The Nature of Heterogeneity in the Context of Distributed Health-Care Information Systems

James G. Nell

NIST

The Nature of Heterogeneity in the Context of Distributed Health-Care Information Systems

James G. Nell
Manufacturing Systems Integration Division
Manufacturing Engineering Laboratory

June 2003

Abstract
This paper offers a discussion of the types of heterogeneity that can occur among processes that share information. The author attempts to establish areas of similarity and difference between the airline, financial and manufacturing industries and the health-care industry. Ways by which information can be shared are discussed, given this heterogeneity, and some recommendations for system design are offered.

The Nature of Heterogeneity
In the Context of Distributed Health-Care-Information Systems

Preface

The Manufacturing Systems Integration Division, MSID, at the National Institute of Standards and Technology, NIST, has concentrated on improving the integration of information systems. A special emphasis has been for large manufacturers in such industries as aerospace and automotive; and for small manufacturers such as those typically involved with producing printed-wiring assemblies for electronic products. MSID has been active in several ways to help the manufacturing industry cope with information-integration challenges. These include:

- Developing and promoting several information-technology standards
- Helping to ascertain industry needs through workshops
- Analyzing the nature of the information-integration problem
- Recommending solutions
- Providing test beds to verify ideas

The health-care industry also has been struggling with the task of improving the information systems that link patients to the various types of providers, to payers, and to vendors of information systems and health-care equipment. The Advanced Technology Program, ATP, at NIST, has been soliciting proposals for new technology to improve information transfer in this domain for several years. MSID has supported this ATP work by reviewing proposals, the work, and results; and by preparing papers to familiarize those in ATP administering the ATP health-care-informatics[1] projects.

As MSID was commencing the health-care work, MSID was sensitive to the fact that each industry is unique with respect to integrating the information that supports that industry. For example, MSID recognized early that the integration that was apparent in the banking and airline industries is, by far, not as complex as the information most manufacturing entities manipulate. To become versed in the particulars of health-care-information needs, MSID reviewed hundreds of relevant readings on the topic[2] and conducted several workshops in which the health-care industry information-technology experts expressed their needs.

Many of these needs are similar to those for manufacturing-enterprise-integration, especially since they relate to small-to-medium-sized enterprises, a business model that exists in most of the health-care industry. At the core of the health-care-informatics problem are the same large challenges that manufacturing has: representing knowledge and semantics and resolving terminology ambiguities. However, the nature of the information to be represented is different, and improving the information integration poses some unique challenges.

This paper presents the work that the author undertook to establish a basis for analyzing information problems of the health-care-industry entities. The core of the work focuses on knowledge about the information-interoperability aspect of enterprise-process that has been accumulated while analyzing manufacturing enterprises. The author indicates where some parallels exist between the informatics of health care and manufacturing, and points out where key differences are. A large part of the paper is devoted to explaining what makes solving the terminology problem so difficult, and how the health-care terminology is different from financial, airline, and manufacturing terminology.

---

[1] The word "informatics" is used by the health-care industry to describe its computer-based-information transfer domain.

[2] See the introductory sentences preceeding the Bibliography, Section 9.

1. Summary

Day-to-day commerce in the health-care industry encounters a problem in which users have a limited ability to exchange information electronically and efficiently. The industry and its suppliers feel a primary cause of this problem is the heterogeneous nature of the information and the information systems that are used. This heterogeneity can consist of differences in hardware, software, communication protocols, and information meaning and format. The author investigated the nature of information, communication, integration, and heterogeneity, and this paper presents the results of that investigation.

Information systems are difficult to integrate and communication among systems is often ambiguous. Systems with dissimilar or heterogeneous elements are even more difficult to integrate. Heterogeneity divides into two parts: incompatibility of physical things and failure to transfer meaning. The prime cause of heterogeneity is that designers of a sending system will have chosen one or more physical and/or semantic elements that are different from, or otherwise incompatible with, those comprising the receiving system.

The physical part--hardware, software, and communication protocols--is the easier part to define, partition, and resolve. Many domains have become agile enough to solve some physical-heterogeneity problems as they are encountered, and standards have resolved other problems.

However, information system designers have found it quite difficult to be agile with respect to the meaning of information transferred--the semantics. The semantic part of heterogeneity is interpreting symbols into mental perceptions. Many problems arise when analysts who observe real-world phenomena, form mental perceptions of what they observe, and record what they observe in text or in some other sort of process model. For communication to be successful, subsequent reviewers of the recorded symbols must form an identical mental concept as the original observer. This means that the original observer must first record the observations correctly, and subsequent reviewers must interpret and act on those symbols exactly as the original observer intended.

Ambiguities are almost impossible to avoid in the natural-language-creation process. Misunderstanding occurs in human conversation even when two closely related people converse about something with which both are familiar. Devising machines and software to obviate miscommunication may be impossible, and humans may always have to be involved to resolve ambiguities. Certainly, however, good information-system design may enable great improvements. Ontologies may offer a way to resolve terminology problems during system operation.

The recommendations expressed in Section 8 frame an approach to mitigating the problem and should indicate a few things the health-care industry can do so that technical advances and standards can be applied to the virtual domain where applicable.

2. The Nature of the Problem

2.1 Enterprises, Products, Processes, and Integration

When discussing information integration it is often convenient to include the concept of enterprise. This provides a flexible way to refer to the domain of discourse and to define the things that are inside the enterprise that can be controlled, and the things that are outside that are more difficult, if not impossible, to control.

For this paper, assume that an enterprise is a group of processes that produces a product or service; to wit, a useful output [1, 2]. An enterprise can be of any size from a multi-corporation endeavor to a single process that may occur in a single-person office. A process is a group of coordinated activities that results in an output. Therefore, an enterprise can be any group of processes of interest at any given time. Most processes have a supplier and all have a customer to whom an output goes. In a virtual-distributed enterprise the enterprise elements are not co-located and the communication paths are managed and used

by computers, often using the Internet. Relationships between these elements are frequently temporal and not formal.

As stated, an enterprise can be any set of related processes we wish to analyze. However, each enterprise has a character and there is some implicit or explicit vision that drives it. Processes and tools help an enterprise accomplish its outputs. Who buys the processes and tools; that is, who selects them, determines if one is better than the other, and decides to invest the money or to wait? How do the tools and processes communicate? How does the enterprise acquire, move, and ship its information and physical things? Normally, there is an executive function that decides these things. The executive function is not necessarily a person at the executive level but a function. This function receives its inputs and constraints from some entity that evaluates enterprise goals and strategies and allocates resources that implement the strategies in an attempt to meet the goals. These functions must operate even more smoothly in a distributed-virtual industry such as health care because there likely is not one set of goals and strategies as there may be in a non-distributed, manufacturing enterprise that is highly dictatorial in nature.

2.2 The Health-Care Enterprise
An evolution of technology has improved the capability for enormous flexibility in the processes that make up the health-care enterprise. For health care, the enterprise will comprise some combination of patients, payers (government, employers, and insurance companies), providers (hospitals, clinics, laboratories, first responders, clinicians, and physicians), vendors (medical equipment and software), and suppliers pharmaceuticals and medical supplies) are those processes that comprise the industry. The industry is sizable; in the United States, there are 3 thousand payers, 7 thousand hospitals, 700 thousand physicians, and 248 thousand groups of physicians. [3] The goal is to be able to reconfigure the processes that provide health-care products and services by improving the information systems, re-assigning resources, and/or changing responsibilities of some people.

As this evolutionary trend has continued, we have encountered a new health-care paradigm, where the term paradigm refers to the rules of the market area. The health-care participants are involved in the same processes, caring for patients' health, however the rules of the market have changed. Pressures to apply newly available capability have caused changes that have been more evolutionary than revolutionary in nature. The trend is toward more information represented in computers and less in physicians' paper files. To make computer use more pervasive has been a challenge. Some challenges are listed below.
- Reduce costs in the processes, especially the administrative costs
- Improve patient safety by reducing errors
- Easier-to-use computer interfaces and data entry
- Converge the terminology of the
  - Largely subjective process of diagnosis description
  - Patient-order system
  - Electronic-medical-records system
  - Medical-information and knowledge bases

The capability improvement has been toward more flexibility and agility. If improvements are well designed, as capability extends, agility and flexibility increases, and efficiency improves. As the state-of-the-art parts of the health-care enterprise invest in that capability, the services and products that these improved enterprises offer can be more customized, with higher quality, and with lower cost. Later, with more improvements to the information system and to the administrative system, the health-care industry can actually operate in the mode in which it is configured; that is, many small providers providing service to patients, having access to more information quickly, and buying medical devices from suppliers. All of this can be done with such agility that there is no advantage to being a large entity or to request more information than one needs.

