions for accessing current and historical laboratory results and interpretations in a patient-centric manner for clinical care by authorized parties.

2007 Use Cases

- Emergency Responder – Electronic Medical Records: Provide timely electronic access and exchange of critical health information to support the assessment, stabilization and treatment of the victims of emergency incidents, as well as, on a treatment non-interference basis, to facilitate family member reunification and expedite next-of-kin notification following such incidents.
- Consumer Empowerment – Access to Clinical Information: Provide the consumer with the ability to access and incorporate their available health information from other sources into their PHR, to create and/or update their list of providers and determine the access permissions that should be granted to those providers for the purpose of accessing information in their PHR, and to transfer information from an existing PHR to another PHR including the transfer of provider lists and provider permissions.
- Medication Management: Enable patient medication and allergies information exchange and the sharing of that information between consumers, clinicians (in multiple sites and settings of care), pharmacists (dispensers) and organizations that provide health insurance and provide pharmacy benefits (payers).
- Quality: Provide the capabilities and functionality needed to measure and report on hospital and clinician quality and the use of quality measures to support clinical decision making in an interoperable healthcare system.

2008 Use Cases

- Remote Monitoring: Enable the patient to gather and communicate remote monitoring information electronically from measurement devices in a home or other non-clinical setting to a clinician's ambulatory EHR system and/or to the patient's PHR.
- Patient – Provider Secure Messaging: Provide the processes and information needed for patient – provider secure messaging which would include both secure messages sent from patients to providers as well as secure messages sent from providers to patients.
- Personalized Healthcare: Enable the processes of personalized healthcare by which healthcare providers can customize treatment and management plans for patients based on their unique genetic makeup including the exchange of genetic/genomic test information, personal and family health history.
- Consultation and Transfer of Care: Enable the electronic exchange of information between clinicians, particularly between requesting clinicians and consulting clinicians,

to support consultations such as specialty services and second opinions and to support transfers of care.

- Public Health Case Reporting: Address population health relating to aspects of Public Health Case (PH Case) reporting and Adverse Event (AE) reporting which may include the reporting of communicable/infectious and non-infectious diseases/conditions and the reporting of AEs associated with post-market vaccines and medications.
- Immunization & Response Management: Enable the exchange of information supporting the distribution and administration of medications, vaccinations, and other specific medical prophylaxis and treatment methods.

4.5.4. Healthcare Information Technology Standards Panel (HITSP)

Of particular relevance to the discussion on the adoption of standards is the Healthcare Information Technology Standards Panel [54]. The mission of HITSP 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. Comprised of a wide range of stakeholders, the Panel 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 the American National Standards Institute (ANSI) in cooperation with strategic partners such as the Healthcare Information and Management Systems Society (HIMSS) [55], 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 Evaluation of Standards Harmonization Process for Health Information Technology Program (ONCHIT-1) brings together the intellectual assets of organizations with a stake in health data standards. With the oversight and backing of the National Health Information Technology Coordinator, ONCHIT-1 guides the collaboration of these organizations toward a healthcare information technology (HIT) standards harmonization process.

The standardization process is based on four iterative functions.

- Use Case Development: Each Use Case Committee member will attend facilitated sessions to develop use cases according to the ONC template. HITSP will review and approve them. There will be collaboration through the project office with other ONC contracts. This annual process is planned to be used for years 2, 3 and beyond.
- Gap Analysis Process: This process will define and resolve standards gaps and overlaps. The Use Case Committees will identify the gaps and overlaps and forward to HITSP for resolution. Once all standards are identified for a use case, it will move into the Implementation Guide Development process.
- Implementation Guidelines Development: Implementation Guides will be developed, approved by HITSP and endorsed by AHIC. Implementation Guides will not be finalized until after a public comment period, a testing process and refinement. Input will be received from ONC and the other contractors.

- Testing: Unit test scripts and test tools will be developed. After completing unit testing, use cases will be paired together to create end to end test scripts in preparation for a Connectathon-like testing event. This validation process will work out the technical details. Results of the testing event will be reviewed/approved by HITSP and used to make required updates to the implementation guides.

The identification of standards by HITSP is recorded in interoperability specifications that specify how and what standards should be used for a particular use case. Seven such interoperability specifications have been released to date:

- Biosurveillance
- Consumer Empowerment – Registration and Medication History
- Electronic Health Records – Lab Result Reporting
- Emergency Responder Electronic Health Record
- Consumer Empowerment – Access to Clinical Information via Media
- Quality
- Medication Management

The vocabulary standards identified within these specifications are:

- American Medical Association's Healthcare Provider Taxonomy
- American Medical Association's Current Procedural Terminology (CPT®) Fourth Edition (CPT-4)
- Health Level Seven (HL7) Version 3.0 Privacy Consent related specifications: RCMR_RM010001 - Data Consent, RCMR_RM010002 - Shared Consent, COCT_RM580000 - Data Consent
- Healthcare Common Procedure Coding System Level II Code Set
- ICD-10-PCS,
- ICD-10-CM,
- ICD-9-CM
- SNOMED-CT
- National Council for Prescription Drug Programs (NCPDP) SCRIPT Standard - 8.1(expected to transition to a version 10.x in the future)
- National Library of Medicine's RxNorm
- National Uniform Billing Committee (NUBC) Uniform Bill Version 1992 (UB-92) Current UB Data Specification Manual Field 22, Patient Discharge Status, Codes
- National Uniform Billing Committee (NUBC) Uniform Bill Version 2007 (UB-04) Current UB Data Specification Manual Field 22, Patient Discharge Status, Codes

4.5.5. 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 [56]. 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) [57], the Healthcare Information and Management Systems Society (HIMSS) [55] and the National Alliance for Health Information Technology (Alliance) [58]. In the following year, additional funding was supplied by the American Academy of Family Physicians (AAFP) [59], the American Academy of Pediatrics (AAP) [60], the American College of Physicians (ACP) [61], the California Healthcare Foundation (CHCF) [62], 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.

There are currently eleven workgroups within CCHIT: Ambulatory EHR, Behavioral Health, Cardiovascular Medicine, Child Health, Emergency Department, Inpatient EHR, Interoperability, Network, Personal Health, Privacy & Compliance, and Security. The criteria set by workgroups are based upon available standards. In cases where there are overlapping standards in a given area, the workgroups take the guidance from the harmonized standards in HITSP Interoperability Specifications.

The 2007 Ambulatory Interoperability Criteria [63] show their close alignment with the HITSP specifications which were based on the CHI recommendations. By category the criteria are:

Category	Source or Reference
Laboratory	HL7 v2.x, LOINC, ELINCS v2.1, SNOMED, HL7 CDA R2, IHE XDS-Lab
Imaging	IHE XDS-I Cross-Enterprise Image Information Sharing Integration profile, HL7 v2.x, IHE XDS-I Procedure
Medications	NCPDP Script 8.1, NCPDP Script 11.1 (not available), X12 270/271/CORE Phase 1 Rules, NCPDP Formulary and Benefit Standard Implementation Guide v1.0, HL7-ASTM CCD, IHE XDS-XPHR, ASTM CCR
Immunizations	HL7-ASTM CCD, IHE XDS-XPHR
Clinical Documentation	IHE Cross-Enterprise Document Sharing (XDS) integration profile, HL7 CDA R2, HL7-ASTM CCD, ASTM CCR, IHE-XPHR
Secondary Uses of Clinical Data	CDC Disease registries, Public Health Information Network (PHIN)
Administrative and	X12 270/271/Core Phase I Rules, IHE PIX profile, IHE PDQ, X12, HL7, HL7 2.4 Patient Administration, X12N 4010, HL7 2.4 Scheduling, X12

Financial Data	278 – Health Care Services Review: Referral Certification and Authorization – Dental, Professional Institutional
Clinical Trials	NCI caBIG, CDISC

4.5.6. IHE — Integrating the Healthcare Enterprise

Started by HIMSS and RSNA, IHE is a spontaneous initiative, organised to improve the integration of systems [64]. It aims at providing a process for a co-ordinated adoption of standards: clinicians and IT staff define needs; vendors develop solutions (a technical Framework). 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 also exist currently with the American College of Cardiology (ACC), American College of Clinical Engineering (ACCE), HL7, and DICOM. Several individual members take part as well.

IHE is an independent private initiative that results from a partnership without formal legal status between these vendors and professional societies. They collectively manage its budget. Participants from the software industry voluntarily participate, for a fee, to the testing and demonstration process, with a return on their investment in the form of a reduced installation efforts and a commercial advantage. Of course users' organisations also contribute at their own expense. This initiative aims at speeding up the rate and quality of integration in healthcare environments, fostering communication among vendors, proving that integration is attainable based on standards, and improving the efficiency and effectiveness of clinical practice.

The needs for the IHE initiative comes from the statement 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 and eventually they may be costly to implement and to maintain. IHE delivers integration profiles built on existing standards. IHE makes it clear that it is 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 SDOs.

IHE is not simply a demonstration project, for IHE demonstrations represent only one means to the end of adoption of integration profiles and standards. These demonstrations are backed up by documentation, tools, testing, and publication of information.

The IHE initiative is both an intra-enterprise and cross-enterprise, bottom-up approach supporting a multi-year, standards based, vendor neutral project that creates a framework to seamlessly convey vital information from application to application, system to system, and setting to setting. The foreseen benefits claimed by the IHE initiative for its participants do not differ from those of standards in general, but the emphasis is put on the practical limitations in the implementation of standards.

An IHE Integration Profile organises a set of coordinated, standards-based transactions between a subset of the functional components of health organisations 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 It is now considering Nuclear Medicine and the exchange of clinical documents across the borders of an enterprise (eHealth).

IHE profiles are devised in an intensive process based on a stepwise approach, according to annual cycles:

- the development of profiles is done at the global level, by an open group of volunteer users and vendors;
- the deployment is organized by (world) regions, and by countries, based on national 'chapters';
- Connect-a-thons are organized at the 'regional' level (a number of national ProRec centres co-operate with IHE).

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

4.5.7. NCRR - National Center for Research Resources

The NCRR [65] was formed on February 15, 1990 when then Secretary of the US Department of Health and Human Services, Dr. Louis W. Sullivan approved the merger of the Division of Research Resources and the NIH Division of Research Services. The mission of the NCRR 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 NCRR 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.

4.5.7.1 CTSA - Clinical and Translational Science Awards

An initiative within the Clinical Research Division, the CTSA program [66], is building a consortium of 60 institutions designed to speed the process by which biomedical discoveries are translated into effective medical care for patients.

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, CTSAs are expected to transform the conduct of translational medicine in the United States. A major hurdle in the way to accomplishing this is the integration of data from patient care systems with that from clinical research systems.

As of January 2008 [67], the CTSA program was still in the formative stages of clearing this hurdle. For example, the Inventory and Resources Project Group is in the process of developing an inventory of informatics resources at CTSAs to assist informatics staff in finding available tools. The Human Studies Metadata Repository Project Group is focused on developing agreed

on standards for a set of metadata about clinical research studies at CTSAs starting with the development of use cases to drive the focus of the group. It is hoped that the creation of a Operations committee that will govern the National Informatics Steering committee by setting strategic directions, prioritize activities and insure timely deliverables will facilitate progress in this critical task.

4.5.8. USHIK - United States Health Information Knowledgebase

The United States Health Information Knowledgebase (USHIK) [68] is a project funded by the Agency for Healthcare Research and Quality (AHRQ) [69] 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 to healthcare standards. 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 contains the information models of:

- Accredited Standards Committee X12 Subcommittee N - Insurance (X12N : Insurance)
- American Dental Association
- Clinical Care Classification (CCC) System
- Consolidated Health Informatics
- Federal Health Information Exchange
- Healthcare Information Technology Standards Panel
- Health Level Seven®
- MHS Functional Area Model - DATA
- National Committee on Vital and Health Statistics
- National Council for Prescription Drug Programs
- National Council for Prescription Drug Programs SCRIPT
- National Health Information Model of Australia
- Health Care Service Data Reporting Guide
- Small Scale Harmonization Project

and stores the meta-data for 13087 data elements.

The web interface allows for 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 status, 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 needing to synchronize large information models.

4.5.9. CMS - Centers for Medicare & Medicaid Services

The CMS [70] is the agency with the US Department of Health and Human Services that has the important responsibilities of 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, there are two programs that the CMS governs that are significant: The Medicaid Information Technology Architecture and in conjunction with the CDC's National Center for Health Statistics (NCHS) the development and maintenance of the International Classification of Diseases Clinical Modifications and Procedure Classification System.

4.5.9.1 MITA - Medicaid Information Technology Architecture

The MITA is an initiative having the goal of transforming the business and information technology of the Medicaid enterprise. The hoped for end result of the project is to be a set of guidelines on which a national architecture of information systems can be built that improve both the quality and efficiency of health care. Critical to the success of the 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 [71]. 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 be required for use (HIPPA required) or as emerging potential candidates for standards. The methodology for adopting standards will proceed according to the following guidelines:

- Identify standards in use by state Medicaid systems
- Align existing standards to data models and messages
- Develop new standards only when no existing standard is available
- Adopt standards in the following order of priority
    - International Standards
    - National Standards
    - Industry/Healthcare Standards
    - MITA/State developed Standards
- Adopt a minimum standard that is useable by the maximum number of state Medicaid enterprises and allow for extensions of standards by states to suit their regulatory environments
- Allow for versioning of standards
- Develop new standards, when needed, in conjunction with states and vendors
- Attempt to submit developed standards to SDOs for adoption

- Maintain mappings of data standards to data models and messages in a MITA repository

The vocabulary standards listed as having a status of either current (required by HIPAA) or emerging relevance to MITA are:

Standard	Description	Status
CDT	Codes on Dental Procedures and Nomenclature	Current
CPT-4	Current Procedural Terminology, Fourth Edition	Current
DRG	Diagnosis Related Groups	Current
HCPCS	Healthcare Common Procedure Coding System	Current
ICD-9-CM Vol. 1 & 2	International Classification of Diseases, 9th Edition, Clinical Modification	Current
ICD-9-CM Vol. 3	International Classification of Diseases, 9th Edition, Procedural Classification	Current
NDC	National Drug Codes	Current
LOINC	Logical Observation Identifiers, Names and Codes	Current – HIPAA 275 will likely require
SNOMED	Systematized Nomenclature of Medicine	Emerging
Non-Medical Code Sets used in X12 HIPAA Transactions	Includes Provider Taxonomy Codes, Claim Status Codes, Country Codes, Facility Codes, and Revenue Codes	Current and Emerging – HIPAA will require

The messaging standards listed as having a status of either current (required by HIPAA) or emerging relevance to MITA are:

Standard	Description	Status
X12N	X12 Insurance Committee	Current
HL7 V2	Health Level 7 Version 2.x messages	Emerging
HL7 V3	Health Level 7 Version 3 messages	Emerging
CCOW	Health Level 7 Clinical Context Object Workgroup, Management Specification Version 1.4	Emerging
CDA	Health Level 7 Clinical Document Architecture	Emerging
DICOM	Digital Imaging and Communications in Medicine	Emerging
ASTM E1762-95	American Society for Testing and Materials Standard Guide for Electronic Authentication of Health Care Information	Emerging
IEEE 1073	Institute of Electrical and Electronics Engineers, 1073 Standard for Medical Device Communications	Emerging
NCPDP	National Council for Prescription Drug Programs Standards	Current
Arden Syntax	Health Level 7 clinical decision support standard	Emerging

As of June 2008, no definitive selections of data standards have been made.

4.5.9.2 International Classification of Diseases 9th and 10th Editions

The CMS and the National Center for Health Statistics (NCHS) [72] oversee the maintenance and production of the ICD-9-CM volumes 1, 2, and 3 and the ICD-10-CM and ICD-10-PCS. 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 in the United States. Used in conjunction with the UB-92 reimbursement form for hospitals and the HCFA-1500 form for physicians, one or more diagnostic codes and the codes for the related treatment procedures are submitted to payers who then match the submission to one of many diagnostic related codes to determine the amount of reimbursement.

The ICD-9-CM classification system was the result of an expansion of the ICD-9 system by the NCHS to include clinical modifications. The ICD-9-CM vols. 1 & 2 are comprised of a list of diagnostic codes in all healthcare setting and ICD-9-CM vol. 3 of a list of inpatient procedure codes reported by hospitals. When WHO released ICD-10 in 1992 it included most, but not all of the additions made in the expansion from ICD-9 to ICD-9-CM vols. 1 & 2. To create ICD-10-CM the NCHS added the codes from ICD-9-CM that were absent from ICD-10 and fit them into the new coding system.

The transition from ICD-9-CM vol. 3 to ICD-10-PCS was necessitated because the earlier system was simply running out of room for adding new procedure codes. The ICD-9-CM vol. 3 codes are rendered as four character strings of the form XX.XX. While only 3,500 procedures are annotated within the system out of a total of 10,000 possible, many of the 2-digit categories were exhausted. For example, cardiology procedures which occupy the code space of 37.XX became full (i.e. more than 100 such procedures were annotated), forcing developers to use a previously unused portion of the code space (i.e. 00.XX) when adding new ones. This resulted in requiring users to search in both spaces to find the procedures they were seeking.

ICD-10-PCS was created anew by 3M Health Systems between 1995 and 1998. The number of procedures included in the new system equals nearly 200,000, a vast expansion on the 3,500 contained in the old. The principles claimed to guide the development are that the system would be:

- Comprehensive – all procedures are classified
- Unique – all substantially different procedures have a unique code
- Expandable – new procedures can be incorporated as new codes
- Hierarchical – procedures can be aggregated into larger categories
- Standardized – all terms are precisely defined
- Exclusively procedural – definitions contain no diagnostic information
- Multi-axial – each code character has a consistent meaning.

Among the improvements of the new system cited by a Rand Corporation evaluation of the costs and benefits of transitioning from ICD-9 to ICD-10 [73] were a reduction of the use of NOS, and NEC qualifiers, avoidance of compound procedure sets, and increased granularity. The same

study cited user issues of difficulties of navigation and memorization of the new system's hierarchies, and a high degree of complexity in creating crosswalks between the new and old systems. In all, ICD-10-CM and ICD-10-PCS have been recommended by agencies such as CHI, HITSP, and NCVHS for use in US government health IT initiatives. The Rand Corporation cited above found that is was probable that the potential cost benefits of transitioning to the new systems outweighed the cost burdens of doing so. To date, no time schedule exists for the adoption of the new system.

4.5.10. 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 [74].

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's stated role in the PHIN is:

- 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 PHIN with other federal initiatives.

The CDC has, in collaboration with state and local health departments 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 vocabularies used within the PHIN.

The focus of the Messaging Services Team [75] 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 [76] 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 questions 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 Public Health Document Architecture.

The purpose of the PHIN VADS [77] 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 are based upon the recommendations of CHI. Files can be downloaded in a variety of standard formats including tab-delimited, Excel, or XML.

4.5.11. Public Health Data Standards Consortium – PHDSC

The PHDSC [78] 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 such 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 the collaborations of the PHDSC include their membership in HITSP, and their participation in the standards development process of HL7, ASC X12, the National Unified Billing Committee (NUBC) [79] and the National Uniform Claim Committee (NUCC) [80].