In the aerospace industry, mostly large manufacturing enterprises, the U.S. Department of Defense has been a key driver toward integrating the enterprise information. DOD did this to save acquisition and support costs of defense-related products. Smaller enterprises can benefit from this trend because the larger manufacturing industries have developed an agility to append a few processes for a short time for a specific

product variation. Often these appendages are small entities that otherwise would not be participating. The virtual-industry concept has become common, and it is applicable to the health-care industry by the very nature of the industry.

A virtual industry is a widely distributed set of independent points of use, or nodes, that operate in a self-organizing mode. The virtual industry does not operate in a strong, tiered hierarchy as do some manufacturing enterprises. It operates more as a taxi system in a city where the nodes, independent taxi drivers and, perhaps, a dispatcher, are highly autonomous. Recognizing this structure, how it operates, and the nature of the transactions among nodes, will help to design a system that improves information access and transfer.

In a system where there is a strong, top-down flow of commands, to prescribe integration successfully would seem to be an easy task. However, the more the humans of traditional enterprises realize that information is a valuable asset, the more that enterprise finds that the people in that enterprise are reluctant to share the information, because it gives them power and advantage over their peers. Contrast that with the self-organizing operations, where the participants may see their power enhanced if they share information because the potential to access all nodes directly in a flat and autonomous system is improved and, thereby, there is potential access to more business.

Assuming that virtual-industry enterprises comprise a widely distributed set of different sized and independent operators or nodes, the need is that these nodes share information in a more efficient way than they normally do. That is, the data systems should be logically centralized and physically distributed. They should be automatic, reliable, fast, accurate, upgrade-able, and extendable. They should use spoken-language entry where feasible; use knowledge bases and expert-system access where practical; and, handle text, video, audio, graphics, and photographs.

To accomplish these things, each sector of the industry must invest in system improvements. With self-organizing enterprises there is no hierarchy of control, no one in charge, and no central enterprise that could dictate, evaluate, approve, and justify investments. There is also no single organization to pay the costs of integration--although government has invested in infrastructure improvement in the past. Each node must help itself, usually by minimizing its own costs. Doing this can increase costs for others outside the node that must inter-operate with that node. There is no incentive to invest for the good of the entire health-care system. These selfish investment and cost minimizing patterns are rampant in traditional manufacturing enterprises. That is why in a traditional industry there are many point solutions that are proprietary and difficult to integrate.

Where a node is a small point in a large enterprise such as a hospital, the node could be the whole enterprise for a small-to-medium-sized enterprise such as a clinic or physician's office. Therefore, investments made to improve the information-sharing capability for a small-to-medium-sized, health-care node allow that node to inter-operate better with the health-care system as a whole. The need for that small enterprise to remain competitive will tend to force a balance between investing in improved capability and the ability to absorb the costs of that capability.

Any planned information-infrastructure improvement that gives the health-care industry more characteristics of a virtual enterprise must consider realistic ways to include the small nodes and their point solutions in the overall resultant environment. For an environment of many independent nodes it appears that the self-organizing approach to information-system design is better suited than the hierarchical approach.

The proprietors of the small virtual nodes probably procured or designed their own systems with the help of their local computer-store expertise. Logically, the systems were selected because they are cost effective and meet the purchaser's specific needs. The small size of the enterprises at these nodes limits severely the amount of additional standards-based software, hardware, and infrastructure that it can or would choose to buy. In fact, a node may decide, as it often does in manufacturing scenarios, to spend not one extra penny to procure a system adhering to standards that, potentially, could benefit the enterprise. These cost-

minimizing tendencies will continue until node managers are rewarded to work to the benefit of the enterprise rather than being rewarded for meeting cost-reduction goals of their node only. For a very small enterprise, these rewards could be in the form of increased access to its customers and to information sources.

How does a health-care enterprise compare with enterprises in other fields of endeavor such as manufacturing, airlines, and finance? For one thing, there is no notion of a controlling hierarchy in a network of very small nodes that exists in the health-care industry. In finance, airline, and manufacturing enterprises there is a notion of a lead firm, a customer, or a program manager to fulfill that role. Another difference is the nature of the information. Early integration of some kinds of information has occurred with airline scheduling systems such as BABS, Apollo, and SABRE.[3] Another early integration occurred in the banking industry with electronic transfers and automated-teller machines. In each, the information is unambiguous numerical quantities. Manufacturing information has taken longer because it is much more complex. Product data is represented by the Standard for the Exchange of Product Model Data, ISO 10303 [75]. Parts of STEP, which has been developed over the past 20 years, are being implemented. STEP consists mostly of computer-aided-system dimensions, tolerances, graphics, and other quantifiable information, and the representation of that information must be exact and unambiguous or the products cannot be manufactured.

In the virtual health-care industry much of the data is more diverse in subject matter, more judgmental, less specific, and more widely distributed. In fact, most of it may be in handwritten form in folders in each proprietor's office. A smaller amount may be entered into stand-alone computers; that is, in computers not connected to the Internet. The entries in some node's files are not likely to be in language that is precise enough to be semantically consistent within the node itself, let alone consistent with the files of another node. The current system works because there are humans interpreting the context and relevancy of the information. Indeed, often interpretation and use is by the human who entered the information in the first place. Therefore, the integration of this information presents significantly more challenge than other industries. The need to integrate, and the benefit of integration, is not any less, however.

The table on the following page summarizes the differences among enterprises in these three domains. The information is intended to point out examples of the differences in the information of the different and to indicate that the job of more fully integrating the health-care information will present challenges that are not present in other industries.

3. The Nature of Information Integration and Heterogeneity

The topics of integration, heterogeneity, semantics, and terminology are valid for any kind of enterprise. However, the nature of the information to be processed is important; therefore, the discussion also covers uniqueness of health-care information.

Integration refers to the degree of enterprise operation in which all necessary things, processes and infrastructure, are in place that enable the correct activity behavior, at the correct time, every time. Information flow is at the heart of successful integration. Information integration refers to integrating the information aspects of the process, perhaps humans, sensors, and/or computer-software applications. When enterprise-element activities are computerized, the tools used to accomplish much of the information work are called applications. This information flow is between the information aspect of one application to the information aspect of another application. The path could be sensor to application, application to application through the communication infrastructure, and between any of these and humans.

Of course, many other things must happen both technically and culturally to allow satisfactory information flow. Proper activity behavior and information flow have existed within and among processes for decades--

---

[3] BABS (British Airways Booking System, now Amadeus), SABRE (Semi-Automatic Business Research Environment formerly American Airlines and IBM), and Apollo (Formerly United Airlines now Galileo International) are huge, integrated, airline reservation systems. Apollo also had maintenance schedules and history for United's Aircraft. See www.amadeusuk.com, www.sabre.com, and www.apollotravel.com

or else enterprises would have produced no products or services. What enterprise integrators are trying to do is to make the information flow repeatable, consistently accurate, and computer sensible.

Table: The Nature of Enterprises in Different Domains

Characteristic	Financial and Scheduling	Manufacturing	Qualitative
Example	Bank or airline	Aerospace team	Health-care entity
Enterprise Architecture	Hierarchical	Hierarchical	Distributed network
Enterprise Type	Intra company (branches)	Extended enterprise (Controlled by lead firm)	Virtual enterprise (Autonomous network)
Nature of Data	Quantitative: centralized and numerical	Quantitative: Distributed, numerical and graphical	Quantitative and qualitative: Widely distributed, numerical, graphics, photographic
Major Issues	Consistency	File-sharing, machine interpretation of product representations	Machine interpretation of qualitative terminology
Privacy Issues	Personal and corporate	Industrial espionage	Personal; fairness of employment and insurance opportunity
Integration Goals	Cost reduction, electronic transfers and reservations	Reduce product cost and lead time, improve accuracy, reduce rework	Reduce errors and health-care costs, better diagnostics
Typical Node Hardware	Main frame; Database; Automated-teller machines	Computer-aided engineering, design, and manufacturing; database	Main frame (payers); PCs (physicians); CAE. CAD, CAM for equipment makers

3.1 Integration
To improve the level of integration, certain improvements to technology are necessary. The categories of integration technology divide into infrastructure and process technologies. Infrastructure improvements benefit more than one process in the enterprise, such as a change from twisted-pair telephone conductor to fiber-optic cable, whereas a process improvement focuses on improving one process, such as upgrading a billing-system to make the information more compatible with other financial or accounting information. If one considers the possible infrastructure and process improvements it is not likely that any enterprise will have invested to the maximum in all possible areas. Usually, however, the enterprise still operates and it may produce quality products or services but at a reduced level of efficiency. To produce any product or service, integration must occur, but the nature of that integration varies from a temporal and not easily repeatable information exchange to an automatic system requiring little or no human interpretation.

Integration is not a monolithic thing or a state that exists suddenly with a specific amount of process and/or infrastructure improvement [1, 4]. Integration is a process that, hopefully, continually improves information flow. It would be more accurate to speak of improving the degree of integration rather than achieving integration. Each enterprise is unique with respect to determining how to improve. Here is where the enterprise character, goals, and strategies come in. The enterprise-executive function selects the level of investment that is appropriate for each process and infrastructure area. These improvement areas appear as many continua of increasing opportunity, functionality, and capability rather than a specific hurdle of capability; and, each enterprise is free to position itself at a particular spot on each continuum. Enterprises base this placement decision on the strategies of the enterprise, the amount of money available for investment, and any other criterion important to the enterprise. The problem is compounded because the technology level of each continuum is continually advancing. Continued investment is necessary to maintain the same relative capability in each continuum.

Each enterprise will have a self-determined priority for which projects will receive investment and where on each continuum the enterprise will operate. Some enterprises will want to be on the leading edge and others may choose to follow the leaders, perhaps to enable them to invest in these technologies later at a lower cost. Neither strategy is incorrect, but these decisions lead to unevenly equipped enterprises that are trying to do business in their marketplace. With well-planned enterprise engineering and standards, these unevenly equipped enterprises still would be able to interoperate to the degree that their investments permit. A low level of integration capability will not have all of the functionality of the high level. This is similar to a personal-computer user who chooses not to invest in the latest integrated office suite, instead, choosing to remain with separate applications that do not share information as well. However, the individual applications could share some information with other enterprises, but at a lower level of functionality.

3.2 The Nature of Information Integration [5 through 8]

Information integration is a goal often expressed by those who want to use, and understand fully, someone else's information. A situation without ambiguity implies that the concept, or meaning, as interpreted by the receiver is exactly as intended by the originator. The process to accomplish this is complex and multidimensional. For a perfect information exchange, the receiver must be able to receive and understand the correct amount of the correct information, at the correct time, every time. This is an ambitious goal. Indeed, humans have not been able to achieve perfect understanding when those who possess the information confront those requesting the information in the same room in a face-to-face meeting. For this paper, a successful communication requires that senders and receivers understand exactly the same concept.

Information is not matter or energy, but it exists among the phenomena of the real world in the order or design of objects whether or not it is perceived or understood.[4] Because the meaning of the information exists only in the minds of the sender and the receiver, the meaning of information is only as perceived by the sender and the receiver. The perceptions differ quite often and clarification becomes necessary because one or both of the two parties inaccurately perceive a phenomenon and form either an imperfect concept due to unintentional differences in abstracting, or they insert an incorrect semantic representation of the concept (syntax), or both. To effect a communication, there must be a sending repository, a medium, a receiving repository, a syntax that is known to both parties, and the information encoded in some kind of semantic-representation package compatible with the syntax.

So if the information is there but the representation of it can vary, extreme care is necessary to assure that the syntax is correct and consistent. Then, the terminology selection is crucial and the meaning of the terminology must be exactly the same to the sender and receiver--neither more nor less. In other words the terminology must denote the meaning, the whole meaning, and nothing but the meaning. The meaning is important for one reason: to effect the desired behavior from the activity in question.[5] A human can perform behavior by using the meaning to "understand" the information and behave accordingly. Or, a machine can, through programming but without "understanding", recognize the syntax and behave per the program. Either way, the terminology must be selected with precision.

3.3 Heterogeneity [5]

Integration becomes more of a problem when heterogeneity exists somewhere in the system. Heterogeneity is a state in which components are not unified or compatible. Heterogeneous systems are those that contain dissimilar ingredients. There are two kinds: physical-system heterogeneity and semantic heterogeneity. Both kinds of heterogeneity are consequences of the design autonomy present in the process of developing

---

[4] T. Stonier in *Information and the Internal Structure of the Universe*, Springer Verlag, 1991, posits that the order in a real-world phenomenon (a measure of its entropy) is directly related to its information content--the more order, the less entropy, the higher the information content. By its very nature, order connotes information content. A phenomenon contains information whether or not it is perceived or understood. Therefore, there is a measurable equivalence between information and the entropy of the universe.

[5] "Semantics does not deal with the meaning of words. The true meaning of a term is to be found by observing what a person does with it, not by what the person says about it." Attributed to P.W. Bridgeman, physicist, in the essay: *How Words Change our Lives*, S.I. Hayakawa, Adventures of the Mind, Alfred A. Knopf, 1961, Edited by Richard Thruelsen and John Kobler.

information resources--especially software and databases. The decisions made during the design of one database most certainly will be different from the decisions made during the independent design of another database. Often the differences are there because the objective is to design the best application to suit the purpose, and some components or approaches offer advantages over others in specific situations. This is a good thing for an enterprise. Often, however, when designing applications, lack of financial incentive to take the enterprise view, available funding, technological obsolescence, and schedule pressures are other reasons for design decisions leading to heterogeneity. These are bad things for information interoperability. On the other hand, there exists no top-down, universal *best set of components* that would tend to reduce heterogeneity. Unfortunately, establishing or requiring this best set of components will be a strong factor in curtailing technological progress.

3.4 Physical Heterogeneity This situation includes things such as differences in hardware, software, operating systems, database-management systems, communication protocols, data models, data-language transaction management, and media. Physical heterogeneity has been studied for years, is well understood, and for the most part, resolvable. Physical heterogeneity may not exist at an application because that is the only type of heterogeneity that is understood and resolved by management. The physical-component of heterogeneity may be legislated out of existence by efforts to standardize on hardware, software, or operating system. This may work well for systems where human interpretation of the data is necessary, such as with word processing.

If computer systems and networks are the basis of the communication medium, as is the case between many processes of commerce, there will be agreements at several levels before an actual communication is attempted. The agreements will consist of protocols and interfaces in the form of formal standards and ad-hoc arrangements that establish a communication channel every time a communication is desired. After this is done the sender and the receiver must interact to agree about the hardware, the software, the protocols to accomplish the communication transfer process, the information format, and the meaning of the elements of information. These elements must be resolved to achieve any electronic communication.

3.5 Semantics of Computers
What semantics do computers store? None, they store information. The so-called semantics of database objects are the concepts elicited by the objects represented in the database. However, the concept is retained only in the human mind. The objects stored in a database are the combinations of signs that humans place there. When computers transfer "semantics" or "meaning" from one process to another "without human interpretation", it actually is a set of signs that one computer transfers to another. The receiving process then performs an algorithmic response with no concept being formed. Only humans using computers understand what the computer's signs mean, whether the concepts represented by the signs in two separate machines are identical, similar, or unrelated. Therefore, while a computer can compare sets of bits, only humans will know which comparisons are appropriate at any given time and what makes each set of bits significant. Herein lies the foundation of semantic heterogeneity.

3.6 Semantic heterogeneity [5, 6, 8, 9, 10]
After resolving physical heterogeneity problems, a very large component of the heterogeneity problem will remain unsolved for computer-sensible applications because system designers failed to recognize and resolve semantic heterogeneity. Solving problems relating to semantic heterogeneity is much more challenging than physical heterogeneity. In fact, it could be true that database designers, or human communication, will never completely resolve semantic heterogeneity.