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 Guide
- Sub-Committee on Payer Typology developed and maintains a payer typology to allows 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) is working on developing an educational strategy on the importance of external causes of injury codes reporting from the emergency rooms chief complaint data to the state and local health agencies.

The mission of the Health Care Service Data Reporting Guide Committee [81] (HCSDR Guide Committee) 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 that are compatible with the 837 Health Claim transaction set standards. These are the standards identified by the Health Insurance Portability and Accountability Act (HIPAA). In addition, the guide includes data elements that are not now needed for the payment of a claim such as race & ethnicity, and patient county code 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 [82] is to create a payer type standard to allow consistent reporting of payer data to public health agencies for health care services and research. The work of this committee is in response to there being no current standard for classifying the sources of payment data and is an acknowledgement that having such a standard is critical for examining the effects of payment policies.

The committee has created the Source of Payment Typology. The typology is said to have an organizational structure similar to that of the ICD classification system and identifies general payer categories which 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 [83] 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 to measure progress on injury and violence prevention and control that is currently hindered by the lack of standardized external cause of injury codes.

To date, the ECIC has collaborated with the National Uniform Billing Committee (NUBC) in the successful petition of ASC X-12 to add ICD-10-CM and ICD-10-PCS. Future efforts of the committee are to identify and propose changes to the X-12 837 guides to provide the ability to utilize ICD-10-CM and ICD-10-PCS codes within the guides and capture additional external cause of injury codes.

4.5.12. National Uniform Billing Committee - NUBC

The NUBC [79] was formed in 1975 by the American Hospital Association (AHA) and consists of equal representation of provider organizations (e.g. AHA and state affiliates, Healthcare Financial Management Association and Federation of American Health Systems) and payer organizations (e.g. Health Care Financing Administration (HCFA), Medicaid, CHAMPUS, 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) more recently has become a member.

The objective of the NUBC was to develop a single billing form and standard data set that could be used nationwide by institutional providers and payers for handling health care claims. The first such form, the UB-82 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 [84] 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 - Codes used to identify conditions relating to a bill that may affect payer processing, such as whether a patient is homeless.
- Occurrence Codes - Codes and an associated date defining a significant event relating to this bill that may affect payer processing, such as an auto accident date.
- Occurrence Span Codes - Codes and the related dates that identify an event that relates to the payment of a claim, such as Skill Nursing Facility level of care dates.
- Revenue Codes - Code which identify a specific accommodation, ancillary service or billing calculation, such as emergency room charges.
- Value Codes -Codes which relate amounts or values to identified data elements necessary to process this claim as qualified by the payer organization, such as accident hour.

4.5.13. National Uniform Claim Committee - NUCC

The NUCC [80] is chaired by American Medical Association and consists of 12 voting members, including HCFA, 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.

5 STANDARDS ON THE CONTENT OF PATIENT SUMMARIES

Three major initiatives are particularly significant for the development of a standard on the content of Patient Summary, involving leading organizations in the standardization arena:

- the Continuity of Care Record (CCR) by ASTM;
- the Care Record Summary (CRS) by HL7;
- the Cross-Enterprise Sharing of Medical Summaries (XDS-MS) by IHE.

The efforts of ASTM-CCR and HL7-CRS are now merging into the HL7/ASTM Clinical Care Document (CCD).

5.1 ASTM Continuity of Care Record (CCR)

The ASTM Continuity of Care Record (CCR) [85] was designed and implemented as a standard for a comprehensive data summary that aggregates data from multiple sources, health care records, medical legal documents, and health care encounters to form a comprehensive overall clinical picture of a patient's current and relevant historical health care status. It is officially balloted and approved as ASTM Standard E2369-05.

The intended uses for the CCR are:

- As a detailed health care summary that a provider can generate and give to the patient at the end of a health care encounter (inpatient, outpatient, or ambulatory care). This can be either in paper or electronic form.
- As a data extract from an EHR, HIS (Hospital Information System), ePrescribing system, data registry, or a PHR (Personal Health Record) so that a patient health care summary can be transferred to another such system.
- To break down barriers to EHR adoption through facilitating the ability of an EHR purchaser to change to another EHR vendor, if desired, by exporting the critical medical information so that all future encounters on the system will have required summary information. The CCR is intended to increase adoption of EHRs by reducing the risk of choosing the 'wrong' EHR and reducing the cost of sales and number of providers who are unable to make decisions by reducing concern over the financial health and future of an EHR company. The CCR can also facilitate incremental pathways to an EHR by allowing a practice or provider to begin with an electronic prescribing system or immunization tracking system and then export the data from those systems when a full EHR is implemented.
- As a model for EHR and PHR data and data objects.
- As a complete patient health care summary to accompany medical legal and administrative documents for patient admission, discharge, or transfer to/from a health care facility.

5.1.1. HL7/ ASTM Continuity of Care Document (CCD) and HL7 Care Record Summary (CRS)

The Continuity of Care Document (CCD) is an HL7 CDA document containing the American Society for Testing and Materials (ASTM) Continuity of Care Record (CCR). CCR is ASTM's active standard in response to the need to organize and make transportable the most relevant and timely facts about a patient's condition. Briefly, the CCR includes patient and provider information, insurance information, patient's health status (e.g., allergies, medications, vital signs, diagnoses, recent procedures), recent care provided, future care (care plan) recommendations, and the reason for referral or transfer.

The HL7 Clinical Document Architecture (CDA) is a document architecture standard designed to represent medical legal health care encounter documents in a standardized format. CDA r2 (Release 2) was balloted and approved in June 2005. The HL7 Care Record Summary (CRS) was proposed as a special use-case of the CDA as a care record summary like the CCR. The CRS ballot did not pass and the CRS has been reconfigured as a Discharge Summary, a medical legal document, and a Referral Document, a quasi-medical legal document, for the IHE HIMSS demonstration and for the usage in XDS-MS (see below).

The intended use for the CDA/CRS is for point-to-point/trading partner-to-trading partner exchange of documents (provider-to-provider or institution-to/from-provider), particularly medical legal documents.

By leveraging the experience of the two groups, the implementation guide on the Clinical Care Document puts the content described by CCR into the architecture provided by CDA.

5.2 IHE Cross-Enterprise Sharing of Medical Summaries (XDS-MS)

The IHE Cross-Enterprise Sharing of Medical Summaries (XDS-MS) is a profile within the IHE Patient Care Coordination (PCC) Framework [86].

In XDS-MS, currently two types of Medical Summary content are specified: one for episodic care, the other for collaborative care. XDS-MS specifies content of Medical summaries by further constraining the HL7 Clinical Document Architecture (CDA) standard and Care Record Summary (CRS) CDA implementation guides. Medical summaries are shared by storing them in ebXML Registry/Repositories. Medical summaries are rendered by using XML Style Sheets.

The further issues involved in sharing medical summaries such as mapping patient identifiers, cross enterprise user authentication, authentication of nodes, obtaining audit trails, making sure that the interacting computers have consistent time are addressed by grouping the Actors of XDS-MS with the relevant Actors of the related IHE Profiles.

Cross-Enterprise Sharing of Medical Summaries (XDS-MS) is a mechanism to automate sharing of Medical Summaries between care providers. The main characteristics of XDS-MS are as follows:

- Two types of Medical Summary content are currently specified: one for episodic care, the other for collaborative care.
- A third type of Medical Summary for permanent care is yet to be defined by IHE.

- XDS-MS specifies content of Medical summaries by building on and further constraining the HL7 Clinical Document Architecture (CDA) standard and Care Record Summary (CRS) CDA implementation guides.
- Document Sources provide an XML stylesheet to render the content of the Medical Summary document.
- Medical summaries are shared within predefined domains (called XDS Affinity Domains) by storing the medical summaries in Registry/Repositories. Note however that IHE also plans the federated XDS Affinity domains; therefore the exchange of medical documents will not be restricted to XDS Affinity Domains in the near future.
- Registry/Repository architectures facilitate the discovery of the Medical Summaries in an XDS Affinity Domain.
- XDS-MS Profile uses the Actors and Transactions of IHE XDS; only the Document types used in XDS-MS are more specific Medical Summaries.

In summary, the IHE Cross-Enterprise Sharing of Medical Summary (XDS-MS) Integration Profile defines the appropriate standards for document transmission and a minimum set of "record entries" that should be forwarded or made available to subsequent care provider(s).

XDS-MS uses Cross Enterprise Document Sharing (IHE-XDS) Profile's Actor's and Transactions; however the document content is defined in a more specific way. The document content builds on and further constrains the CDA-CRS implementation guide.

5.2.1. Possible Categories of Medical Summaries

Medical Summaries are clinical documents that contain the most relevant portions of EHRs. Medical summaries are needed for patient transfers. Patients may be transferred for different reasons and, therefore, the summary documents that accompany these transfers can be categorized into three primary types which differ significantly:

- Episodic summaries have the primary purpose of highlighting the most relevant details of focused periods of time in a patient history, for example, the discharge of a patient from hospital to home. IHE has defined the "Acute Care Discharge to Ambulatory Environment" use case for episodic summaries.
- Collaborative summaries have a focused objective for providing the most relevant information about the patient intended for a specific provider, for example, the referral of a patient from a Primary Care Provider (PCP) to a specialist. IHE has defined the "Ambulatory Specialist Referral" use case for collaborative summaries.
- Permanent summaries have yet a third purpose of summarizing the entirety of a patient's medical history and therefore covers a broader range of patient problems. IHE deferred Permanent care summaries use case as future work.

6 THE STANDARDS OF CLINICAL RESEARCH DATA

6.1 Clinical Data Interchange Standards Consortium (CDISC)

CDISC is a not-for-profit organization founded in 1997 having the mission of developing global, platform-independent data standards that enable information system interoperability to improve medical research and related areas of healthcare [32]. 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 [87]: 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.

Five of these working groups have developed the current production standards:

- The Study Data Tabulation Model (SDTM) for the regulatory submission of Case Report Tabulations. This model now also includes the Standard for the Exchange of Non-clinical Data.
- The Analysis Data Model for the regulatory submission of analysis datasets.
- The Operational Data Model for the transfer of case report form data.
- The Laboratory Model for the transfer of clinical laboratory data.
- The Biomedical Integrated Research Domain Group (BRIDG) model.
- The Case Report Tabulation – Data Definition Specification (CRT-DDS, define.xml)
- The Terminology standard containing terminology that supports all CDISC standards.
- The Glossary standard providing common meanings for terms used within clinical research.

Two standards now under development are:

- The Trial Design model and other medical research protocol standards for the development of machine readable protocol standards
- The Clinical Data Acquisition Standards Harmonization (CDASH) for the development of data acquisition standards

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

The future developments of CDISC [88] include:

- Harmonization of the CDISC standards with the BRIDG domain model. The stated schedule for the harmonization of the standards is to have the SDTM harmonized by the end of 2008, and the ADaM in early 2009. The Protocol standard will be harmonized at the time of its release in 2008 and the LAB standard is already harmonized.

- Continuation of standards development with CDASH version 1.0 scheduled for production release in 2008, Protocol version 1.0 including Statistical Analysis Plans in 2008, extension of the LAB standard to include pharmacogenomics in 2009, improvements to the SDTM to better support the FDA submission process in 2008, modify the SDTM content standard so as to enable the FDA submission content delivered using HL7 transport messages, continued development of the ADaM standard, improvements to the ODM standard to better support industry users, updates to the CRT-DDS standard to align with the ODM and expand support to SDTM and ADaM metadata.
- Once the harmonization of the CDISC standards with the BRIDG model is complete, separate the content from transport standards to provide flexibility in the choice of transport formats (including HL7 formats) for common content to protect users from changes in underlying messaging technologies
- Provide an environment, the CDISC Operational Road Map Environment (CORE), to allow for testing of the standards being developed
- Provide tools that support users in implementing the standards and verifying correct usage
- Execute pilot projects (mostly with the FDA) to gain a better understanding of the needs of regulators and industry
- Link to healthcare by continuing efforts such as those conducted via the collaboration with IHE and BRIDG. Provide a road map by the end of 2008 detailing CDISC's aims in this area for the next 3 to 5 years.

6.2 HL7 Regulated Clinical Research Information Management Technical Committee (RCRIM TC)

The RCRIM TC [89] is chartered with the mission of defining messages, document structures, and terminology to support the systems and processes used in the collection, storage, distribution, integration and analysis of data created during research and regulatory evaluation of the safety and efficacy of therapeutic products or procedures.

Specific domains of interest to the RCRIM TC are: research processes of protocols and data collection, the organization and content of research data, product surveillance information gathered via the US Food and Drug Administration's (FDA) MedWatch reporting forms, regulatory documents including chemistry and manufacturing information and the medication reference terminology used in the coding of product labeling.

The work of the committee is intended to facilitate the availability of safe and effective therapies by improving the processes of regulated clinical research through its development of normative clinical research standards used by government agencies and private and public research groups.

In addition, RCRIM TC maintains a chartered agreement with CDISC, referred to as the CDISC – HL7 Project, having the ultimate goal of linking clinical research data with patient data stored EHR systems, provided, of course, that those EHRs would be using CDISC and HL7 standards. This project is proceeding by stages, the first was an exploratory phase concluded on November 27, 2007 with the conclusion of moving forward with requirements gathering, gap analysis and BRIDG harmonization (see section 6.3) as well as domain modeling and message development.

The messages to be developed (to satisfy the requirements of an FDA use case) in the current phase of the project are: Study Design, Study Participation, Subject Data and Individual Case Safety Report (ICSR). Once developed, the messages will permit informational content expressed in CDISC terminologies to be exchanged using the HL7 message format. The message, the content from CDISC to be contained within it, and the status of the needed harmonization of the CDISC content with the BRIDG model [90] is listed in the table below.

Message Name	CDISC Content	Harmonization Status
Study Design	a) Study Summary (Clinical Trial Registry	a) Harmonization scheduled to be complete by end of 2008
	b) Eligibility Criteria	b) Standards development needed
	c) Trial Design	c) Harmonization scheduled to be complete by end of 2008
	d) Statistical Analysis Plan	d) Standards development needed
Study Participation	e) Collected data/Study data tabulation	e) Harmonization scheduled to be complete by end of 2008
Subject Data	e) Collected data/ Study data tabulation	e) Harmonization scheduled to be complete by end of 2008
	f) Derived data/ Analysis datasets	f) Harmonization scheduled to be complete by early 2009
ICSR	e) Collected data/ Study data tabulation	e) Harmonization schedule to be complete by end of 2008

As of May 28, 2008 a model had been proposed to the RCRIM TC for the Study Design Message.

6.3 Biomedical Research Integrated Domain Group (BRIDG)

The BRIDG project [91] is a collaborative effort of the Clinical Data Interchange Standards Consortium (CDISC) [32], the HL7 Regulated Clinical Research Information Management Technical Committee (RCRIM TC) [89], the National Cancer Institute (NCI) [92], and the US Food and Drug Administration (FDA) [93]. It was formed in 2004 as the result of the cancer Biomedical Informatics Grid (caBIG – see section 6.4) project 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 the dynamic and static semantics of a particular domain-of-interest. In the case of the BRIDG model, the domain is defined as:

Protocol-driven research and its associated regulatory artifacts, i.e. the data, organization, resources, rules, and processes involved in the formal assessment of the utility, impact, or other pharmacological, physiological, or psychological effects of a drug, procedure, process, or device on a human, animal, or other biologic subject or substance plus all associated regulatory artifacts required for or derived from this effort.

Release 1.0 of the BRIDG model was published in June 2007. The scope of the release included the static models (UML Class Diagrams) of the following five projects: Study Table Tabulation Model (CDISC), caXchange/LabHub (NCI/RCRIM TC/CDISC), Regulated Product Submission (RCRIM TC/FDA), Cancer Clinical Trials Object Model (NCI) and the Patient Study Calendar (NCI).

The BRIDG model has been adopted by HL7 as the domain analysis model to be utilized by the RCRIM TC. CDISC has committed to harmonizing their 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 future efforts of the BRIDG project include:

- Provide educational materials that introduce the project to domain experts and that describe the modeling and harmonization practices for use by technical experts
- Comparison of the BRIDG model to other clinical trial models to improve quality and locate opportunities for collaboration
- Creation of tools to support model development, validation and maintenance

The BRIDG model, unfortunately, suffers from a number of inconsistencies [94].

6.4 The Nationwide Health Information Network (NHIN)

The NHIN's mission is to provide a secure, nationwide, interoperable health information infrastructure that will connect providers, consumers, and others involved in supporting health and healthcare [95]. The NHIN intends to enable health information to follow the consumer, be available for clinical decision making, and support appropriate use of healthcare information beyond direct patient care so as to improve health.

In November 2005, the Office of the National Coordinator for Health IT (ONC) awarded four contracts totaling $18.6 million 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 the architecture [96].

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 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 create a "network of networks," providing the interconnection between existing stakeholder networks so 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 is to be performed by HITSP.

We focus here on the architecture developed by CSC as it utilized the components of 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, personal health records (PHRs), and research data.

6.4.1. National Multi-Protocol Ensemble for Self-scaling Systems for Health (NMESH)

The PHR used in NMESH is that developed through the NLM sponsored Personally Internetworked Notary and Guardian (PING) project which has now evolved into the Indivo project of the Children's Hospital Informatics Program (CHIP), a joint collaboration of Harvard Medical and MIT [97]. On September 17, 2007, Dossia, a consortium of large employers including: AT&T, BP America, Cardinal Health, Wal-Mart, Intel Corporation, and Applied Material, announced that it had selected Indivo as the platform it would use to support the PHR of the 5 million employees, retirees, and dependents that it represents.

The monitoring and data analysis component of the NMESH project is the Automated Epidemiological Geotemporal Integrated Surveillance (Aegis) system also developed by CHIP. It provides the ability for large scale population monitoring and for at-a-glance understanding of distribution of disease burdens at monitored areas. The Aegis system is comprised of four components:

- Source manager – handles the reading and processing of data from the various network sites,
- Prediction manager – generates expected values for the observed values in the Aegis database,
- Alarm manager – interprets the observed values in light of the expected values and generates an appropriate alarm, and
- Client manager – interacts and communicates with users through various clients.

Files from source systems are either records of emergency department visits containing chief complaints or for registering ICD-9s. The files sent from source systems to the Aegis system are comma separated text files whose rows contain the following fields:

- Hospital, Medical Record, Account Number, Chief Complaint, ICD9, ED Arrival Date, ED Arrival Time, ED Discharge Date, ED Discharge time, Admit Date, Admit Time, Birthday, Address, City, State/Province, Zip/Postal Code, Country, Gender, and Disposition.

The Aegis system is currently in use as the syndromic surveillance system for the Massachusetts Department of Public Health.

The Shared Pathology Informatics Network (SPIN) component of the NMESH project is a peer-to-peer network, originally a project funded by the NCI, allowing participating institutions to expose de-identified pathology reports to other nodes on the network [98]. This enables sharing of data contained in pathology reports and for retrieving the pathways for the associated pathologic specimens stored as paraffin blocks.

The Indiana Network for Patients Care (INPC) is a node on the SPIN network. Its adaptation of a specialized query tool is illustrative of both the potential of large networked databases and of the ongoing difficulties of linking local systems.