Semantic heterogeneity is concerned with signs, symbols, and mental perceptions. Problems arise when analysts observe real-world phenomena, form mental concepts of what they observe, and convert what they observe into symbol form, into text, or into graphics. The next step often is to assign another person to interpret those symbols and to prepare a document, a drawing, or to design a database. These repositories store these perceptions in another set of symbols. For example, a modeler's mental perceptions are converted, by using symbols, into a model. A database designer forms a perception of what that model supposedly represents and translates that perception into database symbols. Therefore, what is represented

eventually in the database is at least two steps, and two concepts, and two sets of symbols removed from what was observed in the real world.

Furthermore, the probability that a second person, observing the same real-world phenomenon as the first observer will perceive exactly the same as the first observer can be quite low. Further, even if it were the same, the probability that a second database analyst will then represent those perceptions in the database exactly the same is just as remote. Of course the probability of both steps being the same is (low) x (low) = (extremely low). Therefore, we have two different database representations of the same real-world phenomenon. Integrating these two representations can range from tedious to nearly impossible.

Actually, the characters in a database are signs that comprise symbols representing what is really out there. Aristotle talked about symbols having two aspects: the sign and the concept that the symbol represents. The sign is a physical manifestation or thing that is used to refer to the symbol. The concept represented by the symbol exists in the mind of the person who has perceived the signs and thereby recognized the symbol. Another person, upon seeing the signs and recognizing the symbol, also forms an associated concept in his mind. The concept is usually an imagination of a real-world phenomenon. If the two persons imagine different concepts, one or both representations are ambiguous.

Simple signs can be combined to form complex sets of signs that represent complex symbols. The rules by which simple signs can be combined into more complex signs are called syntax. For a language, the syntax refers to the rules by which words are combined to form clauses, phrases, and complete thoughts such as sentences.

The meaning of the symbols that these groups of signs connote and denote are the semantics. In other words, the semantics are the concepts that the mind invokes upon recognizing a symbol based on signs presented to the senses.

How complicated is the semantic heterogeneity? There are many ways that the data in one system can be (semantically) unintelligible to another data system. Semantic heterogeneity has three categories. Reference [5] details the classification of semantic heterogeneity. For example, heterogeneity between object classes consists of differences in extensions (syntactic structure), names (synonyms, homonyms), attributes and methods (present, absent, temporal, -arities, defaults, constraints), domains (semantic and syntactic), and constraints (based on perception and need). Heterogeneity between class structures consists of inconsistencies of level (generalization/specialization and aggregation/decomposition), and discrepancies in the data structure (meta level in one database is data level in another). Heterogeneity between object instances consists of presence/absence discrepancies, multi-valued attribute (discrepancies in number of values used), discrepancies among attributes allowing nulls (null/non-null), and value and accuracy discrepancies.

The heterogeneity discussed above is mostly with regard to database design. The heterogeneity between object classes, synonyms, is one of the most elusive types of heterogeneity. This topic is discussed in more detail in Section 3.7.

To communicate successfully in any kind of enterprise, subsequent reviewers of recorded symbols must represent an identical mental concept as the initial reviewer. If we expect machines to form these bridges, or to act on what a human ordinarily would interpret, we will require formally modeled terminology standards; languages that are very rich and, at the same time, formal; and, a high level of diligence at data-entry points.

3.7 Terminology [9, 11 through 18]
Terminology is the most serious problem to be overcome to improve significantly information integration in medical-informatics applications. Information-system analysts in other product domains, such as manufacturing, are concluding that an enterprise-wide architecture, containing a standard ontology [76] and associated methodologies and tools, will allow automatic evaluation of terminology variants. The

ontology must contain formally modeled medical concepts in knowledge-representation formats, to enable terminology users to build semantic-consistency checks into their systems.

Why, then, is terminology such a large problem? To a large degree, it is because of human nature, organizational legacy, and in a health-care-informatic system, the nature of the medical information itself.

Human Nature: Human nature contributes to the terminology challenge during the language-creation process. When creating language from mental concepts, humans select more often the first word that will fit the need of the moment. Further, our schools have taught us, to make our utterances and writing more interesting, to not repeat words when a synonym is available. Since the English language is rich with words that are not exactly congruent synonyms, since humans will continue to be the data enterers, and since the medical-informatics data is based upon human judgement, the problem cannot be legislated out of existence by a forced system of input-word choices.

Organizational Legacy: Organizations name their processes a certain way and these names tend to persist over time. In enterprises, some processes and human resources are often reorganized and, as a consequence, some functions are reassigned or combined into different processes. Often the result is a mismatch among different enterprises of what activities are included in any given function, and what a process accomplishes. In addition, functions, processes, and activities with the same name in different enterprises may contain different activities. A standard way to query these functions and establish a working ontology is necessary to improve the probability that a message will arrive at the intended destination and result in the desired behavior.

Uniqueness of Medical Information: As discussed on page 6, compared to banking and manufacturing, medical information is more complicated, it tends to be more judgmental, and much of it defies any attempt to measure it. For example, conditions based on patient feelings, such as pain or depression, are difficult to quantify, qualify, represent, or integrate with other similar data.

Although most knowledgeable health-care- and manufacturing-enterprise integrators agree that terminology is a problem, the first attempt at solution has been to create some sort of naming dictionary or categorizing scheme. Unless extreme care is exercised to be absolutely sure that two instances of the same term contain the same meaning, the whole meaning, and only that meaning, ambiguity of terminology will continue. In other words, two concepts are the same if, and only if, a logical representation of the word describing one concept is congruent with a logical representation of the word describing a second concept. Formal terminology modeling methodologies that are machine processible are appropriate for computer-operated ontologies. Carelessly assembled naming conventions will merely push the ambiguity in terminology somewhere else without solving this difficult problem.

Below are some naming efforts, repositories, and languages being developed in the medical-informatics community.
GALEN [3, 19 through 24] (General Architecture for Languages, Encyclopaedias, and Nomenclatures in Medicine): A European-Community Program that has created a generalized and application-independent architecture for medical terminology. There also is a common reference model of medicine.
HL7 [16, 25, 26, 27] (Health Level 7): An ANSI-member organization that creates health-care messaging standards to enable health-care applications to communicate according to standard specifications. There also are vocabulary-domain specification tables and a terminology registry. The application are standards for the clinical-document architecture, a mark-up standard specific structure, and semantics of clinical documents.
ICIDH-2 [11] (International Classification of Impairment, Disability, and Handicap, Version 2.): Part of the World Health Organization suite of terminology. This is an ontology-based means for standardized representation of generic-health and/or functional-status data.
LOINC [28 through 32] (Logical Observations, Identifiers, Names, and Codes): A public domain set of codes and names intended for electronic reporting of laboratory test results. Intended to cover 98 percent of tests performed in an average laboratory.

SNOMED [33 through 38] (Systemized Nomenclature of Human and Veterinary Medicine): A standardized medical nomenclature of basic concepts.
UMLS [39 through 71] (Unified Medical Language System): A National Library of Medicine repository containing 800,000 concepts for biomedical domain. Consists of a hierarchical meta-thesaurus linked by relationships.

A good naming convention probably will not require that everyone in the world name things the same way, only that users map to a standard. Two different processes could use any name they want, map to an agreed-upon standard, thereby improving confidence that the chosen term maps formally to a common sign, and, thus, refers to the same concept. Therefore, to help with the mapping, terms used by each process should be defined in, and put in context by information models. This will help assure that a term used by one process includes the same concepts, only those concepts, and nothing but those concepts, as the corresponding term of the second process. In addition, the mapping process between different terms must be easy to use, and, preferably, automated.

A way to resolve terminology issues that shows promise is by using ontologies. An ontology, for the purposes of this paper, is a conceptual-reference model that is human and machine readable, that describes aspects of interest about some domain. [72] It is similar to a dictionary in that meanings of terms are there, but rather than natural language glosses, the definitions are represented using a formal modeling methodology to insure semantic precision. Ontologies become more important when software agents of enterprise processes are seeking information from other software agents; say, engaging in electronic commerce on the semantic web.

Huhns [73] states that a pre-defined global ontology would be too cumbersome, never large enough, and too difficult to use. Instead, Professor Huhns and others foresee that each software application will contain an ontology that applies to its own domain of discourse. Hopefully, this ontology will be organized in a way consistent with international standards. In addition, standards will prescribe the conversation process that will allow ontologies to determine common semantic representations. The common terms of the transaction in question then will be organized into a temporal ontology at one of the applications. This temporary ontology will contain terms of the transaction of the moment so that information transfer can occur. After the transaction is complete the temporal ontology will be erased to make room for the next transaction and to maintain referential integrity of the ontology of each application.

4. Enterprise Architectures, Frameworks, and Models

The enterprise-reference architecture is knowledge that covers enterprise structure, the enterprise-representation framework, rules and guidelines for enterprise representation, theory of views, modeling methodology, and languages. [74] The architecture covers the enterprise throughout its life cycle: identification, concept, requirements, design, implement, operate, and decommission. Even if there existed a technique to resolve nomenclature or semantic problems, would enterprises be able to understand each other's information given the different architectures? Is it possible, imperative, or desirable that all enterprise structures be the same to achieve better information integration?

In similar business areas, the functions in one enterprise are remarkably similar to those in another, from the high level down to the bottom level. Enterprise engineers design what they feel is an optimum way to accomplish the functions of the subject enterprise, to the degree that two independently generated enterprise architectures probably will not be similar. However, the architectures and frameworks exist for the convenience of each enterprise planner for analysis purposes. The nature of the enterprise architecture is of little importance outside that enterprise. Communication among enterprises with different architectures is much as it is now when we use Internet, faxes, and telephones to communicate with others in different and complex organizational structures.

Enterprise models will allow more consistent modularization so that enterprises, or parts of enterprises, can interchange information about pieces of themselves for the purpose of setting up the parameters for an information transfer. Simulation will be possible allowing evaluation of interoperation with inter-enterprise entities and evaluation of systems with differing granularity. Enterprises will be able to plan migration

paths more effectively. Because information will be an asset, separate from the components of the information system, upgrades will be possible without re-entering information about the products and processes unnecessarily. Enterprises can define paths to make the product and process information tie logically into enterprise goals, strategies, capabilities, and business rules. The models should be scalable so that a high-level model is essentially the same as a lower-level model--that is, use the same modeling constructs for all levels.

5. Role of Standards in Intra- and Inter-Enterprise Operation and Information Transfer

5.1 Enterprise-Operation Standards

People or machines analyze, design, simulate, and operate enterprises with tools. To justify the investment of the tool builder and to make the selling price of tools reasonable, the tools must be sufficiently generalized so that more than one enterprise can use them. The tools of each tool supplier, then, should be extensible, migratable, and interoperable so that one medical enterprise can evaluate a query from another enterprise, gather necessary information, and respond to the query. Each enterprise will be operating at its own chosen points on the various integration continua discussed in Section 3. Well-designed standards will require applications from different vendors to share enough information in a standard format for these applications to interoperate but not so much as to limit functionality unacceptably.

But, how much freedom should a health-care entity have to design the enterprise of its choice? With a large degree of freedom can there still be acceptable integration? Large businesses typically consist of groups, divisions, departments, and sections. Somewhere in that hierarchy the organization becomes function oriented and the higher-level manager establishes incentives to lower-level managers to reduce the costs of their function. When that happens there is no advantage to doing things for the benefit of the enterprise because process and infrastructure improvements tend to sub-optimize information transfers, and even conflict with similar investments in other departments. Subsequent reorganization will attempt to fix the problem and the enterprise designers will be busy. This cultural problem is quite strong and it will exist whether or not technical rationality or standards recommend otherwise.

A standards strategy to resolve semantic heterogeneity contains controversial issues. Some experts say that many standards are required. Others say that open-systems with as few basic standards as possible is better. Standards that pre-define a specific set of capabilities that would resolve heterogeneity and qualify an enterprise as integrated would tend to codify the enterprise strategies and reduce product and process differentiation. The idea of a standard-capability set is contrary to the operation of a free-enterprise market and not feasible for current and foreseen world-commerce situations. This is especially true with a distributed enterprise such as the medical industry. Even if standards that predefine the entire enterprise were feasible, humans will always demand the freedom to improve enterprise operation through redesign. Standardizing the enterprise structure, parts of the enterprise, the products, the information transferred, and the processes probably would not be feasible, because the probability that international standards could constrain enterprise, product, or process structure is very remote.

If standardized enterprises and processes are not feasible, then at what level is a standard appropriate? The domain consists of hardware, software, communications protocols, information, frameworks, and architectures. There are things, the connections between the things, the information, and the information formats.

What seems more usable is to continue what we currently do with a modification. That is, to standardize some interfaces, the infrastructure for defining terminology, and the information formats. To claim conformance with these standards, the enterprise-tool builders would have to use these standard requirements to design software products in such a way that allows the processes to pre-communicate particulars about itself before the actual information is transferred.

Consider the types of interfaces that should be standardized. Of course, infrastructure standards are necessary; such as, those in categories defined in the International Organization for Standardization's *Open*

System Interconnection Reference Model, ISO 7498. This standard defines a communication event at seven layers, which range from the physical connections to the application components of the system. This would include, for example, the standards that enable an Internet communication to occur. Above those is the content of the messages that are sent. Information content, format, and terminology are generally selected by builders of tools that transfer information about processes and products within and among enterprises.

There is an approach being proposed for standards that encourage information interoperability in the manufacturing industry that also may work for the highly distributed processes that occur in the health-care industry. Rather than setting specific requirements for information content, format, and terminology, such standards would require an enterprise process to be aware of its purpose and characteristics, and document, perhaps electronically on line, what its communicating capability is; and then operate itself accordingly. Lower-level standards mentioned above would still define interfaces and protocols. The new standards would be at a level above the highest Open-System Interconnection, or OSI, Application layer. These standards would define how the processes could determine, on the fly, what information the compliant processes must present on line about their capability. This way, a process asking another process for information, a querying process, would be able to evaluate, on line, whether a responding process conforms to listed interface standards. The querying enterprise could then advise its software to make any needed adjustments to its interfaces, software, communication protocols, and information formats; establish a relationship; and operate at an optimum level of integration for both parties. The query and response process would occur with a script and sequence prescribed in the standard.

Standards to accomplish these activities are being designed and proposed in ISO TC 184 SC5.[6] While these standards are developed for industrial-software applications, the approach is applicable to health-care-software applications.

5.2 Information-Transfer Standards
With this pre-communication established, information transfer can begin. What must exist to enable satisfactory information transfer and how much of that should be standard? What material qualifies for consideration as a standard? The easy, but unrealistic, answer is everything. The appropriate answer involves some tradeoffs. One must consider the rate of technology change versus the time it takes to create and change standards. And, with standards, one must evaluate the flexibility one surrenders in the ability to migrate to a newer technology that will help to achieve a higher degree of integration. Another significant factor, in smaller enterprises especially, is the increased cost of buying tools that comply with standards.

Knowing which additional standards or technology are necessary for information transfer depends on the existing situation. For example:
- Is the information at the interface between applications in an open, known format?
- How much of the terminology is in the one-name-one-meaning category?
- How compatible is the hardware?
- How well defined are the software interrelationships?
- Are the communication protocols agreed upon?
- Are the information architectures compatible?

Even with the standards defined, the business rules of each enterprise entity must be compatible, the context must be similar, government regulations for commerce and communication must be agreeable, and the people of the enterprise must truly want improved integration.

6. Implications for Health-Care Systems
The health-care industry needs better integration of its information infrastructure. The health-care entities--payers, clinics, physicians, groups of physicians, hospitals, and government--are characteristic of what is referred to as a network of virtual and extended enterprises.

---

[6] International Organization for Standardization, Technical Committee 184, Industrial Automation Systems, Subcommittee 5, Architectures, Communications, and Integrating Frameworks.

As the health-care industry undertakes its information analysis, it can, hopefully, partition the problem into solvable chunks while avoiding errors that the traditional industrial enterprises have made in the past. Some guidelines are:

- Establish open, standards-based solutions for small portions of the enterprise, thereby making intra-enterprise communication easier.
- Solve semantic problems and not concentrate on short-term and single-process, special-purpose, hardware and software solutions.
- Justify investments to the highest levels in the enterprise for a total-enterprise solution.
- Establish an industry-wide, open, information infrastructure.