The INPC shares clinical data from eleven hospitals in five competing health systems, the county department of health, a large primary care network and a homeless care network. The organizations send data via HL7 v2 messages and are stored in a centralized database with consistent structure and codes. Collectively, the participating hospitals admit 180,000 patients, and serve almost 400,000 emergency room visits and more than 4 million clinic visits annually.

All INPC members currently exchange ED and outpatient visits, inpatient lab results, hospitalization discharge summaries, radiology reports, tumor registry data, anatomic pathology reports, and immunizations. Some institutions are exchanging ambulatory notes, vital signs, visit reasons and diagnoses, medication profile, cardiac testing, radiology images, and gastroenterology reports.

The specialized query tool provides the INPC users with the ability to define cohorts, data sets and set a statistical analysis plan. Cohorts are determined by the user selecting any of 415 variables including those from tumor registries, lab tests, and pathology reports. Since observations are repeated, the user is prompted for each variable selected to filter by the earliest, most recent, maximum, minimum occurrence or to aggregate the repeated values into an average or count.

The definition of the data set, the variables to be retrieved for the defined cohort, proceeds similarly to the definition of the cohort itself. For each variable selected, the user is prompted to select the instance of the variable to be retrieved (first, last, max, min, etc.) as well as the component of the variable, the value, the date, etc.

The specification of a statistical analysis defines the kind of statistical analyses to be run, and the variables to be included in it. Users can choose from five kinds of analyses: a statistical summary, dynamic cross tabulation, logistic regression, simple regression, and survival analysis. For the dynamic cross tabulation, the breakdown variables and their cut points need to be specified. For the other three analyses, users need to enter information that defines the relationship between the independent and dependent variables.

The observation and report identifying codes come in the messages as local idiosyncratic codes. For the needed interoperability these local codes are mapped to a standard code from LOINC. The mapping is a manual process and the time needed to complete one is reported as ranging from a few person days for some EKG systems to six to twelve person *months* for a laboratory with 2000-4000 distinct test observations.

Other issues with the mapping process reported in the CSC Overview are missing data needed for the mapping, laboratories often have difficulty providing details of methods used,

ambiguities in data such as quantitative tests results being reported as “negative” if the result is under specific threshold, but is reported as an integer if the result is over the threshold, the LOINC standard was not mature and had required a considerable expansion of the codes it contained.

6.5 Cancer Biomedical Informatics Grid (caBIG)

caBIG, 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.” [99] 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 [100, 101] 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, and
- SNOMED.

The syntactic component of interoperability is addressed by the caDSR, a metadata repository and registry which holds the potential 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 consisting of two parts: a Data Element Concept – the conceptual definition of the data element and a Value Domain – a description 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.

As an example, the data element concept (DEC) of a data element representing a person’s eye color is composed of two parts: the object class or domain of the data element which in this case would be person, and the property which in this case would be eye color. The value domain (VD) would be color (specified by list or definition). The component-based structure of data elements allows for reuse of the object class, property and value domain to form new data elements as would be the case if the object class of person was combined with the property of hair color to form the new data element concept of person hair color which was then associated

with the value domain of color. The object class, property, and value domain can all be terms from one or more EVS vocabularies.

The capability to integrate external applications to the caCORE components is provided by the caBIO, a set of JavaBeans with open source APIs that can be used to directly access the caDSR. With this infrastructure, federated databases having their metadata registered in caDSR are linked either by the common data elements they share, by the common terms from EVS vocabularies used in distinct data elements, or by the mappings between terms from different EVS vocabularies used in distinct data elements.

The caCORE has an associated toolset including:

- CDE Browser – allows for search and comparison of existing data elements,
- Form Builder – allows for creation and sharing of forms based on existing data elements for use within a user community,
- CDE Curation Tool (requires curatorial authority) – allows for the creation or edit of data element concepts, value domains, and data elements,
- Sentinel Tool – allows for the creation of alerts triggered by changes that have been made within the caDSR,
- Admin Tool (requires appropriate privileges) – allows for the creation of conceptual domains, classification schemas and protocols.

The caBIG project and the caCORE infrastructure is a promising technology in the advancement of interoperability in HIT. However, what is missing from the caBIG 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 many mistakes in the NCI Thesaurus [102].

7 STANDARDIZATION ACTIVITIES ON TERMINOLOGIES AND CODING SCHEMES

7.1 Classifications, nomenclatures and thesauri

7.1.1. Classification

A classification provides a set of classes to arrange individuals, mainly for statistical purposes. Therefore classes are typically mutually exclusive.

Some classifications (e.g. ICD) also provide a taxonomy among classes, to facilitate the clustering of classes for synthetic statistical tables. The main goal of a classification (i.e. the assignment of individuals to classes) should not be confused with the ancillary artifacts (as the clustering of classes into broader classes).

7.1.1.1 An example: Current Procedural Terminology

An example is the Current Procedural Terminology (CPT) published by the American Medical Association (AMA) [103]. The AMA developed the CPT in 1966 to provide a pre-coordinated coding scheme for diagnostic and therapeutic procedures that has since been adopted in the United States for billing and reimbursement. The second edition, published in 1970, contained 5-digit CPT codes, replacing 4-digit codes, for diagnostic and therapeutic procedures in surgery, medicine, and the specialties. The third and fourth editions were released in the 70s. In 1983, the then Health Care Financing Administration (HCFA), now the Centers for Medicare and Medicaid Services, included CPT as part of the HCFA Common Procedure Coding System (HCPCS). In the mid-80s HCFA mandated CPT codes for reporting outpatient hospital surgical procedures and required State Medicaid agencies to use HCPCS.

The maintenance of the CPT code set is the responsibility of the CPT Editorial Panel which is authorized to revise, update, or modify the CPT codes. The Panel is comprised of 17 members. The AMA's board of trustees appoints the Panel members. Five members of the Editorial Panel serve as the panel's Executive Committee. Supporting the CPT Editorial Panel in its work is a larger body of CPT advisors, the CPT Advisory Committee. The members of this committee are primarily physicians nominated by the national medical specialty societies represented in the AMA House of Delegates. Currently, the Advisory Committee is limited to national medical specialty societies seated in the AMA House of Delegates and to the AMA Health Care Professionals Advisory Committee (HCPAC), organizations representing limited-license practitioners and other allied health professionals. Additionally, a group of individuals, the Performance Measures Advisory Committee (PMAC), who represent various organizations concerned with performance measures, also provide expertise.

There are three categories of CPT codes:

- Category I CPT codes are designated for services (or procedures) common in contemporary medical practice and being performed by many physicians in clinical practice in multiple locations. For each, there is a five digit code and a text descriptor.
- Category II CPT codes are focused on performance measurement. They are invented to facilitate data collection by coding certain services and/or test results that are agreed upon

as contributing to positive health outcomes and quality patient care. This category of codes is a set of optional tracking codes for performance measurement. These codes may be services that are typically included in an Evaluation and Management (E/M) service or other component part of a service and are not appropriate for Category I CPT codes. The use of tracking codes for performance measures will decrease the need for record abstraction and chart review, thus minimizing administrative burdens on physicians and survey costs for health plans.

- Category III CPT codes deal with emerging technology. The purpose of this category is to facilitate data collection on and assessment of new services and procedures. These codes are intended to be used for data collection purposes to substantiate widespread usage or in the FDA approval process.

CPT codes specify information about the codes which differentiates them based on their cost as illustrated by the following codes, cited by Cimino et al, for pacemaker insertions:

- 33200 - epicardial, by thoracotomy,
- 33201 - epicardial, by xiphoid approach,
- 33206 - transvenous, atrial,
- 33207 - transvenous, ventricular,
- 33208 - transvenous, AV sequential

CPT codes also describe information about the reasons for a procedure as illustrated by the following codes, cited by Cimino et al, for arterial punctures:

- 36600 - withdrawal of blood for diagnosis,
- 36620 - monitoring,
- 36640 - infusion therapy,
- 75894 - occlusion therapy

In 2000 the AMA initiated the CPT-5 project whose goal was to make improvements in the structure of the CPT codes with the ultimate aim of achieving interoperability with other terminology systems. Some of the stated goals of the project included: greater specificity of terms, improved hierarchy of terms, standardization of language, and structured definitions.

The main features of classifications include:

- Uniqueness of entity-to-code relationship (an event or object such as a patient corresponds to one and exactly one code). Minor exceptions have been verified in some cases where the same classification is required for different purposes, and an additional code may be requested in order to specify details.
- Classes are mutually exclusive. Rare exceptions have been verified when different "clinical schools" conflict
- Propensity for comprehensive expressions including several clinical concepts (designers, not users, decide on the amount of details to be preserved and represented)

- Coverage of a specific clinical field. Coverage usually comprehends only diseases and procedures
- Existence of a wide variety of metalanguage expressions. e.g. “NOS”, “NEC” , “other”, “with mention of”
- Conceived for off-line coding by professional coders; information is considered already available in a document previously written — without constraints — by clinicians

7.1.2. Nomenclature

A nomenclature provides a list of expressions with the goal of capturing in a systematic and reproducible way a set of details.

7.1.2.1 An example: LOINC

An example is LOINC [104], the Laboratory Logical Observation Identifiers Names and Codes database. LOINC is published by Regenstrief Institute and is an adopted standard for laboratory procedures and structured labeling of medications. The original release of the nomenclature was in the spring of 1995.

Initially, LOINC was designed to supply universal identifiers for observations in HL7 messages from laboratories to health care providers and maintenance organizations. These messages have one field, called OBX-3 which carries the observation, or the test performed and another, called OBX-5 which carries the value of the test. Thus LOINC supplies the identifier for the test and other terminologies such as ICD-9, SNOMED, and MEDCIN supply the test value.

According to its developers, LOINC follows the good coding system practices of having no embedded meaning within codes and never reusing or deleting a code. Rather, codes which are no longer to be used are marked as being deprecated but not removed. The structure of LOINC, like the ICD, is multi-axial, meaning that each character of the 6 character LOINC code signifies the codes placement within one of 6 axes. The LOINC major axes are listed below:

Character Place	Description
1	The substance to be measured (analyte)
2	The measured property of the substance e.g. mass concentration
3	Timing, i.e. whether the measure is instantaneous or an average over time
4	System, i.e. type of sample or organ on which observation was made
5	Scale, i.e. value type output by measurement: numerical, ordinal, narrative
6	Method used to produce observation

The scope of the laboratory portion of LOINC includes all observations reported by clinical laboratories, including the specialty areas: chemistry, including therapeutic drug monitoring and toxicology; hematology; serology; blood bank; microbiology; cytology; surgical pathology; and fertility. As of June 2006, there was a draft effort of including codes for genomic tests in LOINC. The review of the effort at that time concluded that changes would probably need to be

made to the then current LOINC conventions for genomic tests and a need was found to exist for a standard approach for reporting genomic sequences based on differences between a standard reference sequence and the patient's sequence. The project would also, it was determined, require the creation of new LOINC codes.

The initial intent of laboratory test coverage was extended to include clinical sub-domains. The clinical LOINC division is concerned with coverage of non-laboratory diagnostic studies, critical care, nursing measures, patient history, physical and survey instruments. The subjects covered to date in clinical LOINC include:

- Blood pressure (systolic, diastolic, and mean)
- Body height
- Body temperature
- Body weight (and measures used to estimate ideal body weight)
- Cardiac ultrasound (echo) imaging
- Cardiac output, resistance, stroke work, ejection fraction, etc.
- Circumference of chest, thigh, legs
- Critical care measures
- Dental
- Electrocardiographic measures
- Emergency department case reports – CDC DEEDS
- Gastrointestinal endoscopy
- Heart rate (and character of the pulse wave)
- Intake and output
- Major headings in operative note
- Major headings of discharge summary
- Major headings of history and physical
- Obstetric ultrasound imaging
- Ophthalmology measurements
- Pathology protocols
- Pulmonary ventilator management
- Radiology reports
- Respiratory rate
- Standardized survey instruments
- Urology ultrasound imaging

The LOINC codes have been met with significant uptake by the health care industry. The American Clinical Laboratory Association (ACLA), an association of large referral laboratories whose members are responsible for more than 60% of US outpatient laboratory test volume, has recommended LOINC for adoption by its members. Quest Diagnostics and LabCorp have adopted LOINC as their code system for reportable test results. In addition, Intermountain Health Care, Kaiser Permanante, Clarian Health, Partners Healthcare System of Boston, Care Group of Boston, Mayo Medical Group, and the Department of Defense are adopting the LOINC codes for laboratory reporting. All US veterinary medicine laboratories have committed to the use of LOINC.

The main features of nomenclatures include:

- Propensity to consider any expression as "unitary" concept, even very long expressions
- Uniqueness of concept-to-code relationship; each preferred term corresponds to a code
- Complex situations expressed by multiple codes; normally a complex concept is represented by a user-defined combination of codes (the user selects the amount of details he wants to preserve)
- Diverse solutions are available for complex expressions ; most Diseases and Procedures have explicit pointers to “elementary axes” (unique code vs multiple codes); different ad hoc arrangements are normally possible for multiple coding
- Extensive coverage of medical terminology; SNOMED has 12 modules; one module is about "modifiers" or "qualifiers"
- Conceived for direct coding by users and case-based processing

7.1.3. Thesaurus

A thesaurus is a system of predefined descriptors, usually designed for indexing and retrieval purposes.

7.1.3.1 An example: the NCI Thesaurus

An example is the NCI Thesaurus (NCIt) [105]. The NCIt 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 NCIt 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 [106].

The NCIt serves several functions, including annotation of the data in the NCI’s repositories and search and retrieval operations applied to these repositories. It is also linked to other information

resources, including both internal NCI systems such as caCore and MGED and also external systems such as the Gene Ontology and SNOMED-CT. It is part of the Open Biomedical Ontologies library (obo website) and is also available under Open Source License on the NCI download area (NCI download area). This makes it an important candidate for the delivery of vocabulary services in cancer-related biomedical informatics applications in the future.

Version 08.04d of the NCIt was released in April of 2008. The content of the NCIt is comprised of concepts from a total of nineteen domains or kinds:

Domain Name	**Domain Description**
Gene	Any definable DNA sequence capable of being transcribed and having biological significance.
Gene_Products	Endogenous RNAs, proteins, protein complexes and riboprotein complexes. Excludes exogenous chemicals.
NCI	Conceptual entities required by NCI operations and systems. Includes administrative, financial, organizational and quasiscientific concepts .
Findings and Disorders	Classification of human conditions that are relevant to cancer. Includes observations, test results, history and other concepts relevant to characterization of human cancer-related conditions. Includes non-neoplastic conditions of special interest
Anatomy	Naturally occurring human biological structures, fluids, and substances. Includes embryonic, gross and micro anatomic structures and surgically created structures, including cellular organelles but excluding single molecules.
EO_Anatomy	Naturally occurring non-human biological structures. Includes embryological, gross and micro anatomic structures in all species used as models of human cancer. Excludes structures smaller than can be visualized by light microscopy.
EO_Findings_and_Disorders	Classification of non-human conditions that are relevant to cancer. Includes observations, test results, history and other concepts relevant to characterization of cancer-related conditions in species used as models of human cancer.
Abnormal_Cell	An enumeration of abnormal cell types that occur in human cells and in cells of experimental models of human cancer.
Molecular_Abnormality	An enumeration of the molecular abnormalities that occur in human cells and tissues and non-human models of human cancer. Includes abnormalities such as translocation, polymorphism, under expression, over expression.
Biological_Process	Events occurring within an organism, between organisms or among organisms and mechanisms underlying such events.
Chemicals_and_Drugs	Organic and inorganic substances, elements, and isotopes used in research or for the prevention, diagnosis, and/or treatment of disease states. Includes biologically active substances that are either synthetically manufactured or endogenous substances extracted and processed to be reintroduced into an organism.
Diagnostic_and_Prognostic_Factors	Characteristics of the organism or of a process that contribute to clinical diagnosis, treatment selection or prediction of clinical outcome.

Technique	Scientific or clinical procedures and methods, including tests
Properties_Or_ Attributes	Modifiers and qualifiers. Any concept in the form of an adjective, adverb, passive voice verb.
Combination_Chemotherapy	Combinations of multiple drugs used in standard and clinical trial treatments. They do not currently specify order, dosage or dosing interval of the individual ingredients.
Reference_Kinds	Concepts incorporated in toto from an external ontology for the purpose of providing range values for NCI concepts and to provide a defined linkage between such NCI concepts and the external ontology.
Equipment	Supplies or apparatus used for cancer-related research, diagnosis or therapy.
Organism	A living entity.
Pathway	A reaction or series of reactions. Includes both anabolic and catabolic reactions, and functional outputs such as transcription, activation, promotion and so forth.

Concepts within the same kind are differentiated through the application of "roles" that are relationships between kinds. As Coronado et al. [106] illustrates, a disease is differentiated using roles such as "Disease_has_primary_anatomic_site", "Disease_has_normal_tissue_origin", "Disease_has_normal_cell_origin" which relates diseases to concepts within the anatomy domain and by using roles such as "Disease_has_abnormal_cell", "Disease_has_cytogenetic_abnormality", and "Disease_has_molecular_abnormality" to other concepts in the findings_and_disorders domain.

The main features of thesauri include:

- Conceived for two kinds of users: the professionals performing the indexing of each new documenting unit, and the clinicians who want to retrieve the documenting units in a large collection;
- Descriptors represent elementary concepts to be used in combination among them
- Typically organized by "facets"; large chapters containing tree structures, often using also associative relations;
- Normally a complex concept should be represented by an ad-hoc defined combination of descriptors (the user selects the amount of details he wants to preserve)
- Occurrence of pragmatic exceptions of different nature and origin
- Extensive coverage of medical field and outside medical field (e.g. Countries)
- Presence of "position codes". In most cases a single concept can be found under different positions, with specific codes (called "contexts" in UMLS/META-1)

7.1.3.2 Problems with the concept-based approach in building thesauri

In analyzing the NCI Thesaurus, Ceusters et al. were particularly interested in how the claimed ontological features of the system work together with its terminological parts [102]. They found that the system suffers from the same problems encountered in so many of the biomedical

terminologies produced in recent years. The NCIT is probably 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, as well as 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. The NCI Thesaurus is a never-ending work in progress, the content of which 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 NCIT to be tested not just for internal consistency but also for consistency with that part of reality which it is intended to represent.

7.2 Unified Medical Language System (UMLS)

The NLM [107] initiated the UMLS research project in 1986 [108, 109] with the goal of overcoming the barriers to health information technology created by the differences of language used in different information sources to refer to the same entity (e.g. 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 and 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. Version 2008AA of the Metathesaurus released in April of 2008 contains 5.7 million distinct normalized names for 1.5 million concepts from 123 families of vocabularies.

The contents of the source vocabularies are linked by meaning within the Metathesaurus. Synonymous, but syntactically different terms from different source vocabularies are collected under a single concept which is then associated with a preferred term and assigned a unique identifier. In addition to synonymy, the Metathesaurus contains other relationships between concepts from the same source vocabulary (intra-source relationships) and well as between concepts from different source vocabularies (inter-source relationships).

Intra-source relationships are generally those from source vocabulary's hierarchical contexts, cross-reference structures, rules for applying qualifiers, or connections between different types of names for the same concept such as between abbreviations and full forms. Some, such as sibling relationships, are to facilitate the construction of user displays. Some are statistical, computed by determining the frequency with which concepts co-occur in databases, such as those storing patient discharge data or by determining the frequency with which concepts co-occur as key concepts within the same article.

The primary inter-source relationships are the relationships between synonymous terms added by the editors of the Metathesaurus but others are added by users who discover "like" or "similar"

relationships. Still others are created during the creation of mappings between source vocabularies.

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 and a set of 54 Semantic Relationships between these Types. Semantic Types are broadly 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.

7.3 SNOMED-CT

7.3.1. Overview

SNOMED-CT® [15] was developed by the College of American Pathologists (CAP) [110], and grew out of the merger, expansion, and restructuring of SNOMED RT® (Reference Terminology) [111] and the United Kingdom National Health Service Clinical Terms (also known as the Read Codes) [112]. CAP and the NHS Information Authority have been collaborating on the development of SNOMED-CT since April of 1999. Alpha testing started in 2001 [113].

July 1, 2003, the CAP has signed a US$32.4 million, five-year sole source contract with the National Library of Medicine (NLM), which is part of the National Institutes of Health within the US Department of Health and Human Services (HHS), to license English and Spanish language editions of SNOMED-CT [114]. The agreement provides free-of-charge access to SNOMED-CT core content and all version updates, starting in January 2004. Access is 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).

SNOMED-CT is based on "*concepts*." Each concept represents a unit of thought or meaning and is labelled 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 that have logical connections to it. Relationships are used to provide a computer readable description, and sometimes a definition of the concepts. These connections allow SNOMED-CT to be searched, retrieved, reused or analysed 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 organised into hierarchies, with approximately 1.36 million relationships between them, and more than 1,068,278 terms. SNOMED-CT is officially available in English and Spanish language editions.

7.3.2. Positive views on SNOMED-CT

The main merits of SNOMED-CT for clinical documentation are its broad terminological coverage which has been shown repeatedly in the course of its development and in various areas.

Already in 1997, SNOMED (at that time SNOMED International) and the Read Codes proved to have the best term and concept coverage (more than 60% of the terms and more than 50% of the concepts) in comparison to other systems [115]. In that study, a distributed US-national experiment using the Internet and the UMLS Knowledge Sources, lexical programs, and server was carried out to determine the extent to which a combination of existing machine-readable health terminologies cover the concepts and terms needed for a comprehensive controlled vocabulary for health information systems. Using a specially designed Web-based interface to the UMLS Knowledge Source Server, participants searched the more than 30 vocabularies in the 1996 UMLS Metathesaurus and three planned additions to determine if concepts for which they desired controlled terminology were present or absent. For each term submitted, the interface presented a candidate exact match or a set of potential approximate matches from which the participant selected the most closely related concept. The interface captured a profile of the terms submitted by the participant and for each term searched, information about the concept (if any) selected by the participant. The term information was loaded into a database at NLM for review and analysis and was also available to be downloaded by the participant. A team of subject experts reviewed records to identify matches missed by participants and to correct any obvious errors in relationships. The editors of SNOMED International and the Read Codes were given a random sample of reviewed terms for which exact meaning matches were not found to identify exact matches that were missed or any valid combinations of concepts that were synonymous to input terms. The 1997 UMLS Metathesaurus was used in the semantic type and vocabulary source analysis because it included most of the three planned additions. Sixty-three participants submitted a total of 41,127 terms, which represented 32,679 normalized strings. More than 80% of the terms submitted were wanted for parts of the patient record related to the patient's condition. Following review, 58% of all submitted terms had exact meaning matches in the controlled vocabularies in the test, 41% had related concepts, and 1% were not found. Of the 28% of the terms which were narrower in meaning than a concept in the controlled vocabularies, 86% shared lexical items with the broader concept, but had additional modification. The percentage of exact meanings matches varied by specialty from 45% to 71%. Twenty-nine different vocabularies contained meanings for some of the 23,837 terms (a maximum of 12,707 discrete concepts) with exact meaning matches. Based on preliminary data and analysis, individual vocabularies contained < 1% to 63% of the terms and < 1% to 54% of the concepts. Only SNOMED International and the Read Codes scored higher.

In 2002, the authors of [116] performed a simple semi-quantitative evaluation of ICD-10, CDAM, MedDRA, MeSH, READ, SNOMED and UMLS to provide objective criteria for the choice of a coding system for the computer representation of clinical trials in the context of evidence-based decision support and for the integration of the messages produced by these activities with clinical information and electronic patient record systems. Criteria included coding coverage, size, integration and language coverage. The results of the comparison lead them to choose SNOMED as the most appropriate coding system for their needs.

In developing a customized enterprise-wide vocabulary for clinical terminology in 2003, the authors of [117] implemented SNOMED CT as a base vocabulary in their system, while

facilitating the addition of site-specific clinical terms or concepts not represented in SNOMED CT. They evaluated the breadth of SNOMED CT terms and concepts for the coding of diagnosis and problem lists by clinicians within a computerized physician order entry (CPOE) system. Clinicians selected diagnosis and problem list terms from a lexicon based on SNOMED CT, submitting requests for clinical terms that were not found in the controlled vocabulary. For each "missing" term, they assigned one of four mapping types, representing the relationship of this new terminology entry to the SNOMED CT reference terminology. Their results show that the majority of diagnosis/problem list terms (88.4%) were found in SNOMED CT. Of the 145 missing terms, only 20 represented significant concepts missing from SNOMED CT, resulting in concept coverage of 98.5%.

In the same year, the ability of SNOMED-CT to represent simple and compositional concepts in FDA approved oncology drug indications was assessed [118]. Oncology drug indications were decomposed into single and compositional concepts. SNOMED-CT's coverage of single concepts and the semantics needed to create compositional concepts were evaluated using automated and manual techniques. SNOMED-CT covered 86.3% of single concepts present in oncology drug indications; 11.3% of indications were covered completely. Coverage was best for concepts describing diseases, anatomy, and patient characteristics. Medications accounted for 50.5% of missing concepts. Excluding drug names, 45.2% of indications were completely represented. SNOMED-CT's semantics completely represented 60.1% of compositional expressions. The authors concluded that SNOMED-CT's overall coverage of the concepts in oncology drug indications was good but that improvements or alternatives are needed for medications and semantics.

In 2004, the Veterans Health Administration (VHA) reported similar results for the first phase of their evaluation of the use of SNOMED-CT, which examined the coverage of SNOMED-CT for problem list entries [119]. Clinician expressions in the VHA problem lists are quite diverse compared to the content of the current VHA terminology Lexicon. They selected a random set of 5054 narratives that were previously "unresolved" against the Lexicon. These narratives were mapped to SNOMED-CT using two automated tools. Experts reviewed a subset of the tools' matched, partly matched, and un-matched narratives. The automated tools produced exact or partial matches for over 90% of the 5054 unresolved narratives. They concluded that SNOMED-CT has promise as a coding system for clinical problems and planned to perform subsequent studies on the coverage of SNOMED for other clinical domains, such as drugs, allergies, and physician orders.

In 2005, [120] reported on the adequacy of 5 controlled medical terminologies (International Classification of Diseases 9, Clinical Modification (ICD9-CM); Current Procedural Terminology 4 (CPT-4); Systematized Nomenclature of Medicine, Clinical Terms (SNOMED-CT); Logical Identifiers, Names, and Codes (LOINC); Medical Entities Dictionary (MED)) for representing concepts in ophthalmology. Twenty complete ophthalmology case presentations were sequentially selected from a publicly available ophthalmology journal. Each of these cases was parsed into discrete concepts, and each concept was classified along 2 axes: (1) diagnosis, finding, or procedure and (2) ophthalmic or medical concept. Electronic or paper browsers were used to assign a code for every concept in each of the 5 terminologies. Adequacy of assignment for each concept was scored on a 3-point scale. Findings from all 20 case presentations were combined and compared based on a coverage score, which was the average score for all concepts in that terminology. The cases resulted into 1603 concepts. SNOMED-CT had the highest mean

overall coverage score (1.625+/-0.667), followed by MED (0.974+/-0.764), LOINC (0.781+/-0.929), ICD9-CM (0.280+/-0.619), and CPT-4 (0.082+/-0.337). SNOMED-CT also had higher coverage scores than any of the other terminologies for concepts in the diagnosis, finding, and procedure categories. Average coverage scores for ophthalmic concepts were lower than those for medical concepts.

In 2006, Richesson et al. [121] investigated the coverage of clinical research concepts provided by SNOMED-CT. During the study, a total of 616 unique concepts were derived from a set of 17 case report forms used in conducting longitudinal, observational studies of vasculitis. The concepts were categorized as being either clinical findings or procedures and were classified by the presence and nature of SNOMED CT coverage. The results were that while 88% of the clinical items were covered by SNOMED CT only 23% were fully covered, meaning that all aspects of the concept could be represented without post-coordination. The types of post-coordination required were either to clarify the context of a data element or to more fully capture the content of complex concepts such as disease-related findings.

7.3.3. Negative views on SNOMED-CT

Despite these positive assessments of the performance of SNOMED-CT on tests of coverage, there are also negative assessments along primarily three lines: term formation principles, SNOMED-CT as an ontology, and practical usefulness.

With respect to content, [122] conclude that SNOMED CT is currently weak in the areas of specific recording for domestic violence, but as a consequence of prior development, supports those areas of interest that are neutral to this particular exercise (e.g. recording of background information physical and non-physical morbidity).

7.3.3.1 Linguistic problems with SNOMED

In [123], Ceusters et al. analysed the procedure axis of SNOMED International (1998) from the perspective of controlled language principles for the construction of controlled vocabularies, thereby identifying several sources of confusion and ambiguity, including:

- inappropriate use of synonymy (e.g. inconsistently used in preferred terms of "*ear drum*" and "*tympanic membrane*"),
- misleading use of homonymy ("*ventricle*" used both for "*cardiac ventricle*" and "*cerebral ventricle*"),
- incomprehensible concatenation of noun clusters, for example in the term: "*open treatment of craniofacial separation, Lefort III type with wiring and/or local fixation, complicated, fixation by head cap, halo device, multiple surgical approaches, internal fixation, and/or wiring of teeth"*
- attenuated or ambiguous dependency of modifiers (e.g. in: "*epiphyseal arrest by stapling, combined, proximal and distal tibia and fibula and distal femur*").

It was accordingly argued that term-formation in SNOMED could benefit from the use of a controlled language to make the meaning of terms clearer.

Bodenreider *et al.* used lexical techniques to study the (in)consistent use of modifiers such as "bilateral"/"unilateral", and "congenital"/"acquired" in SNOMED International [124]. Every

occurrence of “bilateral X” or “congenital X” would indeed call for a “unilateral X” and “acquired X” respectively, but this requirement was met in very few cases.

A more formal representation of SNOMED class descriptions was at that time not available. Such a formal representation did become available with SNOMED-RT (2000), which however at the same time opened up the possibilities for new types of mistakes. In [125], Campbell reports having found only 0.6% “editorial mistakes” in the portion of SNOMED-RT that he analysed. Whether this surprisingly low figure is accurate is hard to assess. The actual sample size is not given in the paper, but based on his report to the effect that 128 clinical statements from the University of Nebraska Medical Center Lexicon (UNMC Lexicon) were analysed, and that single statements translated into an average of 2.61 SNOMED-RT canonical representation triples, the absolute upper bound of SNOMED-RT statements verified must have been 334. Even given that the 128 statements were a representatively selected sample of 1% from the UNMC Lexicon, it is hard to defend the thesis that these 334 SNOMED triples – constituting less than 0.28% of RT as then constituted, were also representative of SNOMED as a whole. In contrast to Campbell’s positive statement of his results, 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*” [126].

7.3.3.2 Ontological problems in SNOMED-CT

In [127] 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 reading a term (e.g. via a browser) attach to it a meaning that is not intended by the system (which can be verified by analyzing in detail the formal statements through which the term is defined) [128, 129].

The analysis was carried out on the January and July 2003 versions of SNOMED-CT®. For purposes of further reference, the authors assigned an identifying label of the form “Ja-#”, “Ju-#”, or “Jau-#” to each reported mistake or inconsistency. These labels indicate the presence of the corresponding error in the January, July or in both versions of the system, respectively.

7.3.3.2.1 Human error

Some mistakes must have their origin in inattentiveness on the part of human beings during the manual phases of the process of creating and error-checking SNOMED-CT®. The following are some of the types of errors that were found under this heading.

Improper assignment of is-a relationships

The class “265047004: diagnostic endoscopic examination of mediastinum NOS” is subsumed by “309830003: mediastinoscope”. Thus a procedure is classified as an instrument (Jau-1). The former is marked as “limited” (meaning: “of limited clinical value”) as it is based on a classification concept or an administrative definition. Yet SNOMED-CT® still considers entries with this status as valid for current use and as active.

Another example has a procedure wrongly subsumed by a disease. Thus the class "275240008: Lichtenstien repair of inguinal hernia" is directly subsumed by "inguinal hernia" (Jau-2).

A specific subtype of this sort of mistake consists of the *improper treatment of the partial/complete distinction*. 9 classes were found whose terms included the qualifier "complete" yet were subsumed by altogether 17 classes qualified as "partial". 6 "partial" classes were, in the other direction, subsumed by 11 "complete" parents. As an example, "359940006: partial breech extraction" is subsumed by "177151002: breech extraction", which is in turn subsumed by "237311001: complete breech delivery" (Jau-4).

The reason for these mistakes turned out to be the assignment of a term of the form "complete X" to a SNOMED-CT® class with the preferred name "X", where "X" then also subsumes "partial X". Mistakes of this type can be detected only when external ontological information is used – in this case information to the effect that classes qualified as "partial X" are disjoint from classes qualified as "complete X".

Other subtypes of erroneous assignment of *is-a* relations can be classified under the heading: *improper treatment of negation.* Thus "203046000: Dupuytren's disease of palm, nodules with no contracture" is subsumed by "51370006: contracture of palmar fascia" (Jau-3).

Further evidence for mistakes along these lines come from Bodenreider *et al.*, who performed a quantitative analysis of SNOMED-CT®, assessing its conformance to a number of principles of good practice in classification [130]. The methodology applied was not suited to the finding of mistakes, but quite sensitive in detecting missing information. As an example, 51% of the assigned parent-child relationships were found to lack differentiating criteria, so that the semantic difference between child and parent remains for these cases unexplained. Furthermore, 31.5% of classes with children have only one child, which suggests for each such class that either at least one child term is missing (from which the available term would then be differentiated), or that there is no semantic difference between parent and child.

Improper assignment of non-is-a relationships

The class "51370006: contracture of palmar fascia" is linked by the *Finding Site* relationship to the class "64799002: plantar aponeurosis structure". Probably as a consequence of automated classification, the latter is wrongly subsumed by "disease of foot" (reflecting the fact that "plantar aponeurosis structure" is subsumed by "structure of foot") (Jau-5). A similar phenomenon is observed in relation to "314668006: wedge fracture of vertebra", which is subsumed by "308758008: collapse of lumbar vertebra" (Ja-6). Although this erroneous subsumption is no longer present in the July version, the wrong association via *Finding Site*: "bone structure of lumbar vertebra" has been retained (Jau-7). Equally the class "30459002: unilateral traumatic amputation of leg with complication" is classified as an "open wound of upper limb with complications" due to an erroneous association with *Finding Site*: "upper limb structure" (Jau-8).

Errors of this kind can be detected only by adding to SNOMED an external ontology such as BFO [131]. Their prevention is more difficult, since they are due simply to inattention on the part of the terminologists or ontologists working on the system.

7.3.3.2.2 Technology-induced mistakes

A first example of a mistake of this type has been referred to already above (Jau-5): wrong subsumption because of relationships inappropriately assigned. Other errors are probably induced by tools performing lexical or string matching. We can hardly imagine that a human being would allow "9305001: structure of labial vein" to be directly subsumed by both "vulval vein" and "structure of vein of head". The error probably comes from an unresolved disambiguation of the word "labia" that is used for both lip (of the mouth) and vulval labia (Jau-9). Error detection is possible for this sort of case only through the exploitation of an external ontology such as BFO and an associated external reference anatomy such as the FMA [132]. Error prevention would then require the terminology authoring system to enforce corresponding class disjointness at run-time.

7.3.3.2.3 Shifts in meaning from SNOMED-RT® to -CT®

The meanings of some SNOMED-CT® terms have changed with respect to the corresponding terms in SNOMED-RT© even where these terms have the same numerical identifier. Above all, the adoption of [133]'s idea of SEP-triplets (structure-entire-part) led to a large shift in the meanings of nearly all anatomical terms. One might argue that in RT anatomical terms such as "heart" were never supposed to mean "entire heart", but rather always: "heart or any part thereof"; in CT this distinction has been made explicit.

Many other terms appear also to have changed in meaning even though they have the same unique identifier in both RT and CT. A notable example is "45689001: femoral flebography" which in RT relates only to ultrasound but in CT involves in addition the use of a contrast medium (Jau-10).

There are also changes in the meaning of terms which are less easy to detect or classify. As an example, the meaning of "leg" has changed from SNOMED-RT© to SNOMED-CT®. In RT "leg" was invariably intended to mean "lower leg"; in CT the situation is unclear. The term "34939000: amputation of leg" means in RT: "amputation of lower leg" and in CT: "amputation of any part of the lower limb, including complete amputation" (Jau-11). The authors observed also numerous examples of inconsistent use of "leg" within CT itself: "119675006: leg repair" refers explicitly to "lower leg structure", while "119673004: leg reconstruction" refers explicitly to "lower limb structure" (Jau-12).

7.3.3.2.4 Redundant concepts

8,746 SNOMED-CT® concepts were identified as the seat of redundancies, which is to say: cases where no apparent difference in meaning can be detected between one concept and another one on the basis of the terms that were assigned to it. (This is in reality a severe underestimation, since the authors had set the parameters for matching lexical variants very conservatively, sacrificing recall for precision.) These are all pairs or larger pluralities of terms among which differences in meaning could be identified neither ontologically nor linguistically. Many of them are, it is believed, the result of incomplete or inadequate integration of the Read terms into SNOMED-CT®. An astonishing example is "210750005: traumatic unilateral amputation of foot with complication", which co-exists in SNOMED-CT® with "63132009: unilateral traumatic amputation of foot with complication". (Jau-13)

Of the same nature is the co-existence of "41191003: open fracture of head of femur" and "208539002: open fracture head, femur" (Jau-14), which fit differently into the class hierarchy but in such a way that the technology used in the development of SNOMED-CT® was unable to find the redundancy involved: the former is directly subsumed by "fracture of femur", the latter by "fracture of neck of femur".

Some redundancies become overt only when a larger part of the subsumption hierarchy is examined. Thus one can question to what extent "172044000: subcutaneous mastectomy for gynecomastia" is different from its immediate subsumer "59620004: mastectomy for gynecomastia" when the latter is itself immediately subsumed by "70183006: subcutaneous mastectomy" (Jau-15). All these errors are easily detectable, again, by using semantic distance based algorithms.