In addition, the solution technologies must:

- Employ a system that permits users to enter data only once, and use it many times at other nodes in the system.
- Offer alternatives to users at the smallest and most harried nodes that supplant entering the information into the system using only a keyboard.
- Not require users to memorize anything or to use large and complicated look-up tables.
- Plan for an agile system that can migrate with technology improvements.
- View legacy systems as a part of the solution rather than part of the problem.
- Use systems that are intuitive to use for an office staff with average skills.
- Offer the capability to enter such graphics as photos, line diagrams, x-rays, and scans.

The system must enable transfer of consistent meaning among users. Even in normal conversation, paying attention to consistent semantics is a difficult job. It is a real challenge to build a semantically consistent system and, at the same time, build a system compatible with the above, and to make it affordable (since no one, especially a small entrepreneur, can afford to pay dearly for these improvements).

According to an old adage, a product "can be good, fast, and cheap--pick any two". This admonishment is no longer valid. Using a total-quality approach, system developers are discovering that products with quality designed into them (good), mean more efficient production and less time in rework (fast), and, lower selling price as well as higher margins, because production and warranty costs are lower (cheap). Therefore, users can demand and receive all three attributes in a single system.

7. Solution Possibilities

The most logical approach probably is to begin at the largest nodes, such as a physicians group of linked small-to-medium-sized enterprises, a payer, or a large entity such as a hospital. The communicating medium could be the Internet, such as secure e-mail and a new domain: "node@server.hce" (health-care enterprise). Another approach is to post required information on some secure application of the World-Wide Web. One problem with the largest nodes is the nature of the information there. Much of it probably is still on large mainframe-computing systems, access to which is not always very easy by the small computer users--but it is possible. Once the network is established and benefits are seen to exist, there will be a market for inexpensive tools that will enable more and more of the small nodes to participate.

Assume that it is feasible to interconnect these data systems to improve the quality of health-care informatics. How would such a flatly organized set of information be managed? How would the interconnection process start? There needs to be some forum where consensus among the node proprietors can be developed. This group must select a minimum of new-technology-based, and standards-based, things required to begin operation. Early compatibility with the existing information-technology infrastructure, such as the World-Wide Web, Internet, and common protocols, will allow the network to grow. The system designers should plan to use the heterogeneous system that has been in place for a long time.

System designers should not be daunted by the specter of legacy systems and legacy data--these systems should be included directly into the analysis. When one considers the rate at which new technology is developed and introduced, the short-life cycle of a product in information technology, and that everyone

does not invest in the same version of every equipment or software technology, one may conclude that legacy systems will always exist. Therefore, it is not a one-time hurdle to eliminate legacy systems. Similarly, to freeze a system technology by standards or mandate, or shortening standards-development periods to provide more flexibility are short-sighted strategies. A better long-term strategy is to characterize the nature of the enterprise and its information system, be aware of how the enterprise operates in its current environment, simplify those operations, and create an extendible and inter-operating system that serves the enterprise in its current operating environment. Things are the way they are because something about them works best in each separate context.

A key ingredient to mitigate the terminology problem is a logic-based ontology defined by international standards. Each software application would contain such an ontology pertaining to its own vocabulary as discussed at the end of Section 3. Systems such as agents could be developed to reconcile definitions of terms that do not match, terms that have different definitions, or terms that are missing in one of the applications. The eventual solution to the terminology problem, to be practical, has to be automated and not require humans to select from a look-up table.

A short-sighted approach to the terminology challenge is to design a system that obviates a lot of the ambiguity through a closely controlled language-usage mechanism. The crunch with such a system will occur at the point of data entry into the system. Consider these. How much patience will a physician have if forced to use a keyboard, at a computer terminal, while interviewing a patient, with seven examination rooms occupied and a full waiting room? Then, if the data-entry system forces the physician, or other data-entry person, to use proper standard vocabulary, and continually reminding the data-entry person that "Sorry, the terminology you are using is not acceptable--do you wish to review the list of standard terminology?" The authorized list, surely, will be quite lengthy. Probably, there also will be no time later to enter the data. Anyone that enters data in this environment soon will become frustrated and no longer use that system

8. Recommendations
These recommendations are intended as a guideline for approaching the task to better integrate the information shared among health-care enterprises. They have been grouped into four categories: enterprise definition, enterprise-information analysis, information-system design, and information-technology standards.

8.1 Enterprise Definition
- Consider the health-care-information domain to be an enterprise with characteristics similar to a small-to-medium-sized, industrial enterprise, operating in a virtual-enterprise mode.
- Define the nature and scope of the information aspect of the enterprise applications. To design the information-sharing environment, define what is inside the enterprise and what is outside. Define the places in enterprise processes where information enters, is processed, and exits. These are the nodes of this information system.

8.2 Enterprise-Information Analysis
- Define an information-generating-and-use diagram to pinpoint where, what, why, and how the information is created and where, what, why, and how the information is used. This will identify customers and suppliers and the enterprise resources that must eventually be accommodated in the integrating process. It may even identify that some information is generated and not used, thereby effecting a simplification of the information structure.
- Between the analysis and automate phases be sure to simplify where possible to avoid unnecessary complication due to unnecessary legacy activities.
- Prioritize the areas of the enterprise, the information nodes, according to where the biggest payback exists. Plan to start the analysis and implementation at these areas.
- Develop a profile of what information is public, what is private, and who really owns the information. Develop a security plan, a privacy plan, and a right-to-access plan. Avoid overburdening the new

system by trying to solve security and privacy problems that are considered tolerable risk in the present systems.

8.3 Information-System Design
- Determine whether a centralized or a distributed system will meet needs better. Consider user input so that the resultant system is user friendly and does not require technical knowledge of information systems to use it. Similarly, consider vendor input to improve chances of implementation.
- Consider networks and communication protocols that already exist and work well, such as the Internet and the World-Wide Web. Determine gaps in technology and standards to cover unsatisfactory areas. Inject needs for health-care improvement into relevant programs. One such is the National Industrial Information Infrastructure Program, NIIIP[7], which is attempting to enhance these existing networks for electronic commerce in the shipbuilding industry. Another is the Integrated Manufacturing Technology Roadmapping Initiative, IMTI[8], that is trying to identify the best solutions to integration problems and fit them into a milestone format to further understanding of the major issues.

8.3 Information-Technology Standards
- Become aware of and participate in standards and technology projects in related domains that can resolve needs, such as the manufacturing-enterprise domain (ISO TC184, Industrial-Automation Systems and Integration), and the information-technology domain (ISO/IEC JTC1).
- Decide which transaction components standards should constrain: process structure, information content, interfaces between processes, and/or known hook-up points so that the software knows with what it is dealing.
- Decide which standards in place will increase the market for tools, thereby justifying tool-development costs.

---

[7] See http://www.NIIIP.org
[8] See http://www.imti21.org

9. Bibliography

Over the past several years, the author has assembled a Health-Care Information Reference List that contains about 260 readings related to health-care informatics. This bibliography is a subset of that list. The numbers in braces in the items that follow, for example {X}, correspond to the numbers assigned to the documents in that reference list.