7.3.3.2.5 Missing full definitions

The graph expansion algorithm described by the authors was able to detect many cases in which a SNOMED-CT® class is declared to be "primitive" where it could easily have been fully defined. The typical scenario for this type of mistake is one in which a class node introduced by the expansion algorithm subsumes precisely one class in SNOMED-CT®. Examples are "302829009: adenoma of nipple", and "63348002: excision of benign tumor of breast". The latter is especially surprising, given that "46116005: excision of malignant tumor of breast" is itself correctly declared "fully defined". This again poses questions as to the appropriateness of the methodology that is applied in building SNOMED-CT®.

7.3.3.2.6 Mistakes due to lack of an underlying ontological and anatomical theory

Lack of sound mereotopology

It is difficult to imagine that a single connected object can be a proper part of two regions that are topologically disconnected. Despite this, "45684006: structure of tibial nerve" is directly subsumed by both "thigh part" and "lower leg structure", which explicitly refer to the upper and lower parts of the lower limb, respectively (Jau-16).

Omission of obvious relationships

Certainly no large terminology can be expected to be complete. However, one can wonder why "248182008: cracked lips" *is-a* "301346001: finding of appearance of lip" but "80281008: cleft lip" *is-a* "disease" and has no relation at all to "finding of appearance of lip" (Jau-17). Such omissions have the consequence that many sound inferences cannot be made. As another example: "181452004: entire uterus" *part-of* "362235009: entire female internal genitalia", which itself is *part-of* "362236005: entire female genitourinary system". This means, however, that SNOMED-CT® does not allow the inference to "181452004: entire uterus" *part-of* "181440006: female genital tract", since the latter has no relationships with "female internal genitalia", and nor will it allow inferences e.g. to the effect that pregnancy involves the uterus (Jau-18).

Mistakes of this kind can be found, again, only by resorting to additional ontological and anatomical information. For this reason, SNOMED-CT's relational organization is still best conceived as a convenient mechanism for browsing through the terminology in order to find

better descriptors, but not as a representation of how the corresponding instances are related to each other in reality [134].

7.3.3.3 Issues of usefulness

In June 2004, the Advanced Medical Technology Association, the American Hospital Association, the American Medical Association, and the Federation of American Hospitals expressed their concern with the recommendation of the President's Information Technology Advisory Committee (PITAC), Health Care Delivery and IT Subcommittee, calling for the adoption of SNOMED-CT as an alternative to ICD-10-CM [135]. Their main argument was that although it appeared during the NCHVS hearings that SNOMED-CT was supposed to be well designed for the support of the electronic health record, it contains too many terms to feasibly collapse them for American Hospital Association statistical analysis and reimbursement systems. Further, using SNOMED-CT to define all clinical concepts contained in the medical record - discharge summary, operative report pages and pages of progress notes -can only lead to a cumbersome and inefficient process. Reimbursement systems would be difficult to design using SNOMED-CT because of its extreme granularity that often lends itself to redundancy of codes. On the other hand, ICD-10-CM and ICD-10-PCS provide for a more precise selection of codes for summarizing the patient's record.

8 STANDARDIZATION ACTIVITIES ON ONTOLOGIES

8.1 Ontologies in Life Sciences and Health

Ontologies are currently a hot research topic in Life Science and Health, their main purpose being, as it is hoped, to assure semantic interoperability of systems. More than in other domains, it seems, there is a divide between researchers approaching the issue from an information science and software engineering perspective, and those taking a philosophical stance.

The former group understands by 'ontology' a formal representation of a (partial) intensional definition of a conceptualization of an application domain [136], i.e. as a first order vocabulary with semantically precise and formally defined logical terms that stand for concepts and their inter-relationships of an application domain, and thus more as a knowledge representation (and as such not different from the old frame-based or semantic-network variations). This community works with minimalist "models" that then are used as templates to look at 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* [137]. 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, as stated in [138] (pp. 30-31) that '*Model-theoretic semantics does not pretend, and has no way to determine what certain words and statements "really" mean. (…) It (= model theoretic semantics) offers no help in making the connection between the model (the abstract structure) and the real world*'.

It is this lack of explicit reference that disturbs those who take an analytical-philosophical stance, and for whom the term "ontology" denotes rather a representation of reality. This community argues that an ontology should correspond to reality itself in a manner that maximises descriptive adequacy within 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 of the 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 formalised in a way that allows it to support automatic information processing. By 'terminology', they then mean a set of terms that within the linguistic and professional community by which they are used, are accepted designators for the entities represented in the ontology. This is a more precise notion than the one endorsed in linguistic-terminological circles in which terminologies are perceived as a class of systems, either in software or on paper, that contain the terms which specialists in a specific domain are supposed to use when exchanging information. Their purpose is twofold: to allow an unambiguous understanding of what is conveyed, and to stabilise as much as possible the terminology within a specific domain. In this notion, there is not the requirement that there ought to be for each term a referent in reality.

With respect to *patient data* for instance, an ontology enables explicit references to be made to the real instances (particular cases) to which the statements in the electronic health record may refer only implicitly, and to describe in a formally rigorous way the relationships that occur between these instances [139]. A good biomedical ontology thus reflects the most general

categories in reality, i.e. those categories which contain all of the categories into which biomedical data is organized. This makes it possible to link together the general terms that are provided by clinical or biological terminologies.

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 [102, 127, 140-144]. 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.

8.2 Standardization efforts concerning biomedical ontologies

8.2.1. Bioinformatics Data Structures - Framework and Overview (BSR/IEEE 1953-200x)

The scope of this project is to develop a framework for standards and protocols, incorporating existing standards where appropriate, to support the bioinformatics sciences with common definition, storage and exchange of information between them [145]. The project will define efforts in the area of nomenclature, databases, access protocols, benchmarks, and validation suites for a variety of bioinformatics data (e.g., genomics, proteomics, transcriptomes, gene ontology, structural ontology, biological pathways, pharmacogenomics and more).

8.2.2. Standard for Sequence Ontology (BSR/IEEE 1953.1-200x)

The Sequence Ontology is a current working procedure in the Bioinformatics community, this work will formalize that methodology into a standard [146]. The Sequence Ontology (SO) is designed for three different, but related, purposes. The first of these is to provide a structured controlled vocabulary for the description of features that may be described by their spatial location upon sequences and thus annotate these sequences; the second is to provide a structured controlled vocabulary for the description of genes in terms of their sequence characteristics; the third is to provide a structured vocabulary for the description of chromosome and sequence variation within organism. The SO will also provide associated tools for applying and using the vocabularies to support the exchange of genomic sequence annotation.

Work is primarily done through the Sequence Ontology Project (SO), a joint effort by genome annotation centers, including: WormBase, the Berkeley Drosophila Genome Project, FlyBase, the Mouse Genome Informatics group, and the Sanger Institute [147]. They are a part of the Gene Ontology Project and their aim is to develop an ontology suitable for describing biological sequences.

8.2.3. 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 [148] and MGED [149].

Within the Open Biomedical Ontologies (OBO) framework [150], 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 [151]. This standardization initiative is called the OBO Foundry.

8.2.3.1 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.

The members of the OBO Foundry commit in advance to developing their ontologies in cooperation with each other. They agree that when disagreements arise in ontology development, the rationale for these disagreements will be documented and efforts will be made to resolve them in the spirit of scientific inquiry. The primary objective is to establish ontology standards for individual domains of inquiry. We are striving for community acceptance of a single reference ontology for each domain, rather than encouraging rivalry among ontologies, which would defeat the purpose of ontology development. Our goal is to allow multiple ontologies based on common principles to be used in combination, for example when anatomy and process ontologies are combined through additional relationships.

A Web page will be created in which the ontologies being created on the basis of these principles will be listed, together with the names of those groups involved in the OBO Foundry experiment. Further principles will be added over time in order to bring about a gradual improvement in the quality of included ontologies.

By joining the OBO Foundry, the authors of an ontology commit to its maintenance and to soliciting community feedback for its improvement. They also give an assurance that they will work with other groups to ensure that, for any particular domain, there is community convergence on a single reference ontology, which will incorporate multiple perspectives wherever necessary. Application ontologies developed for specific purposes will then be referred back to this reference ontology.

The initial set of principles, which will be extended over time, is listed below.

1. The ontology is open and available to be used by all without any constraint other than (1) its origin must be acknowledged and (2) it is not to be altered and subsequently redistributed under the original name or with the same identifiers.

The OBO ontologies are for sharing and are resources for the entire community. For this reason, they must be available to all without any constraint or license on their use or redistribution. However, it is proper that their original source is always credited. Furthermore, after any

alterations by external users, they must not be redistributed using the original name or with the same identifiers.

2. The ontology is in, or can be instantiated in, a common formal language. The languages supported by OBO are listed at http://obo.sf.net/

The reason for this requirement is that the same tools can then be usefully applied. The use of standard formal languages facilitates shared software implementations.

3. The ontology possesses a unique identifier space within OBO.

The source of terms from any ontology can be immediately identified by the prefix of the identifier of each term. It is, therefore, important that this prefix be unique. Each term in the ontology must have a unique identifier comprising (1) a unique identifier or namespace for the ontology together with (2) a unique term identifier.

4. The ontology provider has procedures for identifying distinct successive versions.

All maintained ontologies change over time and it is important that there is a rigorous way to refer to a particular version and to identify changes, deletions or additions of terms with respect to previous versions, as well as the reasons for such changes. The CVS repository of OBO will maintain all versions.

5. The ontology has a clearly specified and clearly delineated content.

An ontology should have a clearly specified subject matter, and the name of the ontology should make this subject- matter clear. An ontology devoted to, say, cell components should not include terms like: ‘database’, or ‘microscope’, or ‘photograph’. (These terms might, though, belong in other ontologies.)

6. The ontology includes textual definitions for all terms.

Many biological and medical terms may be ambiguous, so terms must be defined in such a way that their precise meaning is clear to a human reader. Some high-level terms may be declared to be primitive. Definitions should be in a form that is intelligible to human users. Equivalent computationally intelligible definitions should also be supplied (see criterion regarding relationships).

7. The ontology uses relations which are unambiguously defined following the pattern of definitions laid down in the OBO Relation Ontology.

The reason for this requirement is so that the meaning of particular relationships (e.g., is_a, part_of) is the same in all ontologies. This requirement is designed to facilitate integration of and reasoning over a plurality of ontologies.

8. The ontology is well-documented.

An ontology should have good overall documentation, which is clearly written for non-experts in ontologies. This documentation should provide a clear description of the domain of the ontology, of how it can be used to support reasoning, and how changes and additions to the ontology can be proposed. (Examples of best practices in ontology documentation will be provided.)

9. The ontology has a plurality of independent users.

Interoperability of ontologies is not an end in itself, but is designed to facilitate new types of biomedical informatics experiment, involving computer-assisted combination of data derived

from different sources. The data which drives such experiments will exist only when ontologies are set to work in different contexts and by independent sets of users.

8.3 International Standardization bodies with ontology-related activities

8.3.1. International Organization for Standardization (ISO)

The work conducted within the ISO Technical Committee 37 (ISO TC 37) - Terminology and other language and content resources - includes the standardization of principles, methods and applications relating to terminology and other language and content resources in the contexts of multilingual communication and cultural diversity [152]. Within this context are worth mentioning:

- ISO 704:2000 Terminology work – Principles and methods
- ISO 860:1996 Terminology work – Harmonization of concepts and terms
- ISO 1087-1:2000 Terminology work – Vocabulary – Part 1: Theory and application
- ISO 15188:2001 Project management guidelines for terminology standardization
- ISO 1087-2:2000 Terminology work – Vocabulary – Part 2: Computer applications
- ISO 12620:1999 Computer applications in terminology – Data categories
- ISO 16642:2003 Computer applications in terminology – Terminological markup framework
- ISO 2788:1986 Documentation – Guidelines for the establishment and development of monolingual thesauri

8.3.2. The International Federation of Library Associations and Institutions (IFLA)

The IFLA is the leading international body representing the interests of library and information services and their users [153]. IFLA is at the basis of the Functional Requirements for Bibliographic Records [154] that identifies and clearly defines the entities of interest to users of bibliographic records, the attributes of each entity, and the types of relationships that operate between entities. It consists of a conceptual model that serves as the basis for relating specific attributes and relationships (reflected in the record as discrete data elements) to the various tasks that users perform when consulting bibliographic records.

These specifications, together with contributions from the Dublin Core Metadata Initiative [155], the INDECS E-Commerce Metadata Model [156], the CIDOC Conceptual Reference Model [157] and the CIMI consortium (which ceased operations at the end of 2003) [158] led to the **ABC-Ontology** which has been developed within the Harmony international digital library project to provide a common conceptual model to facilitate interoperability between metadata ontologies from different domains [159].

8.3.3. The European Committee for Standardization (CEN)

Relevant for ontology in healthcare and life sciences is CEN/TC 251 which takes care of standardization in the field of Health Information and Communications Technology (ICT) to

achieve compatibility and interoperability between independent systems and to enable modularity. This includes requirements on health information structure to support clinical and administrative procedures, technical methods to support interoperable systems as well as requirements regarding safety, security and quality [31]. CEN/TC 251 has 4 working groups of which the objectives of CEN/TC 251 WGII are most close to ontology related issues. They include:

1. The semantic organization of information and representation of knowledge so as to make it of practical use in the domains of health informatics and telematics.
2. The provision of information and criteria to support harmonization in health care and terminological consistency within TC251

The focus of the work is on:

1. terms, concepts and interrelationships of concepts
2. structures for concept systems including those for multi-axial coding schemes
3. representation of clinical knowledge in health information systems
4. guidelines for the production of coding systems
5. systematization of the semantic structure behind the names of compositions and headed sections of the health care record

The production of coding schemes is usually outside the working group's scope.

A fundamental problem for CEN/TC 251 WGII in relation to ontology is its strict adherence to the paradigms put forward by ISO TC37 which are claimed not to be suitable for serious ontology work [160].

8.3.4. IEEE groups

8.3.4.1 IEEE P1600.1: Standard Upper Ontology Working Group (SUO WG)

The SUO WG is developing a standard that will specify an upper ontology to support computer applications such as data interoperability, information search and retrieval, automated inferencing, and natural language processing [161].

8.3.4.2 The Foundation for Intelligent Physical Agents (FIPA)

FIPA is a non-profit IEEE Computer Society standards organization aimed at producing standards for the interoperation of heterogeneous software agents [162]. It was officially accepted by the IEEE as its eleventh standards committee on 8 June 2005.

FIPA was originally formed as a Swiss based organization in 1996 to produce software standards specifications for heterogeneous and interacting agents and agent based systems. Since its beginnings, FIPA has played a crucial role in the development of agent standards and has promoted a number of initiatives and events that contributed to the development and uptake of agent technology. Furthermore, many of the ideas originated and developed in FIPA are now coming into sharp focus in new generations of Web/Internet technology and related specifications.

8.3.5. World Wide Web Consortium (W3C)

The World Wide Web Consortium (W3C) is an international consortium where Member organizations, a full-time staff, and the public work together to develop Web standards. W3C's mission is to lead the World Wide Web to its full potential by developing protocols and guidelines that ensure long-term growth for the Web [36]. Some of its work is directly relevant to ontology and the resulting products are discussed elsewhere in this document. Most prominent is its work on OWL [163].

8.3.6. Object Management Group (OMG)

As described earlier, the Object Management Group (OMG) is an open membership, not-for-profit consortium that produces and maintains computer industry specifications for interoperable enterprise applications [164]. Its membership includes virtually every large company in the computer industry, and hundreds of smaller ones. Most of the companies that shape enterprise and Internet computing today are represented on the Board of Directors.

Central in their work is the multi-platform Model Driven Architecture (MDA), recently underway but already well known in the industry. OMG's own middleware platform is CORBA, which includes the Interface Definition Language OMG IDL, and protocol IIOP. The Object Management Architecture (OMA) defines standard services that will carry over into MDA work shortly. OMG Task Forces standardize Domain Facilities in industries such as healthcare, manufacturing, telecommunications, and others.

8.3.7. Organization for the Advancement of Structured Information Standards (OASIS)

OASIS is a not-for-profit, international consortium that drives the development, convergence, and adoption of e-business standards [165]. The consortium produces more Web services standards than any other organization along with standards for security, e-business, and standardization efforts in the public sector and for application-specific markets. Founded in 1993, OASIS has more than 5,000 participants representing over 600 organizations and individual members in 100 countries.

OASIS Standards are approved within an OASIS Committee, submitted for public review, implemented by at least three organizations, and finally ratified by the Consortium's membership at-large.

Approved OASIS standards that have some relationship with ontologies include:

Darwin Information Typing Architecture (DITA [166]): an architecture for creating topic-oriented, information-typed content that can be reused and single-sourced in a variety of ways. It is also an architecture for creating new topic types and describing new information domains based on existing types and domains. The process for creating new topic types and domains is called specialization. Specialization allows the creation of very specific, targeted document type definitions while still sharing common output transforms and design rules developed for more general types and domains, in much the same way that classes in an object-oriented system can inherit methods of ancestor classes. DITA topics are XML conforming. As such, they are readily viewed, edited, and validated with standard XML tools, although some features such as content referencing and specialization may benefit from customized support.

9 GAPS AND OVERLAPS IN STANDARDS

In an appendix to Richesson et al. [167] the authors present a table of the gaps and overlaps of recommended standards in the domain of clinical research. Grouped by constructs the authors derived from the work plan of CHI, the table is a valuable overview of the current standards environment of that selected domain. Our work attempts to use this as a starting point and improve upon it in the following ways: 1) expand the range of the recommendations beyond those within the domain of clinical research, 2) provide a more granular presentation of the constructs and 3) present information on the qualifications of the recommendations made by the organizations.

As many others have noted, the standardization of the messaging and vocabularies of the domain of clinical research must be done in concert with the standardization of the domain of clinical care if the hoped for improvements in translational medicine, population health and biosurveillance are to be achieved. Thus we have included recommended standards for the domain of clinical care as well as clinical research. The increase in the level of granularity from 27 broad constructs in the Richesson table to 44 sub-domains presented here allows for a better depiction of the scope of coverage of a recommendation. For example, it is possibly misleading to characterize an organization has having recommended a standard for a given area of medical care or research when in fact, the recommendation explicitly covers only a sub-area. We attempt to remedy this with explicitly defining the scope of a recommendation. Lastly, we attempt to add detail to the gap analysis provided in the Richesson table by providing the qualifications upon recommendation made by the recommending organizations.

Domain	Sub-Domain	Organization - Recommended Standard	Scope	Standard Ownership	Conditions on Recommendation
Administration & Finance	Billing and Financial	CHI, HITSP, CCHIT - HIPAA approved codes sets as follows: ICD-9-CM, NDC codes, HCPCS, CPT-4 codes, CDT codes, ABC codes, and DRG codes.	Claim Submission for reimbursement, Health Care Claim Payment/Advice, Eligibility Determination, Prior Authorization and Referral, Enrollment/Disenrollment, Coordination of Benefits, Claims Status Inquiry, Appeals, Certificate of Medical Necessity, and Employer Identifiers.	NCHS, FDA, CMS, AMA, ADA, Alternative Link, and CMS respectively	None
Administration & Finance	Messaging	CHI - HL7 V2.3+ HITSP, CCHIT - HL7 V2.4, X12	Order Entry, Scheduling, Medical record/Image Management, Patient Administration, Observation Reporting, Financial Management, Patient Care, Public Health Notification	Health Level 7 - HL7 V2.3+, V2.4 ASC - X12	The CHI recommendation notes that the HL7 v2.3+ standard does not provide adequate constraints to promote semantic interoperability between sender and reciever systems and advises an aggressive move towards HL7 Version 3.0 as a remedy.
Chronic Disease	Messaging	CHI - IEEE 11073 Series	Device to device connectivity for data exchange	The Institute of Electrical and Electronics Engineers	CHI recommended for federal intra-agency use only, characterizes the standard as emerging
Clincial Care	Medical Devices and Supplies	None	As scoped by CHI this standard would be used to inventory medical devices and supplies and document their utilization by health services establishments and to regulate medical device and supply availability and utilization in the community by public health agencies.	None	The recommendation of CHI was to encourage the Global Medical Device Nomenclature (GMDN) and the Universal Medical Device Nomenclature System (UMDNS®) to merge, and, once merged, CHI would re-evaluate/adopt the resulting terminology.

Clincial Care	Physiology	None	As scoped by CHI this standard would be used to describe or infer human physiology at least at the organ system, cellular, and biochemical levels	None	The CHI workgroup could not identify an acceptable terminology that covered cellular physiology. It found the VA NDF-RT medication physiologic effect axis to be too narrowly focused toward drug physiology to be of general applicability. The clinical terms in LOINC® and the appropriate hierarchies within SNOMED CT® were found to have significant weaknesses for use in this domain.
Clinical Care	Adverse Events	None	A standard would be used to document any adverse change in health of a participant in a clinical trial while they are the treatment (study medication, application of the study device, etc.) or within a pre-specified period of time after their treatment has been completed.	None	ICH has recommended MedDRA as the standard for this domain. FDA uses MedDRA in its Adverse Event Reporting System.
Clinical Care	Allergy	CHI, HITSP, CCHIT - HL7 v2.4+, Unique Ingredient Identifier (UNII), RxNorm, NDF-RT, SNOMED CT	Allergy, but not adverse reations	Health Level 7, FDA, NLM, VA, and IHTSDO respectively	The CHI recommendation was conditional upon satisfaction of the following conditions: 1) At least 90% of the UNII codes for active ingredients for approved and marketed prescription drugs have been completed by the FDA, 2) UNII codes are made publicly available through a simple download mechanism, 3) Updates to UNII codes are made publicly available on at least a quarterly basis, 4) The plan for UNII code development and maintenance beyond active ingredients for approved and marketed prescription drugs is documented, and 5) The updated version of the NDF-RT, 2006 or later, is made publicly available through a simple download mechanism.

Clinical Care	Anatomy	CHI - Systematized Nomenclature of Medicine Clinical Terms (SNOMED CT) and National Cancer Institute's Thesaurus (NCIt)	Anatomical location of a procedure or of an injury; Anatomical description of a specimen; Subcelluar anatomy; Measured or inferred physiology of organ or organ system; Measured or inferred physiology of cell; Morphology. Not a standard for patient physiology.	IHTSDO and NCI, respectively	The CHI recommendation cites a lack of coverage of the subcellular structures required for research by SNOMED CT and this is the reason for the augmentation with the NCI Thesaurus. Also, the recommendation claims that SNOMED CT does not suit the needs of a general practioner for a stand-alone, limited clinically oriented terminology for anatomy that relates to the more complex ones. Standardized methods of post-coordination of terms are not available making the full use of the anatomy terminology limited.
Clinical Care	Clinical Encounters	CHI -HL7 V2.4+ HITSP, CCHIT - HL7 CDA R2, HL7-ASTM CCD	CHI - HL7 V2.4+ for Clinical Encounters, Admission Information, Patient Transfer Information, Discharge Information, Provider Information, Accident Information, Death and Autopsy Information HITSP, CCHIT - HL7 CDA R2 for Clinical Encounters, Consultation Letters, Discharge Information, HL7-ASTM CCD for Patient Transfer Information	Health Level 7	The CHI recommendation was based on the 92 data fields of the HL7 message for the clinical encounter. Of the fields requiring the use of value sets, 5 were found to have gaps:1) Explicit support for home health, field and virtual encounters, 2) Support for clinical services that do not meet definition of clinical encounter, 3) National Provider System identifiers for practitioners and healthcare organizations, 4) Standard location identifiers, and 5) Standard hospital service names
Clinical Care	Demographic	CHI - HL7 V2.4+ code sets HITSP, CCHIT - CDC Race and Ethnicity Code Sets	The scope of the standard is that which can be used to set the requirements for collecting and storing specific patient demographic data.	Health Level 7 - HL7 V2.4+ code sets CDC - CDC Race and Ethnicity Codes Sets	The CHI recommendation lists the following gaps in the HL7 V2.4+ Code Sets: Marital Status contains overlaps, Gender should be more restrictive, No element for Insurance Status and its associated values exists, Living status data element contains gaps of coverage for those situations when information related the presence of another individual in a residence is needed to determine the

					supervisory care.
Clinical Care	Diagnosis and Problem List	CHI - SNOMED CT	Clinical Diagnosis/Problems, Subject Symptoms/Observed Findings, Synonyms but not Nursing Diagnoses, Modifiers and Descriptors, Dental or Alternative Medicine.	IHTSDO	The CHI recommendation is not conditional, however, the entirety of SNOMED CT is not being recommended but rather only the content located in the SNOMED CT® concept groupings of "disorders" and "findings".
Clinical Care	Functioning and Disability	CHI - Clincal LOINC, International Classification of Functioning, Disability and Health (ICF), SNOMED CT, HL7 v2.4+, and HL7 CDA	Functioning and Disability Content, Patient/Client Assessment Forms that include disability and functioning content	The Regenstrief Institute, World Health Organization, IHTSDO, and Health Level 7 respectively	None
Clinical Care	History and Physical	None	As scoped by CHI this standard would cover History of Present Illness, Review of Systems, Past Medical/Surgical History, Family History, Social History, Non-Medication Allergies, Vital Signs, Physical Exam Observations, Physical Exam Findings	None	The reason for the failure to identify a standard on the part of the CHI workgroup was the observed variability, in the format and content of a History & Physical. They concluded that It is presently not standardized and is typically dependent upon the clinical judgment of the practitioner.

Clinical Care	Non-Laboratory Procedures and Interventions	CHI - SNOMED CT	Procedure by site (e.g. on body part), Procedure by method, Procedure by intent (e.g. therapeutic, preventive, palliative), Procedure by focus, Regime / Therapy, Procedure by device, but not Dental, Alternative Medicine, Laboratory Procedures (addressed elsewhere) or Administrative / Management procedure	IHTSDO	The CHI recommendation does not include the entirety of SNOMED CT but rather only the content that pertains to interventions and procedures, found within the procedure axis of SNOMED CT and it explicitly excludes the content contained in the SNOMED CT hierarchies of Procedures by method (covered in the laboratory domain), Administrative procedures (covered in the financial billing domain) or Laboratory Procedures (covered by the Laboratory domain)
Clinical Care	Nursing	CHI - SNOMED CT	Assessment / Observations, Plan / Goals, Diagnosis, Interventions, and Evaluation / Outcome but not Intensity of Nursing Care	IHTSDO	The CHI recommendation does not include the entirety of SNOMED CT but rather only the content that contains nursing concepts as modeled / integrated from the source nursing terminologies such as the Georgetown Home Health Care Classification, the Omaha System and the Nursing Interventions Classification (NIC); Intervention Concepts from the Perioperative Nursing Data Set (PNDS); Nursing diagnosis and problem concepts from NANDA, PNDS, HHCC, and Omaha.
Clinical Care	Subject and Provider Identifiers	HITSP - National Provider Identifier (NPI), Presciber Number	The NPI is an identification number assigned to health care providers The DEA Prescriber Number is assigned to health care providers allowing them to prescribe controlled substances	CMS - NPI DEA - Prescriber Number	None

Clinical Care	Text Based Reports	CHI - HL7 CDA Release 1.0 HITSP, CCHIT - HL7 CDA R2	CHI recommendation limited to Text-Document structure and syntax, Electronic Signature, Document Section Headings, and Clinical Document Types/Titles.	Health Level 7	None
Clinical Research	Case Report Design	None	Clinical Trial data recording document	None	None
Clinical Research	Genes	CHI - Human Gene Nomenclature (HUGN)	Genes, but not proteins	Human Genome Organization (HUGO)	The CHI recommendation is not conditional however it stresses that HUGN is a highly focused standard dealing with the nomenclature of human genes and thus even when complete HUGN will not be a terminology for DNA sequences that are transcribed to RNA but not subsequently translated to proteins, or for highly conserved DNA sequences with unknown biological functions.
Clinical Research	Proteins	None	As scoped by CHI this standard would be used to describe proteins involved in the pathogenesis, drug resistance, or identification of infectious diseases and would provide a Protein Nomenclature	None	The CHI workgroup could not identify an acceptable terminology that covered proteins.
Clinical Research	Protocol Deviations	None	Events occurring within a clinical trial that are not defined within the trial's protocol	None	None
Clinical Research	Study Descriptive Information	None	Study Title, Purpose, Protocol Summary, Basic Eligibility Criteria, Study Site Location(s), Study Site Contact Information	None	None

Clinical Research	Study Eligibility Criteria	None	Age, Sex, Disease, Disease Stage, Health Status, Current and Past Treatments, Race, Ethnicity, etc.	None	None
Clinical Research	Study Events	None	Schedule of Events within a Clinical Trial	None	None
Clinical Research	Subject Disposition Descriptors	None	A standard in this domain would be used for recording a participant's status relative to a clinical trial (e.g. eligible, enrolled, lost to follow-up)	None	None
Imaging	Data from multiple media sources	CHI - Digital Imaging Communications in Medicine (DICOM)	Incorporation of multimedia information into patient records including Images, Audio information, Waveform data, and Video information but not Telemedicine	National Electrical Manufacturers Association (NEMA)	None
Imaging	Messaging	CHI - Digital Imaging Communications in Medicine (DICOM) HITSP, CCHIT - IHE XDS-I Cross-Enterprise Image Information Sharing integration profile	Encoding of images and image-related information and how such information objects are exchanged between instruments.	National Electrical Manufacturers Association (NEMA) IHE - IHE XDS-I (includes DICOM standard)	CHI recommendation notes that the standard lacked the ability to permit understandable exchange of both the image and text portions of a DICOM message.

Immunizations	Immunizations	CHI - HL7 V2.3.1+, Clinical Vaccine Formulation (CVX), Manufacturer (MVX) HITSP, CCHIT - HL7-ASTM CCD	CHI - HL7 2.3.1, and future versions, for the messaging standard and for the immunization registry terminology, the CVX and MVX codes from HL7, maintained by the CDC. HITSP, CCHIT - HL7 ASTM CCD for import of immunization history from a PHR	Health Level 7, CDC	The CHI recommendation is not conditional, however it acknowledges that the recommended standards are adequate only for the limited purpose of exchanging immunization information and will not be so for future needs including decision support; ensuring more consistent interpretation of categorizations and term relationships both within and among organizations; facilitating the ability to assess immunization coverage for populations; and allowing healthcare organizations to better integrate their various IT applications into one system.
Lab	Laboratory Result Contents	CHI, HITSP, CCHIT - SNOMED CT for Result Contents HITSP, CCHIT - LOINC CCHIT - ELINCS v2.1	SNOMED CT for Ordinal results, Anatomical Pathology report codes, Living Organism codes, Hematology result codes, Immunohematology (Blood Bank) result codes, Units, Other descriptive laboratory test result codes, Standard Comments, Abbreviations, Nonhuman Specimen Type but not Numerical results including titers, Normal result and other flag indicators orOut of range results LOINC for senstivitiy ELINCS v2.1 for microbiology lab results	IHTSDO	None

Lab	Laboratory Results Names	CHI, HITSP, CCHIT - Laboratory LOINC	Lab test result name, not Lab test ordering or Lab test result value	The Regenstrief Institute	1. Ease introduction of LOINC codes by providing consistent and widespread appearance of XXX codes. 2. Introduce a hierarchy to LOINC to allow for standard aggregation of terms across the healthcare system, ease in identifying needed terms, and identification of terms to assign within an institution. 3. Improve content coverage, definitions, and unrecognized synonymy.
Lab	Laboratory Test Order Names	CHI, HITSP, CCHIT - LOINC	Laboratory test name for clinical pathology orders, anatomical pathology orders, test panel names for clinical pathology orders, and test panel names for anatomical pathology orders.	The Regenstrief Institute	The CHI recommendation was conditional upon the following improvements being made: 1) Introduction of a hierarchy to LOINC® would allow for standard aggregation of terms, 2) Improvements to the naming of panels to allow the laboratory to specify the exact test to be run, 3) Improvements to definitions and unrecognized synonymy, 4) Integrate genomic tests by allowing users to search using a disease specific key word strategy and expand the coverage of tests to include gene array and proteomic based laboratory tests, and 5). Make available a map LOINC to CPT codes to facilitate the production of administrative data from clinical applications.
Lab	Messaging	HITSP, CCHIT - HL7 v2.4, HL7 v2.5	Send an order and receive results for a laboratory test	Health Level 7	Further work is needed to define the ordering messages, should include an order number for tracking

Lab	Units	CHI - HL7 V2.x codes for Units HITSP - Unified Codes for Units of Measure (UCUM)	Ordinal results, Anatomical Pathology report codes, Living Organism codes, Hematology result codes, Immunohematology (Blood Bank) result codes, Units, Other descriptive laboratory test result codes, Standard Comments, Abbreviations, Non-human Specimen Type but not Numerical results including titers, Normal result and other flag indicators or Out of range results	Health Level 7 - HL7 V2.x codes for Units The Regenstrief Institute - UCUM	None
Medications	Active Ingredients	CHI, HITSP - FDA Established Name for active ingredient & Unique Ingredient Identifier (UNII) codes.	Active Ingredients	FDA	CHI recommendation cites issue with salt forms or base forms
Medications	Clinical Drugs	CHI, HITSP, CCHIT - Semantic Clinical Drug (SCD) of RxNorm	Clinical Drugs, i.e. Active ingredients, their strength, and the dose form of the drug	National Library of Medicine	Recommendations cite incomplete coverage of RxNorm terminology on multi-ingredient OTC drugs and contrast media.
Medications	Drug Classifications	CHI, HITSP - National Drug File Reference Terminology (NDF-RT)	CHI limits to Drug Classifications by Physiologic Effect and Mechanism of Action	Veterans Administration	Recommendation does not include other important classification scales of drugs such as intended therapeutic use, chemical structure, pharmacological properties, and FDA approved indications
Medications	Drug Product	CHI, HITSP, CCHIT - National Drug Code (NDC)	Drug Product	FDA	CHI recommendation cited issues with coverage incorrectly listings
Medications	Manufactured Dosage Form	CHI - CDER Data Standards Manual	Manufactured Dosage Form	FDA	None

Medications	Messaging	CHI, HITSP, CCHIT - SCRIPT	Retail pharmacy transactions	National Council for Prescription Drug Programs (NCPDP)	None
Medications	Package	CHI - CDER Data Standards Manual	Package	FDA	None
Medications	Special Populations	CHI - HL7 V2.4 gender, race & ethnicity codes HITSP - CDC Race and Ethnicity Code Sets	Sub-groups of the population using medications for the treatment or prevention of medical conditions e.g. by Gender, Age, Race/Ethnicity	Health Level 7 - HL7 v2.4 codes CDC - CDC Code Sets	CHI notes that HL7code sets include race classifications being socio-cultural rather than scientifically derived and ethnicity code set having incomplete coverage
Medications	Structured Product Label	CHI - LOINC® Clinical SPL	Structured Product Label	The Regenstrief Institute	At the time of the CHI recommendation the SPL terminology was in the process of being incorporated into LOINC® Clinical.
Public Health	Chemicals	CHI - Substance Registry System (SRS)	Chemicals of importance to health care outside of medications	Environmental Protection Agency (APA)	The CHI recommendation was conditional upon 1) establishment of communication protocol so that medical needs are addressed in a timely fashion, 2) development of a mechanism for matching other date against the SRS, 3) making available a view of SRS data in a format for healthcare use , and 4) Requirement for registering an Object Identifier (OID) if it is to be used in HL7 messaging.
Public Health	Population Health	None	As scoped by CHI this standard would be use for public health Reporting and population health statistics.	None	The reason for the failure to identify a standard on the part of the CHI workgroup was the diversity of terminology needs found during their investigation of population health reporting needs.

10 ACRONYMS

AAAI..... American Association for Artificial Intelligence

AAMC..... American Association of Medical Colleges

AAMI..... Association for the Advancement of Medical Instrumentation

ACC ..... American College of Cardiology

ACDB..... Ambulatory Care Data Base

ACM..... Association for Computing Machinery

ACMI ..... American College of Medical Informatics

ACR ..... American College of Radiology

ADA ..... American Dental Association

ADG ..... Ambulatory Diagnostic Groups

AEP ..... Appropriateness Evaluation Protocol

AFNOR ..... French standards organization

AHA ..... American Hospital Association

AHCPR ..... Agency for Healthcare Policy and Research

AHIC..... American Health Information Community, USA

AHIC..... Australian Health Information Council, Australia

AHIMA ..... American Healthcare Information Management Association

AIM ..... Advanced Informatics in Medicine, former EU programme

AIMBE..... American Institute for Medical and Biological Engineering

AMA ..... American Medical Association

AMIA..... American Medical Informatics Association

ANSI ..... American National Standards Institute

ANSI HISB ..... American National Standards Institute's Healthcare Informatics Standards Board

Arden Syntax..... (Standard for defining Medical Logic Modules)

ARPA..... Advanced Research Projects Agency

AS4..... Laboratory coding set ASTM E1238

ASC X12..... US SDO for EDI

ASC..... Accredited Standards Committee

ASCII ..... American Standards Committee Information Interchange

ASN.1..... Abstract Syntax Notation One

ASTM..... American Society for Testing and Materials

BIN..... Belgium National standards body

BIPM ..... Bureau international des poids et mesures

BNF..... Bachus Naur Form

BSI ..... British Standards Institute (UK NSB)

BSR ..... Basic Semantic Repository

CAP ..... College of American Pathologists

Carelink ..... eHealth Competence Centre in Sweden

CAS ..... Clinical Abstract Codes

CCD ..... Clinical Care Document, merging of HL7-CDA and ASTM-CCR

CCHIT ..... Certification Commission for Health Information Technology, USA

CCOW ..... Clinical Context Object Workgroup, in HL7

CCR ..... Continuity of Care Record

CDA ..... Clinical Document Architecture, in HL7

CDC ..... Centers for Disease Control and Prevention, in USA

CDT ..... Current Dental Terminology

CEN TC 251 ..... Technical Committee for Health Informatics in CEN

CEN ..... Comitee Europeen de Normalisation

CENELEC ..... European Committee for Electrotechnical Standardization

CfH ..... Connecting for Health, England

CGs ..... Conceptual graphs

CHI ..... Consolidated Health Informatics, e-gov USA

CHIME ..... The College of Healthcare Information Management Executives

CHIN ..... Community Health Information Network

CLIPS ..... C Language Integrated Production System

COACH ..... Canadian Organization for Advancement of Computers in Health, Canada

CORBA ..... Common Object Request Broker Architecture.