1. Industrial Modernization Incentives Program (IMIP), Phase 1, Enterprise Analysis, Richard Engwall, et al, Final report, Contract # F19628-92-C-0021, CRDL # A004, Electronics Systems Center, Air Force Materiel Command, USAF, January 31, 1993 {78}
2. Industrial Automation Systems--Concepts and Rules for Enterprise Models, ISO 14258, International Organization for Standardisation (ISO), 1998-09-01, Reference number: ISO 14258:1998(E) {74}
3. Introduction to EDI and Health Care for HIPAA, Rachel Foerster, Foils from Presentation at NIST, 2002-July-31, Rachel Foerster and Associates LTD, Beach Park, IL, http://www.rfa-edi.com {195}
4. Enterprise Organizations and the New Enterprise Paradigms, K Kosanke, CIMOSA Association, JG Nell, NIST, IFAC World Congress, Elsevier Science Limited, ISBN 0-08-043248-4, 1999-July {73}
5. Semantic Heterogeneity in Multi-database Systems, Manuel Garcia-Solaco, Felix Saltor, and Malu Castellanos, Universitat Politecnica de Catalunya, Spain. Chapter 5 of: Object-Oriented Multi-database Systems, A solution for Advanced Applications, Edited by: Omran A. Bukhres, Ahmed K. Elmagarmid, Purdue University 1996. {81}
6. A Collision of Semantics, Geoffrey A. Howe, Database Programming and Design, February 1993. {82}
7. On Resolving Schematic Heterogeneity in Multi-database Systems, Won Kim, Injun Choi, Sunit Gala, and Mark Sheevel, UniSQL Inc., In Distributed and Parallel Databases 1, 1993. {83}
8. So Far (Schematically) yet So Near (Semantically), Amit Sheth and Vipul Kashyap, Bellcore and Rutgers University, Interoperable Database Systems, 1993. {84}
9. From Data to Knowledge through Concept-oriented Terminologies; James J Cimino; Journal of the American Medical Informatics Association, Volume 7 Number 3, May/June 2000 {124}
10. Clinical Terminology: Why is it so Hard? AL Rector; Methods of Information in Medicine; 1999: 38:239-252 {125}
11. Expression of a Domain Ontology Model in Unified Modeling Language for the World Health Organization International Classification of Impairment, Disability, and Handicap, Version 2, Alexander P Ruggieri, Peter L Elkin, Harold Solbrig, Christopher G Chute, ibidem. {110}
12. Development of an Ontology to Model Medical Errors, Information Needs, and the Clinical Communication Space; Peter D Stetson, Lawrence K McKnight, Suzanne Bakken, Christine Curran, Tate T Kubose, James J Cimino; Journal of the American Medical Informatics Association, 2001 AMIA Symposium, 2001 November-3/7; ISBN 1-56053-536-9 {111}
13. Change Management of Shared and Local Version of Health-Care Terminologies; DE. Oliver, Y. Shahar, Methods of Information in Medicine; 2000: 39:278-290 {120}
14. Clinical Terminology: Why is it so Hard? AL Rector; Methods of Information in Medicine; 1999: 38:239-252 {125}
15. Concepts of Language and Language of Concepts, Methods of Information in Medicine, J Cimino, F K Schattauer, Verlagsgesellschaft mbH. 1998 {126}
16. HL7 Website: www.hl7.org/special/splash.cfm {200} and HL7 Messaging Standard Recommended as Core National PMRI Standard under HIPAA, Karen Van Hentenryck. {200}
17. The HL7 Clinical Document Architecture, Robert Dolin, et al, JAMIA, Vol. 8 #6, 2001-Nov/Dec. {202}
18. Netscape wasn't Enough, Rob Hof, Business Week, June 24, 1996, p6 {154}
19. Letter to George Beeler, et al, October 31, 1995. Standards-related documents for Open EDI efforts for IEEE P1157.1, Data Modeling. {193}
20. Letter to George Beeler, et al, November 29, 1995. Additional standards-related documents for Open EDI efforts for IEEE P1157.1, Data Modeling {194}
21. GALEN's Model of Parts and Wholes: Experience and Comparisons, Jeremy Rogers, Alan Rector, JAMIA (Journal of the American Medical Informatics Association) Proceedings: Converging Information, Technology, and Health-care, 2002-Nov. {196}

22. Natural Language Processing: Techniques Associated with the open GALEN Ontology for Semi-Automatic Textual Extraction of Medical Knowledge: Abstracting and Mapping Equivalent Linguistic and Logical Constructs, M Biczyk do Amaral, A Roberts, AL Rector, ibidem. {197}

23. Having our Cake and Eating it too: How the GALEN Intermediate Representation Reconciles Internal Complexity with Users Requirements for appropriateness and Simplicity, WD Solomon, A Roberts, JE Rogers, CJ Wroe, AL Rector, ibidem. {198}

24. Goals for Concept Representation in the GALEN Project; AL Rector, WA Nowlan, Andrzej Glowinski; AMIA Symposium on Computer Applications in Medical Care; October 30-November 3, 1993 {199}

25. Toward Vocabulary Domain Specifications for Health Level 7, Coded-data elements, Suzanne Bakken, et al, Journal of American Medical Informatics Association (JAMIA), Vol. 8 #6 2001-Nov/Dec. {201}

26. The HL7 Clinical Document Architecture, Robert Dolin, et al, JAMIA, Vol. 8 #6, 2001-Nov/Dec. {202}

27. Implementing HL7: From the standards specification to Production Application, G Schadow, V Föhring, Th Tolxdorff, Meth Information Med, F K Schattauer, Verlagsgesellschaft mbH. 1998 {203}

28. LOINC and RELMA; The Regenstrief Institute, Web page as of 2002-06-12: http://www.regenstrief.org/loinc. {206}

29. Integration of Nursing Assessment Concepts into the Medical Entities Dictionary Using the LOINC Semantic Structure as a Terminology Model; Bethany J Cieslowski, David Wajngurt, James J Cimino, Suzanne Bakken; Journal of the American Medical Informatics Association, 2001 AMIA Symposium, 2001 November-3/7; ISBN 1-56053-536-9. {207}

30. A Method for the Automated Mapping of Laboratory Results to LOINC; Lee Min Lau, Kate Johnson, Kent Monson, Siew Hong Lam, Stanley M Huff; Journal of the American Medical Informatics Association, 2000 AMIA Symposium, 2000 November-4/8; ISBN 1-56053-480-X. {208}

31. Evaluation of the Clinical LOINC Semantic Structure as a Terminology Model for Standardized Assessment Measures; Suzanne Bakken, James J Cimino, Robert Haskell, Rita Kukafka, Cindi Matsumoto, Garrett K Chan, Stanley M Huff; Journal of the American Medical Informatics Association, Volume 7, Number 6, Nov/Dec 2000. {209}

32. Coupling Vocabularies and Data Structures: Lessons from LOINC; Roberto A Rocha, Stanley M Huff; 1996 AMIA Annual Fall Symposium Proceedings; October 26-30, 1996 {210}

33. The SNOMED Model: A Knowledge Source for the Controlled Terminology of the Computerized Patient Record; YA Lussier, DJ Rothwell, RA Cote; Methods of Information in Medicine; 1998: 37:161-164 {211}

34. Managing Information with SNOMED, Understanding the Model, D.J. Rothwell and R.A. Cote, Medical College of Wisconsin and Universite de Sherbrooke, 1996. {212}

35. The SNOMED DICOM Microglossary: Controlled Terminology Resource for Data Interchange in Biomedical Imaging; W Dean Bidgood, Jr; 1996; Copyrighted by the author. {213}

36. SNOMED-Based Knowledge Representation; DJ Rothwell; Methods of Information in Medicine; F K Schattauer, Verlagsgesellschaft mbH. 34:1995 {214}

37. Associating Semantic Grammars with the SNOMED: Processing Medical Language and Representing Clinical Facts into a Language-Independent Frame; do Amaral Marcio B and Satomura Y; MEDINFO'95, Proceedings of the Eighth World Congress, 23-27 July 1995 {215}

38. Developing a Standard Data Structure for Medical Language--the SNOMED Proposal; DJ Rothwell, RA Cote, JP Cordeau, MAS Boisvert; AMIA Symposium on Computer Applications in Medical Care; October 30-November 3, 1993 {216}

39. Circular Hierarchical Relationships in the UMLS: Etiology, Diagnosis, Treatment, Complications and Prevention; Olivier Bodenreider; Journal of the American Medical Informatics Association, 2001 AMIA Symposium, 2001 November-3/7; ISBN 1-56053-536-9. {217}

40. Battling Scylla and Charybdis: The Search for Redundancy and Ambiguity in the 2001 UMLS Metathesarus; James J Cimino; Journal of the American Medical Informatics Association, 2001 AMIA Symposium, 2001 November-3/7; ISBN 1-56053-536-9. {218}

41. Managing Name Ambiguity in the UMLS Metathesaurus; Laura Roth, William T Hole; Journal of the American Medical Informatics Association, 2001 AMIA Symposium, 2001 November-3/7; ISBN 1-56053-536-9. {219}

42. A Metaschema of the UMLS Based on a Partition of its Semantic Network; Michael Halper, Zong Chen, James Geller, Yehoshua Perl; Journal of the American Medical Informatics Association, 2001 AMIA Symposium, 2001 November-3/7; ISBN 1-56053-536-9. {220}