COSTA RT ..... Coding Symbols for Thesaurus of Adverse Reaction Terms.

CPRI ..... Computer-based Patient Record Institute, USA

CPT ..... Common Procedural Terminology. Current Version 4.

CRS ..... Care Record Summary, based on HL7 CDA

CRS ..... Care Records Service, NHS England

DARPA ..... Defence Advanced Research Projects, predecessor of ARPA

DG-INFSO ..... DG Information Society, European Commission (formerly DG-XIII)

DICOM ..... Digital Image COMmunication

DIN ..... German Standards organization

DITA ..... Darwin Information Typing Architecture

DMP ..... Dossier Médical Personnel, formerly “dossier médical partagè", France

DNS ..... Domain Name System.

DoD ..... Department of Defence, USA

DRG ..... Diagnostic Related Groups

DSM-IV ..... Diagnostic and Statistical Manual of Mental Disorders.

DTD ..... Document Type Definition

ECMA ..... European Computer Manufacturers Association

EDI ..... Electronic Document Interchange

EFMI ..... European Federation for Medical Informatics

EHR ..... Electronic Health Record

EHRcom ..... CEN standard on EHR communication (EN 13606)

EHTEL ..... European Health Telematics association

EN ..... fully balloted CEN standard

EU ..... European Union

EuroRec..... European Institute for Health Records

FDA..... Food and Drug Administration, USA

FIPA ..... Foundation for Intelligent Physical Agents

F-Logic ..... Frame Logic

FTP..... File Transfer Protocol

GAO..... General Accounting Office, USA

GEM ..... Guideline Element Model

Gematic ..... eHealth Competence Centre in Germany

GFP ..... Generic Frame Protocol

GLIF ..... Guidelines Interchange Format

GP ..... General Practitioner

GRAIL ..... GALEN representation and integration language

GUI..... Graphical User Interface

HCFA..... (former) Health Care Financing Administration, USA

HIE ..... Healthcare Information Exchange, USA

HIMSS ..... Healthcare Information and Management Systems Society

HIPAA ..... Health Insurance Portability and Accountability Act, USA

HIS ..... Hospital Information System

HISPC ..... Health Information Security and Privacy Collaboration, USA

HISPP..... Healthcare Informatics Standards Planning Panel, USA

HITSP ..... Health Information Technology Standards Panel, USA

HL7 ..... Health Level Seven

HMO ..... Health Maintenance Organisations, USA

HTML ..... Hypertext Markup Language

ICCS..... International Classification of Clinical Services

ICD..... International Classification of Disease

ICD10..... Latest version, International Classification of Disease

ICD-9-CM..... ICD9 with Clinical Modification, USA

ICHPPC..... International Classification of Health Problems in Primary Care

ICPC..... International Classification of Primary Care

ICT ..... Information and Communication Technology

IEC ..... International Electrotechnical Committee

IEEE..... Institute of Electrical and Electronic Engineers

IETF ..... Internet Engineering Task Force

IFCC ..... International Federation for Clinical Chemistry and Laboratory Medicine

IFIP..... International Federation of Information Processing

IFLA ..... International Federation of Library Associations and Institutions

IHE ..... Integrating the Healthcare Enterprise

IMIA..... International Medical Informatics Association

IOM..... Institute of Medicine, USA

IS ..... fully balloted ISO standard

ISO TC 215 ..... Technical Committee on Health Informatics in ISO

ISO ..... International Standards Organization

ITU ..... International Telecommunication Union

IUPAC ..... International Union of Pure and Applied Chemistry

KIF ..... Knowledge Interchange Format

KRS ..... knowledge representation system

LAN ..... Local Area Network

LIS..... Laboratory Information System

LOINC ..... Laboratory Observation Identifiers Names and Codes

MDF ..... Message Development Framework, in HL7

MED ..... Medical Entities Dictionary

MEDDRA ..... Medical Dictionary for Drug Regulatory Affairs

MEDIX..... Medical Data Interchange Standard, IEEE P1157

MESH..... Medical Subject Headings, by NLM, USA)

MGMA..... Medical Group Management Association

MIB ..... Medical Information Bus, IEEE 11073, Standard for Medical Device Communications

MIB ..... Medical Informatics Bus IEEE P1073

MRI ..... Medical Records Institute

NAHIT ..... National Alliance for Health Information Technology, USA

NANDA ..... North American Nursing Diagnoses Association

NCCLS..... National Committee for Clinical Laboratory Standards

NCHS..... National Center for Health Statistics

NCI..... National Cancer Institute, USA

NEHTA ..... National E-Health Transition Authority, Australia

NEMA ..... National Electrical Manufacturing Association, USA

NHIG..... National Health Information Group, Australia

NHIMAC ..... National Health Information Management Advisory Council, Australia

NHIN..... Nationwide Health Information Network, USA

NHS..... National Health Service, England

NHSIA ..... (former) National Health Service Information Authority, England

NIC..... Nursing Intervention Classification

NICTIZ ..... eHealth Competence Centre in the Netherlands

NIST..... National Institute of Standards and Technology, USA

NLM..... National Library of Medicine, USA

NSB..... National Standardization Bodies, members of CEN

NSF ..... UNational Science Foundation, USA

OASIS ..... Organization for the Advancement of Structured Information Standards

OBO ..... Open Biomedical Ontologies

OCL ..... Object Constraint Language

OCML ..... Operational Conceptual Modelling Language

OIL ..... Ontology Inference Layer

OKBC..... Open Knowledge Base Connectivity

OLE..... Object Linking and Embedding

OMG ..... Object Management Group

OML ..... Ontology Markup Language

ONCHIT ..... Office of the National Coordinator for Health Information Technology, USA

OSF ..... Open Systems Foundation

OSI..... Open Systems Interconnect

OTA ..... Office of Technology Assessment, USA

OWL ..... Web Ontology Language

PACS..... Picture Archiving and Communication System

PCC..... Patient Care Coordination Framework, in IHE

PCP ..... Primary Care Provider

PDS ..... Personal Demographics Services, England

PHR ..... Personal Health Record

PITAC ..... President's Information Technology Advisory Committee, in USA

POD ..... Predefined Operational Domain (see Affinity Domain in IHE-XDS)

RDF ..... Resource Description Framework

Read Codes ..... United Kingdom National Health Service Clinical Terms

RHIO ..... Regional Health Information Organization, USA

RIDE ..... "A Roadmap for Interoperability of eHealth Systems in Support of COM 356 with Special Emphasis on Semantic Interoperability", a EU project

RIM ..... Reference Information Model, in HL7

RIS ..... Radiology Information System

RSNA ..... Radiological Society of North America

RuleML ..... Rule Markup Language

SDO ..... Standards Development Organization

SIG ..... Special Interest Group

SMTP ..... Simple Mail Transfer Protocol

SNOMED CT ..... SNOMED Clinical Terms (merging of SNOMED RT and UK Read Codes - Clinical Terms)

SNOMED RT ..... SNOMED Reference Terminology

SNOMED ..... Systematized Nomenclature of Human and Veterinary Medicine

SOA ..... Service Oriented Architecture

SUO WG ..... IEEE P1600.1: Standard Upper Ontology Working Group

TC ..... Technical Committee

UMDNS ..... Universal Medical Device Nomenclature System, by ECRI

UML ..... Unified Modeling Language, by OMG

UMLS ..... Unified Medical Language Systems, Metathesaurus by NLM, USA

UN/EDIFACT ..... United Nations/Electronic Data Interchange for Administration, Commerce and Transport

URL ..... Universal Resource Locator

VA ..... Veterans Administration, USA

W3C ..... World Wide Web Consortium

WAN ..... Wide Area Network

WEDI ..... Workgroup on Electronic Data Interchange

WEEB ..... Western Europe EDIFACT Board

WHO ..... World Health Organization

WONCA ..... World Organization of National Colleges, Academies and Academic Associations of General Practitioners

WSML ..... Web Service Modeling Language

WWW ..... World Wide Web

XAD ..... XDS Affinity Domains

XDS ..... Cross-Enterprise Sharing, in IHE

XDS-MS ..... Cross-Enterprise Sharing of Medical Summaries, in IHE PCC

XML ..... Extensible Mark-up Language

11 RELEVANT WEB SITES

AHIC - American Health Information Community (the Community)	http://www.hhs.gov/healthit/ahic.html
AHIMA - American Health Information Management Association	http://www.ahima.org
AMA - American Medical Association	http://www.ama.assn.org
ANSI - American National Standards Institute	http://www.ansi.org
ANSI –HITSP - Healthcare Information Technology Standards Panel	http://www.ansi.org/standards_activities/standards_boards_panels/hisb/hitsp.aspx?menuid=3
ASTM - American Society for Testing and Materials	http://www.astm.org
CEN - Comité Européen de Normalisation	http://www.cenorm.be
CEN/TC 251 - Technical Committee 251 "Health Informatics"	http://www.centc251.org
CIHI - Canadian Institute for Health Information	http://www.cihi.ca
CORBA - Common Object Request Broker Architecture	http://www.corba.org
DICOM - Digital Imaging and Communications in Medicine Standard	http://www.xray.hmc.psu.edu/dicom/dicom_intro/DICOMIntro.html
ebXML - Electronic Business using eXtensible Markup Language	http://www.ebxml.org
EDI - Electronic Data Interchange	http://www.premenos.com
EHTO - European Health Telematics Observatory	http://www.ehto.be
ETSI - European Telecommunications Standards Institute	http://www.etsi.org
HIMSS - Health Care Information and Management Systems Society	http://www.himss.org
HL7 - Health Level Seven	http://www.hl7.org
HON - Health on the Net Foundation	http://www.hon.ch
IEC - International Electrotechnical Commission	http://www.iec.ch
IEEE - Institute of Electrical and Electronics Engineers	http://www.ieee.org
ISO - International Standards Organization	http://www.iso.ch
ITU - International Telecommunications Union	http://www.itu.int
LOINC - Logical Observation Identifiers Names and Codes	http://www.loinc.org
NAHIT - National Alliance for Health Information Technology	http://www.nahit.org
NAHIT, Directory of eHealth Standards	http://www.nahit.org/hitsdir/pgLCA

NEMA - National Electrical Manufacturers Association	http://medical.nema.org
NIH - The Combined Health Information Database	http://www.chid.nih.gov
NLM - U.S. National Library of Medicine	http://www.nlm.nih.gov
OASIS - Organisation for the Advancement of Structured Information Standards	http://www.oasis-open.org
OMG - Object Management Group	http://www.omg.org
ONCHIT - Office of the National Coordinator for Health Information Technology, USA	http://www.hhs.gov/healthit/
SNOMED International	http://www.snomed.org
Telemedicine Glossary and Links	http://www.hscsyr.edu/telemed/glossary.html
W3C - World Wide Web Consortium	http://www.w3.org
WHO - World Health Organization	http://www.who.ch

12 REFERENCES

[1] eyeGene Project. The eyeGene Project Homepage. [cited 2008 July 14]; Available from: http://www.nei.nih.gov/resources/eyegene.asp

[2] Zerhouni E. The NIH Roadmap. Science. 2003 3 October 2003;302:63-72.

[3] Ride Project. A Roadmap for Interoperability of eHealth Systems [cited 2008 July 11]; Available from: http://www.srdc.metu.edu.tr/webpage/projects/ride/

[4] Cimino JJ EG, Zeng Q. . Supporting infobuttons with terminological knowledge. . J Am Med Inform Assoc Symp Suppl 1997:528-32.

[5] International Organization for Standardization. ISO Web Site. [cited 2008 June 20]; Available from: http://www.iso.org/

[6] European Committee for Standardization. European Committee for Standardization. 2008 [cited 2008 July 14]; Available from: http://www.cen.eu/cenorm/homepage.htm

[7] Digital Imaging and Communications in Medicine. DICOM Homepage. [cited 2008 July 14]; Available from: http://medical.nema.org/

[8] Health Level 7. Health Level Seven Version 3.0. 2007 [cited 2007 July 23]; Available from: http://www.hl7.org/

[9] American Society for Testing and Materials International. ASTM Homepage. 2007 [cited 2007 July 23]; Available from: http://www.astm.org

[10] Institute of Electrical and Electronics Engineers. Institute of Electrical and Electronics Engineers. 2008 [cited 2008 July 10]; Available from: http://www.ieee.org/portal/site

[11] World Health Organization. International Classification of Diseases (ICD). 2008 [cited 2008 July 16]; Available from: http://www.who.int/classifications/icd/en/

[12] World Health Organization. International Classification of Diseases for Oncology, 3rd Edition (ICD-O-3). 2008 [cited 2008 July 16]; Available from: http://www.who.int/classifications/icd/adaptations/oncology/en/

[13] Logical Observation Identifiers Names and Codes. LOINC Homepage. [cited 2008 July 14]; Available from: http://www.loinc.org/

[14] Northrop Grumman Corporation. Welcome to MedDRA and the MSSO. 2008 [cited 2008 July 15]; Available from: http://www.meddramsso.com/MSSOWeb/index.htm

[15] International Healthcare Terminology Standards Development Organization. IHTSDO SNOMED CT Webpage. [cited 2008 June 30]; Available from: http://www.ihtsdo.org/snomed-ct/.

[16] National Center for Health Statistics. International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM). 2008 [cited 2008 July 15]; Available from: http://www.cdc.gov/nchs/about/otheract/icd9/abticd9.htm

[17] Jenders RA. The Arden Syntax for Medical Logic Systems: An ANSI Standard Sponsored By Health Level Seven. 2006 [cited 2008 July 16]; Available from: http://cslxinfmtcs.csmc.edu/hl7/arden/

[18] InterMed Consortium. Towards A Representation for Sharable Guidelines. 2008 [cited 2008 July 16]; Available from: http://www.glif.org/glif_main.html

[19] Michael G. Guideline Elements Model. 2008 [cited 2008 July 16]; Available from: http://gem.med.yale.edu/default.htm

[20] Ceusters W, Smith, B. Tracking Referents in Electronic Health Records. In: Engelbrecht R. GA, Lovis, C., ed. *Connecting Medical Informatics and Bio-Informatics: Proceedings of MIE 2005 The XIXth International Congress of the European Federation for Medical Informatics*. Amsterdam: IOS Press 2005:71-6.

[21] Smith B, Ashburner, M., Rosse, C., Bard, J., Bug, W., Ceusters, W., Goldberg, L., Eilbeck, K., Ireland, A., Mungall, C., Leontis, N., Rocca-Serra, P., Ruttenberg, A., Sansone, S., Scheuermann, R., Shah, N., Whetzel, P., Lewis, S.,. The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration. Nature Biotechnology. 2007;25(11):1251-5.

[22] e-Health Standardization Focus Group. Current and future standardization issues in the e-Health domain: Achieving interoperability 2004; Bruxelles: CEN/ISSS; 2004.

[23] Smith B, Ceusters, W.,. HL7 RIM: An Incoherent Standard. Studies in Health Technology and Informatics. 2006;124:133-8.

[24] Smith B. HL7 Watch. [cited 2008 June 30]; Available from: http://hl7-watch.blogspot.com/

[25] American National Standards Institute. ANSI Homepage. [cited 2008 June 30]; Available from: http://www.ansi.org/.
[26] American National Standards Institute. ANSI Essential Requirements: Due process requirements for American National Standards. New York; 2008 January.
[27] International Electrotechnical Commission. International Electrotechnical Commission. 2008 [cited 2008 July 12]; Available from: http://www.iec.ch/
[28] International Telecommunication Union. International Telecommunication Union 2008 [cited 2008 July 10]; Available from: http://www.itu.int/net/home/index.aspx
[29] International Organization for Standardization Technical Committee. ISO/IEC JTC 1. [cited 2008 June 30]; Available from: http://isotc.iso.org/livelink/livelink?func=ll&objId=755080&objAction=browse&sort=name.
[30] ISO/TC215. Technical Committee of ISO for Health Informatics. 2008 [cited 2008 July 11]; Available from: http://isotc.iso.org/livelink/livelink?func=ll&objId=529137&objAction=browse&sort=name
[31] CEN/ISSS. CEN Technical Committee 251. [cited 2008 June 30]; Available from: http://www.centc251.org/.
[32] Clinical Data Interchange Standards Consortium. CDISC Homepage. [cited 2008 July 14]; Available from: http://www.cdisc.org/index.html
[33] International Council of Nurses. International Council of Nurses - Homepage. 2008 [cited 2008 July 14]; Available from: http://www.icn.ch/
[34] International Medical Informatics Association. International Medical Informatics Association. 2008 [cited 2008 July 15]; Available from: http://www.imia.org/
[35] United Nations Economic Commission for Europe. United Nations Economic Commission for Europe. 2008 [cited 2008 July 16]; Available from: http://www.unece.org/
[36] World Wide Web Consortium. W3C Homepage. [cited 2008 June 30]; Available from: http://www.w3.org/Consortium/Overview.
[37] HL7. HL7 Reference Information Model. 2007.
[38] Dolin R, Alschuler, L., Boyer, S., Beebe, C., Behlen, F., Biron, P., Shabo, A. HL7 Clinical Document Architecture, Release 2. J Am Med Inform Assoc. 2006;13:30-9.
[39] American College of Radiology. ACR Homepage. [cited 2008 July 14]; Available from: http://www.acr.org
[40] National Electrical Manufacturers Association. NEMA Homepage. [cited 2008 July 14]; Available from: http://www.nema.org
[41] Japan Industries Association of Radiological Systems. JIRA Homepage. [cited 2008 July 14]; Available from: http://www.jira-net.or.jp/e/index.htm
[42] ANSI Healthcare Information Standards Board. ANSI-HISB page. [cited 2008 July 14]; Available from: http://ansi.org/hisb/
[43] Digital Imaging and Communications in Medicine. DICOM Strategic Document. [cited 2008 July 14]; Available from: http://medical.nema.org/dicom/geninfo/Strategy.pdf
[44] IEEE Standards Association. IEEE-SA Homepage. [cited 2008 July 14]; Available from: http://standards.ieee.org
[45] The Object Management Group. OMG Homepage. [cited 2008 June 30]; Available from: http://www.omg.org/.
[46] 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 April 30, 2004;69(84):24059-61.
[47] Office of the National Coordinator. Office of the National Coordinator: mission. 2008 [cited 2008 June 29]; Available from: http://www.hhs.gov/healthit/onc/mission/
[48] Office of the National Coordinator for Health Information Technology (ONC). The ONC-Coordinated Federal Health IT Strategic Plan: 2008-2012. . [cited 2008 June 30]; Available from: http://www.hhs.gov/healthit/resources/HITStrategicPlan.pdf
[49] Office of Management and Budget. eGov Inititative on Consolidated Health Informatics. [cited 2008 June 30]; Available from: http://www.whitehouse.gov/omb/egov/c-3-6-chi.html .
[50] Consolidated Health Informatics. CHI. [cited 2008 June 30]; Available from: http://www.hhs.gov/healthit/chi.html
[51] Richesson RL, Krischer, J. Data Standards in Clinical Research: Gaps, Overlaps, Challenges and Future Directions. Journal of the American Medical Informatics Association. 2007;14(6):687-96.