43. A Semantic Navigation Tool for the UMLS; Olivier Bodenreider; Journal of the American Medical Informatics Association, 2001 AMIA Symposium, 2001 November-3/7; ISBN 1-56053-536-9. {221}

44. Partitioning the Semantic Network of the UMLS; Zong Chen, Michael Halper, James Geller, Yehoshua Perl; Journal of the American Medical Informatics Association, 2001 AMIA Symposium, 2001 November-3/7; ISBN 1-56053-536-9. {222}

45. Exploring the Semantic Relations of Radiological Terms using UMLS Resources; David A Campbell, Donald P Harrington, James J Cimino; Journal of the American Medical Informatics Association, 2001 AMIA Symposium, 2001 November-3/7; ISBN 1-56053-536-9. {223}

46. Automated Knowledge Extraction from MEDLINE Citations; Eneida A Mendonca, James J Cimino; Journal of the American Medical Informatics Association, 2001 AMIA Symposium, 2001 November-3/7; ISBN 1-56053-536-9. {224}

47. Knowledge Acquisition to Qualify Unified Medical Language System Interconceptual Relationships; Franck Le Duff, Anita Burgun, Mireille Cleret, Bruno Pouliquen, Victoire Barac'h, Pierre le Beux; Journal of the American Medical Informatics Association, 2001 AMIA Symposium, 2001 November-3/7; ISBN 1-56053-536-9. {225}

48. UMLS Concept Indexing for Production Databases: A Feasibility Study; Prakash Nadkarni, Roland Chen, Cynthia Brandt; Journal of the American Medical Informatics Association, Volume 8 Number 1, Jan.Feb 2001. {226}

49. Discovering Missed Synonymy in a Large Concept-Oriented Metathesaurus; William T Hole, Suresh Srinivasan; Journal of the American Medical Informatics Association, 2000, AMIA Symposium, 2000 November-4/8; ISBN 1-56053-480-X. {227}

50. Using UMLS Semantics for Classification Purposes; Olivier Bodenreider; Journal of the American Medical Informatics Association, 2000 AMIA Symposium, 2000 November-4/8; ISBN 1-56053-480-X. {228}

51. The UMLS Knowledge Source Reserve: A Versatile Internet-Based Research Tool; Alexa T. McCray, Amir M Razi, Anantha K Bangalore, Allen C Browne, P Zoe Stavri, American Medical Informatics Association, Proceedings, Symposium, October 26-30, 1996 {229}

52. Finding the Findings: Identification of Findings in Medical Literature using Restricted Natural Language Processing; Charles A Sneiderman, Thomas C Rindflesch, Alan R Aronson, American Medical Informatics Association, Proceedings, Symposium, October 26-30, 1996 {230}

53. The UMLS Knowledge Source Server and the NLM/AHCPR Large Scale Vocabulary Test; Amir Razi, Anantha Bangalore, Guy Divita, Allen Browne, P Zoe Stavri, Alexa T McCray, American Medical Informatics Association, 1996 {231}

54. UMLS-Based Access to CPR Data; Erik M van Mulligen; American Medical Informatics Association, 1996 {232}

55. The Representation of Meaning in the UMLS; AT McCray, SJ Nelson; Methods of Information in Medicine, F K Schattauer, Verlagsgesellschaft mbH. 34:1995 {233}

56. The UMLS Knowledge Source Server; AT McCray and A Razi; MEDINFO'95, Proceedings of the Eighth World Congress, 23-27 July 1995 {234}

57. A conceptual Graphs Modeling of UMLS Components; M Joubert, F Milton, J-J Robert; MEDINFO'95, Proceedings of the Eighth World Congress, 23-27 July 1995 {235}

58. Formal Properties of the Metathesaurus; Mark S Tuttle, Nels E Olson, Keith E Campbell, David D Sherertz, Stuart J Nelson, William G Cole; AMIA Symposium on Computer Applications in Medical Care; November 5-9, 1994 {236}

59. A Conceptual Model for Information Retrieval with UMLS; Michel Joubert, Marius Fieschi, Jean-Jacques, Robert; AMIA Symposium on Computer Applications in Medical Care; October 30-November 3, 1993 {237}

60. From ICD9-CM to MeSH Using the UMLS: A How-to Guide; James J Cimino, Stephen B Johnson, Ping Peng, Anthony Aguirre; AMIA Symposium on Computer Applications in Medical Care; October 30-November 3, 1993 {238}

61. Using the UMLS Semantic Network as a Basis for Constructing a Terminological Knowledge Base: A Preliminary Report; Giuseppe Carenini and Johanna D Moore; AMIA Symposium on Computer Applications in Medical Care; October 30-November 3, 1993 {239}
62. UMLS Knowledge for Biomedical Language Processing; Alexa T McCray, Alan R Aronson, Allen c Browne, Thomas C Rindflesch, Amir Razi, Suresh Srinivasan; Bull Medical Library Association; April 1993 {240}
63. Coach: Applying UMLS Knowledge Sources in an Expert Searcher Environment; Lawrence C Kingsland III, Anna M Harbourt, Edmund J Syed, Peri L Schuyler; Bull Medical Library Association; April 1993 {241}
64. Using a Network Menu and the UMLS Information Sources Map to Facilitate Access to Online Reference Materials; Jeffrey I Clyman, Seth M Powsner, John A Paton, Perry L Miller; Bull Medical Library Association; April 1993 {242}
65. The UMLS Project: Making the Conceptual Connection between Users and the information they Need; Betsy L Humphreys, Donald AB Lundberg; Bull Medical Library Association; April 1993 {243}
66. The UMLS Metathesaurus: Representing Different views of Biomedical Concepts; Peri L Schuyler, William T Hole, Mark S Tuttle, David D Sherertz; Bull Medical Library Association; April 1993 {244}
67. Structuration and Acquisition of Medical Knowledge using UMLS in the Conceptual Graph Formalism; F Volot, P Zweigenbaum, B Bachimont, M Ben Said, J Bouaud, M Fieschi, JF Boisvieux; AMIA Symposium on Computer Applications in Medical Care; October 30-November 3, 1993 {245}
68. The METAS_1 Engine: A Database Methodology Used in Building the UMLS Metathesaurus; DD Sherertz, NE Olson, MS Tuttle, WD Sperzel, MS Erlbaum, and LF Fuller; MEDINFO'92, Proceedings of the Seventh World Congress on Medical Informatics, 6-10 September 1992 {246}
69. The Semantic Foundations of the UMLS Metathesaurus; MS Tuttle, SJ Nelson, LF Fuller, DD Sherertz, MS Erlbaum, WD Sperzel, NE Olson, ON Suarez-Munist; MEDINFO'92, Proceedings of the Seventh World Congress on Medical Informatics, 6-10 September 1992 {247}
70. Beyond UMLS: Computational Semantics for Medical Records; A Gangemi, M Galanti, E Galeazzi, Mori A Rossi; MEDINFO'92, Proceedings of the Seventh World Congress on Medical Informatics, 6-10 September 1992 {248}
71. The Unified Medical Language System Project: A Distributed Experiment in Improving Access to Biomedical Information; BL Humphreys, DAB Lindberg; MEDINFO'92, Proceedings of the Seventh World Congress on Medical Informatics, 6-10 September 1992 {249}
72. Ontologies and their Role in Knowledge Management and E-Business Modelling; Hans Akkermans; Enterprise Inter- and Intra-Organizational Integration--Building International Consensus; Edited by: Kosanke, Jochem, Nell, and Ortiz; Kluwer Academic Publishers, ISBN 1-4020-7277-5; 2003.
73. Semantic Bridging of Independent Enterprise Ontologies; Michael N. Huhns and Larry M. Stephens; Enterprise Inter- and Intra-Organizational Integration--Building International Consensus; Edited by: Kosanke, Jochem, Nell, and Ortiz; Kluwer Academic Publishers, ISBN 1-4020-7277-5; 2003.
74. ISO 15704, 2000, Industrial automation systems--Requirements for enterprise-reference architectures and methodologies, TC 184 SC 5 WG 1.
75. ISO 10303, 1993 Industrial automation systems--Standard for the exchange of product model data, TC 184 SC 4.
76. ISO 18629, 2001, Industrial automation systems--Process Specification Language, PSL, TC 184 SC4/5 JWG 8.