[52] American Health Information Community. AHIC Homepage. [cited 2008 June 30]; Available from: http://www.hhs.gov/healthit/ahic.html
[53] Office of the National Coordinator for Health Information Technology (ONC). Harmonized Use Case for Electronic Health Records (Laboratory Result Reporting), . [cited 2008 June 30]; Available from: http://www.ansi.org/standards_activities/standards_boards_panels/hisb/hitsp.aspx?menuid=3
[54] Healthcare Information Technology Standards Panel. Welcome to www.HITSP.org. 2008 [cited 2008 June 30]; Available from: http://www.hitsp.org/
[55] Healthcare Information and Management Systems Society. Healthcare Information and Management Systems Society 2008 [cited 2008 July 15]; Available from: http://www.himss.org/ASP/index.asp
[56] Certification Commission for Healthcare Information Technology. CCHIT Homepage. [cited 2008 July 14]; Available from: http://www.cchit.org
[57] American Health Information Management Association. American Health Information Management Association 2008 [cited 2008 July 15]; Available from: http://www.ahima.org/
[58] National Alliance for Health Information Technology. NAHIT Homepage. [cited 2008 July 14]; Available from: http://www.nahit.org
[59] American Academy of Family Physicians. AAFP Homepage. [cited 2008 July 14]; :[Available from: http://aafp.org
[60] American Academy of Pediatrics. AAP Homepage. [cited 2008 July 14]; Available from: http://www.aap.org
[61] American College of Physicians. ACP Homepage. [cited 2008 July 14]; Available from: http://www.acponline.org
[62] California Healthcare Foundation. CHCF Homepage. [cited 2008 July 14]; Available from: http://www.chcf.org
[63] Certification Commission for Healthcare Information Technology. CCHIT Ambulatory EHR 2007 Interoperability Criteria. [cited 2008 July 14]; Available from: http://www.cchit.org/files/Ambulatory_Domain/CCHIT_Ambulatory_INTEROPERABILITY_Critieria_2007__Final__16MAR07.pdf
[64] Integrating the Healthcare Enterprise. Welcome to Integrating the Healthcare Enterprise 2008 [cited 2008 July 15]; Available from: http://www.ihe.net/
[65] National Center for Research Resources. National Center for Research Resources. 2008 [cited 2008 July 15]; Available from: http://www.ncrr.nih.gov/
[66] Clinical and Translational Science Awards. CTSA Homepage. [cited 2008 July 14]; Available from: http://www.ctsaweb.org
[67] Clinical and Translational Science Awards. CTSA Consortium Oversight Committee Attachment B. January 15, 2008 [cited 2008 July 14]; Available from: http://www.ctsaweb.org/uploadedfiles/Attachment%20B_Informatics%20Breakout1.pdf
[68] US Health Information Knowledgebase. USHIK Homepage. [cited 2008 July 14]; Available from: http://dcg.dnsalias.net/registry/x/
[69] Agency for Healthcare Research and Quality. AHRQ Homepage. [cited 2008 July 14]; Available from: http://www.ahrq.gov/
[70] Centers for Medicare & Medicaid Services. CMS Homepage. [cited 2008 July 14]; Available from: http://www.cms.hhs.gov/
[71] Centers for Medicare & Medicaid Services. Medicaid IT Architecture Framework 2.0. [cited 2008 July 14]; Available from: http://www.cms.hhs.gov/MedicaidInfoTechArch/04_MITAFramework.asp
[72] National Center for Health Statistics. NCHS Homepage. [cited 2008 July 14]; Available from: http://www.cdc.gov/nchs/
[73] Libicki M, Brahmakulam, I. The Costs and Benefits of Moving to the ICD-10 Code Sets. *Rand Corporation Report* 2004.
[74] Public Health Information Network. PHIN Homepage. [cited 2008 July 14]; Available from: http://www.cdc.gov/PHIN/
[75] Public Health Information Network. PHIN Messaging Services Fact Sheet. [cited 2008 July 14]; Available from: http://www.cdc.gov/phin/library/documents/pdf/111759_PHINmessaging.pdf
[76] Public Health Information Network. PHIN XForms Question Framework Fact Sheet. [cited 2008 July 14]; Available from: http://www.cdc.gov/phin/library/documents/pdf/111759_Xform.pdf
[77] Public Health Information Network. PHIN Vocabulary Services Fact Sheet. [cited 2008 July 14]; Available from: http://www.cdc.gov/phin/library/documents/pdf/111759_vocabulary.pdf

[78] Public Health Data Standards Consortium. PHDSC Homepage. [cited 2008 July 14]; Available from: http://www.phdsc.org/
[79] National Uniform Billing Committee. NUBC Homepage. [cited 2008 July 14]; Available from: http://www.nubc.org
[80] National Uniform Claim Committee. NUCC Homepage. [cited 2008 July 14]; Available from: http://www.nucc.org
[81] Public Health Data Standards Consortium. PHDSC Health Care Service Data Reporting Guide Committee. [cited 2008 July 14]; Available from: http://www.phdsc.org/about/committees/hcs.htm
[82] Public Health Data Standards Consortium. PHDSC Payer Typology Committee. [cited 2008 July 14]; Available from: http://www.phdsc.org/about/committees/payer.htm
[83] Public Health Data Standards Consortium. PHDSC External Cause of Injury Codes Committee. [cited 2008 July 14]; Available from: http://www.phdsc.org/about/committees/ecic.htm
[84] Claredi. HIPAA Code Sets. [cited 2008 July 14]; Available from: https://www.claredi.com/hipaa/codesets.php
[85] ASTM International. ASTM E2369 - 05 Standard Specification for Continuity of Care Record (CCR). 2008 [cited 2008 July 12]; Available from: http://www.astm.org/Standards/E2369.htm
[86] IHE. PCC TF-1/XDS-MS. 2008 [cited 2008 July 16]; Available from: http://wiki.ihe.net/index.php?title=PCC_TF-1/XDS-MS
[87] Clinical Data Interchange Standards Consortium. CDISC Workgroups and Standards. [cited 2008 July 14]; Available from: http://www.cdisc.org/standards/index.html
[88] Clinical Data Interchange Standards Consortium. CDISC Technical Roadmap 2008-2010. [cited 2008 July 14]; Available from: http://www.cdisc.org/downloads/CDISC_Road_Map_Spring2008.pdf
[89] Health Level 7. RCRIM Homepage. [cited 2008 July 14]; Available from: http://www.hl7.org/Special/committees/rcrim/index.cfm
[90] Food and Drug Administration. CDISC - HL7 Project Document Post HL7 RCRIM Meetings 17-20 September 2007. [cited 2008 July 14]; Available from: http://www.fda.gov/oc/datacouncil/cdischl7projectdocument20071015.pdf
[91] Fridsma D, Evans, J., Hastak, S., Mead, C. The BRIDG Project: A Technical Report. J Am Med Inform Assoc. 2008;15:130-7.
[92] National Cancer Institute. NCI Homepage. [cited 2008 July 14]; Available from: http://www.cancer.gov/
[93] United States Food and Drug Administration. FDA Homepage. [cited 2008 July 14]; Available from: http://www.fda.gov/
[94] Ceusters W. How to build an Ontology? *Workshop on Clinical Trial Ontology*. NIH Campus, Bethesda, MD 2007.
[95] U.S. Department of Health & Human Services. Nationwide Health Information Network (NHIN): Background. 2008 [cited 2008 July 15]; Available from: http://www.hhs.gov/healthit/healthnetwork/background/
[96] Rishel W, Riehl, V., Blanton, C. Summary of the NHIN Prototype Architecture Contracts. *Gartner Report* 2007.
[97] Kohane I, Mandl, K. Final Report for the National Multi Protocol Ensemble for Self-Scaling Systems for Health (NMESH). 2007.
[98] McDonald C, Dexter, P., Schadow, G., Chueh, H., Abernathy, G., Hook, J., Blevins, L., Overhage, J., Berman, J. SPIN Query Tools for De-Identified Research on a Humongous Database. AMIA 2005 Annual Symposium; 2005; 2005. p. 515-9.
[99] cancer Biomedical Informatics Grid. caBIG Primer. [cited 2008 July 14]; Available from: https://cabig.nci.nih.gov/overview/cabig-primer/html/primer.html
[100] 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):1910-6.
[101] 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;41:106-23.
[102] Ceusters W, Smith, B. . A Terminological and Ontological Analysis of the NCI Thesaurus. . Methods of Information in Medicine 2005;44:498-507.
[103] American Medical Association. CPT Background and Categories of CPT Codes. [cited 2008 July 14]; Available from: http://www.ama-assn.org/ama/pub/category/12886.html

[104] 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):624-44.
[105] National Cancer Institute. NCI Terminology Browser. [cited 2008 July 14]; Available from: http://nciterms.nci.nih.gov/NCIBrowser/Dictionary.do
[106] de Coronado S, Haber, M., Sioutos, N., Tuttle, M., Wright, L. NCI Thesaurus: Using Science-Based Terminology to Integrate Cancer Research Results. In: Fieschi M, Coiera, E., Li, Y-C., editor. Medinfo; 2004: IOS Press; 2004. p. 33-7.
[107] National Library of Medicine. NLM Homepage. [cited 2008 July 14]; Available from: http://www.nlm.nih.gov/
[108] 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):1-11.
[109] Bodenreider O. The Unified Medical Langauge System (UMLS): integrating biomedical terminology. Nucleic Acids Research. 2004;32:267-70.
[110] College of American Pathologists. CAP Homepage. [cited 2008 June 30]; Available from: http://www.cap.org/apps/cap.portal.
[111] Spackman K, Campbell. K., Cote R. . SNOMED RT: a reference terminology for health care. . Proc AMIA Annu Fall Symp 1997:640-4.
[112] NHS Connecting for Health. Connecting for Health NHS Website. [cited 2006 March 6]; Available from: http://www.nhsia.nhs.uk/terms/pages/faq.asp?om=m1.
[113] PR Newswire. SNOMED International and the UK's National Health Service have begun alpha test of SNOMED Clinical Terms. [cited 2008 June 30]; Available from: http://www.prnewswire.co.uk/cgi/news/release?id=71229.
[114] PR Newswire. US healthcare industry moves toward the reality of electronic health records. [cited 2008 June 30]; Available from: http://www.pressreleases.be/script_UK/newsdetail.asp?nDays=m&ID=9455.
[115] 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):484-500.
[116] Strang N, Cucherat, M., Boissel, J. . Which coding system for therapeutic information in evidence-based medicine? Comput Methods Programs Biomed. 2002;68(1):73-85
[117] 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 2003:699-703.
[118] 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 2003:115-9.
[119] 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)):540-4.
[120] Chiang M, Casper, J., Cimino, J., Starren, J. Representation of ophthalmology concepts by electronic systems Adequacy of controlled medical terminologies. . Ophthalmology. 2005 February;112(2):175-83.
[121] 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;13:536-46.
[122] NHS. NHS Information Standards Board Chairman's report. 2004 Nov 23.
[123] 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 Feb;48(1-3):85-101.
[124] Bodenreider O, Burgun, A., Rindflesch, TC. . Assessing the consistency of a biomedical terminology through lexical knowledge. . Int J Med Inf 2002 67(1-3):85-95.
[125] Campbell J. Semantic features of an enterprise interface terminology for SNOMED RT. . Medinfo. 2001;10(Pt 1):82-5.
[126] 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. Medical Concept and Language Representation; 1999.; Phoenix; 1999. p. 70-81.
[127] Ceusters W, Smith, B., Kumar A., Dhaen C. Ontology-Based Error Detection in SNOMED-CT®. Proceedings of Medinfo; 2004; 2004.
[128] 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 10-14 May 2003.

[129] Ceusters W, Smith, B., Kumar A., Dhaen C. Mistakes in Medical Ontologies: Where Do They Come From and How Can They Be Detected? In: Pisanelli DM, ed. *Ontologies in Medicine Proceedings of the Workshop on Medical Ontologies, Rome October 2003*: IOS Press 2004.
[130] Bodenreider O, Smith, B., Kumar, A., Burgun, A. Investigating subsumption in DL-based terminologies: A case study in SNOMED CT. In: Hahn U, Schulz, S., Cornet, R., editor. Proceedings of the First International Workshop on Formal Biomedical Knowledge Representation (KR-MED 2004); 2004; 2004. p. 12-20.
[131] Bittner T, Smith, B. . Normalizing Medical Ontologies Using Basic Formal Ontology. Tagungsband der 49 Jahrestagung der Deutschen Gesellschaft für Medizinische Informatik, Biometrie und Epidemiologie. 2004.
[132] Rosse C, Mejino, J. . A reference ontology for biomedical informatics: the Foundational Model of Anatomy. . J Biomed Inform 2003;36(6):478-500.
[133] Hahn U, Schulz, S., Romacker, M. . Part-whole reasoning: a case study in medical ontology engineering. . IEEE Intelligent Systems and Their Applications 1999;vol 14 (nr 5):59-67.
[134] Smith B, Ceusters, W.,. An Ontology-Based Methodology for the Migration of Biomedical Terminologies to Electronic Health Records. AMIA 2005; 2005 October 22-26; Washington, DC; 2005. p. 669-73.
[135] PITAC Letter. PITAC Letter. [cited 2008 June 30]; Available from: http://www.nitrd.gov/pitac/reports/20040604_aha.pdf.
[136] Guarino N. Formal Ontologies and Information Systems. In: Guarino N, editor. FOIS98; 1998: IOS Press; 1998. p. 3-15.
[137] Baader F, McGuinness, D., Nardi, D. , ed. The Description Logic Handbook. : Cambridge University Press 2003.
[138] Farrugia J. Model-Theoretic Semantics for the Web. 12th International Conference on the WWW 2003: ACM; 2003. p. 29-38.
[139] Ceusters W, Smith, B. Strategies for referent tracking in electronic health records. Journal Biomedical Informatics. 2006;39(3):362-78.
[140] Kumar A, Smith, B. . The Unified Medical Language System and the Gene Ontology. KI2003: Advances in Artificial Intelligence 2003(Lecture Notes in Artificial Intelligence 2821):135–48.
[141] Smith B, Williams, J., Schulze-Kremer, S. . The ontology of the gene ontology. In: Musen M, editor. AMIA 2003 Annual Symposium; 2003; Washington D.C.; 2003. p. 609-13.
[142] Grenon P, Smith, B., Goldberg, L. Biodynamic Ontology: Applying BFO in the Biomedical Domain. In: Pisanelli DM, ed. *Ontologies in Medicine*. Amsterdam: IOS Press 2004:20-38.
[143] Smith B, Rosse, C. . The role of foundational relations in the alignment of biomedical ontologies. . In: Fieschi M, Coiera, E., Li, Y-C. , editor. MEDINFO 2004 Proceedings of the 11th World Congress on Medical Informatics; 2004; San Francisco, CA: IOS Press; 2004. p. 444-8.
[144] Kumar A, Schulze-Kremer, S., Smith, B. . Revising the UMLS Semantic Network. . In: Fieschi M, Coiera, E., Li, Y-C., editor. MEDINFO 2004 Proceedings of the 11th World Congress on Medical Informatics; 2004; San Francisco, CA: IOS Press; 2004. p. 1700-4.
[145] BSR/IEEE 1953-200x. BSR/IEEE 1953-200x. [cited 2008 June 30]; Available from: http://www.nssn.org/Search/DetailResults.asp?docid=43517&selnode=.
[146] BSR/IEEE 1953.1-200x. BSR/IEEE 1953.1-200x. [cited 2008 June 30]; Available from: http://www.nssn.org/Search/DetailResults.asp?docid=43930&selnode=.
[147] Sequence Ontology Project. The Sequence Ontology Homepage. [cited 2008 June 30]; Available from: http://song.sourceforge.net/.
[148] Gene Ontology Project. The Gene Ontology Homepage. [cited 2008 June 30]; Available from: http://www.geneontology.org/.
[149] Microarray and Gene Expression Data Society. MGED Homepage. [cited 2008 June 30]; Available from: http://www.mged.org/.
[150] Open Biomedical Ontologies Project. OBO Homepage. [cited 2008 June 30]; Available from: http://obo.sourceforge.net/.
[151] 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):R46.
[152] International Organization for Standardization Technical Committee 37. ISO TC 37. [cited 2008 June 30]; Available from: http://www.iso.org/iso/iso_technical_committee.html?commid=48104.
[153] International Federation of Library Associations and Institutions. The IFLA Homepage. [cited 2008 June 30]; Available from: http://www.ifla.org/III/index.htm.

[154] International Federation of Library Associations and Institutions. Functional Requirements for Bibliographic Records. [cited 2008 June 30]; Available from: http://www.ifla.org/VII/s13/frbr/frbr.pdf.
[155] Dublincore Metadata Initiative. The Dublincore Homepage. [cited 2008 June 30]; Available from: http://dublincore.org/.
[156] Rust G, Bide, M. . The INDECS Metadata Model. 1999 [cited 2008 June 30]; Available from: http://www.doi.org/topics/indecs/indecs_framework_2000.pdf
[157] Committee on Documentation of the International Council of Museums. CIDOC Conceptual Reference Model. [cited 2008 June 30]; Available from: http://cidoc.ics.forth.gr/.
[158] Consortium for Interchange of Museum Information. CIMI Homepage. [cited 2008 June 30]; Available from: http://xml.coverpages.org/cimi.html.
[159] Lagoze C, Hunter, J. . The ABC Ontology and Model. Journal of Digital Information. Journal of Digital Information. 2001;2(2 Artical Number 77).
[160] Smith B, Ceusters, W., Temmerman, R.,. Wusteria. In: Engelbrecht R. ea, ed. *Medical Informatics Europe*. Amsterdam: IOS Press 2005:647-52.
[161] Standard Upper Ontology Working Group. SUO Homepage. [cited 2008 June 30]; Available from: http://suo.ieee.org/index.html.
[162] Foundation for Intelligent Physical Agents. FIPA Homepage. [cited 2008 June 30]; Available from: http://www.fipa.org/.
[163] OWL Working Group. OWL WG Homepage. [cited 2008 July 14]; Available from: http://www.w3.org/2007/OWL/wiki/OWL_Working_Group
[164] Object Management Group. OMG Homepage. [cited 2008 June 30]; Available from: http://www.omg.org/.
[165] Organization for the Advancement of Structured Information Standards. OASIS Homepage. [cited 2008 June 30]; Available from: http://www.oasis-open.org/home/index.php.
[166] Organization for the Advancement of Structured Information Standards. Darwin Information Typing Architecture. [cited 2008 June 30]; Available from: http://www.oasis-open.org/committees/tc_home.php?wg_abbrev=dita.
[167] Richesson R, Krischer, J. Table 1. Broad Constructs and Current U.S. Standards for Clinical Research Data, Including Overlaps and Gaps. 2007 [cited 2008 July 14]; Available from: http://www.jamia.org/cgi/data/M2470/DC1/1