The Semantic Web As "Perfection Seeking:" A View from Drug Terminology

Mark S. Tuttle, *Apelon, Inc.*, mtuttle@apelon.com
Steven H. Brown, MD, *Vanderbilt University / Veterans Administration*
Keith E. Campbell, MD, PhD, *Inoveon, Inc.*
John S. Carter, *University of Utah / Apelon, Inc.*
Kevin D. Keck, *Keck Labs*
Michael Lincoln, MD, *University of Utah / Veterans Administration*
Stuart J. Nelson, MD, *National Library of Medicine*
Michael Stonebraker, PhD, *Massachusetts Institute of Technology*

Abstract. To date, the Semantic Web has viewed formal terminology, or ontology, as either immutable, or something that can change but that has no past and no future – only a present. Change, or process – such as "perfection seeking," is outside the scope of the proposed "semantics," except in so far as it is represented in attributes. In contrast, current U.S. Government efforts to formalize drug (medication) terminology are being driven by the need to manage changes in this terminology asynchronously and longitudinally. For example, each year the FDA (Federal Drug Administration) approves about 150 new drugs and thousands of changes to the "label" of existing drugs, the VHA (Veterans Health Administration) must manage new drugs, label changes, and tens of thousands of drug "packaging" changes, and the NLM (National Library of Medicine) must maintain a current index of references to proposed or approved medications in the world's biomedical literature. We propose that an emerging multi-federal-agency reference terminology model for medications, mRT, be used to drive development of the necessary repertoire of "semantic" change management mechanisms for the Semantic Web, and that these "process" mechanisms be organized into an ontology of change.

1. Overview – Using mRT to drive the development of Semantic Web change management

Creating standards, especially standards that create information industry infrastructure, is difficult, time-consuming and at constant risk for irrelevance and failure. One way to mitigate this risk, and secure the participation of the diverse interest groups required to make such standards a success is to focus on process – as in the process that produces and maintains a good standard. This is in contrast to an approach that says some existing artifact selected from a list will be THE standard, and all the others will NOT be the standard. An observation that we attribute to Betsy Humphreys from the National Library of Medicine in the context of biomedical terminology standards is that it doesn't matter where you start, i.e., it doesn't much matter which terminology or terminologies one selects as a starting point; instead what does matter is the process by which the proposed standard evolves to achieve and sustain the desired degree of quality, comprehensiveness, and functionality. The process is what determines where the standard ends up.

Seen in this light, change, even a large amount of change, will be a feature of successful formal terminologies, or ontologies. We hope to demonstrate the feasibility and utility of this approach. The challenge in the context of the Semantic Web is to choose a representation for change that makes it explicit. Viewed this way the Semantic Web would be "perfection seeking,"[1] and the ongoing changes would be part of the semantics. The challenge with this approach is the formulation of the units of change and the creation of an ontology of these change units. This follows a Semantic Web notion expressed by Tim Berners-Lee in a discussion of Metadata Architecture [1] "… metadata itself may have attributes such as ownership and an expiry date, and so there is meta-metadata but we don't distinguish many levels, we just say that metadata is data and that from that it follows that it can have other data about itself. This gives the Web a certain consistency." Making change part of the Semantic Web would preserve that consistency.

One way to focus the development of the desired units, inter-relationships, and uses is to solve real problems and gain experience from deployments of these solutions; we propose to do this by formulating, deploying and evaluating what we now call "The New Drug Transaction." This transaction needs to supply diverse, operating healthcare and biomedical information systems with the requisite formal definition of a new drug, given a reference model, and do so at Web scale. The main challenge is how to do this in a way that first avoids breaking working applications that use the drug terminology and second preserves the longitudinal value of existing and future patient descriptions of medication use.

More generally, healthcare and biomedicine undergo constant change – some of it perfection seeking and some of it clerical – and the relevant terminology needs to change in parallel. Again, the challenge is to the extent possible to accommodate change without breaking what already works, and without losing the value of historical data.

A simple, large-scale model of longitudinal change management is that used by MEDLINE, the National Library of Medicine's citation database (http://www.ncbi.nlm.nih.gov/entrez/query.fcgi). The formal "semantics" of MEDLINE are supported by MeSH (Medical Subject Headings), a concept-based biomedical terminology that is updated annually (http://www.nlm.nih.gov/mesh/meshhome.html). Each year, rules are written that transform citations indexed using the previous year's MeSH into citations indexed using the new version of MeSH. In this way, by "re-writing history," old citations can be retrieved as appropriate using current terminology. As can be appreciated, formulating the rules requires manual intervention and testing, but more than 11 million citations, each tagged with about a dozen index terms selected from some 18,000 concepts, are maintained longitudinally in this way. While MEDLINE has always been a pre-eminent and exemplar information retrieval system, the notion of "history re-writing" implies a loss of information; the declining cost of secondary storage may eliminate one of the reasons for such information loss, a theme that will be re-examined below.

[1] Peri Schuyler, then head of the MeSH (Medical Subject Headings) at the NLM, used this term in the context of the UMLS (Unified Medical Language System) Project in 1988.

2. Background - The Semantic Web is a generalization of formalization efforts already underway in healthcare and biomedicine

In his recent Scientific American article Berners-Lee argues that the Semantic Web is infrastructure, and not an application [2]. We couldn't agree more. To us, this view is a top-down and horizontal approach to Semantic Web objectives, and it is this kind of disciplined thinking that made the Web the success that it is today.

In parallel with this effort, progress toward related goals is occurring in healthcare and biomedicine and we think of this progress as bottom-up and vertical. Thus, at present, healthcare and biomedicine have a repertoire of standard terminologies and standard messages and, in some instances, their use is or will be mandated by law.[2] While current deployments of these artifacts lack the formality required for the Semantic Web they nevertheless represent a rehearsal of many of the processes that the Semantic Web will require. Further, as will be described in a later section, the shortfalls of current healthcare terminology and message standards are driving a new generation of healthcare terminologies and messages that do have some of the desired formal properties. All this is part of a gradual evolution in healthcare information technology that is changing its focus from "systems" to "data," [3] [4] a trend predicted in [5]. The authors believe that the major forcing function for this evolution is the need to "scale" healthcare information technology to ever larger enterprises and collections of individuals and enterprises; while this trend began before the Web, the presence of the Web has accelerated the change.

What is missing in healthcare and biomedicine is a way to link its relevant progress and experience with that occurring in the Semantic Web community. The Web influences healthcare information technology, but the Web is little influenced by lessons learned in healthcare IT. We believe that medications represent a domain in which these two activities can be joined productively. The potential significance of such a joining cannot be over-estimated. Healthcare now costs the U.S. more than $1 trillion/year, and medications are the largest single category of cost and the fastest growing category of cost.[3] They are also involved in a significant number of life-threatening medical errors. [6]

At a deeper level, we believe that the Semantic Web is an opportunity to shrink the "formalization gap" described by Marsden S. Blois, PhD, MD (1918-88). Blois argued that overcoming this gap was the fundamental challenge of medical informatics: 'This discontinuity in formalization between a manual (human) medical information process and the machine code necessary to accomplish comparable ends begins at a very high descriptive level and it is not itself a concern of computer science. If this concern is to be given a name at all, it must be regarded as concerning medical applications, and it is increasingly being referred to as "medical information science" in the United States, and as "medical informatics" in Europe. It will be the task of this new discipline to better understand and define the medical information processes we have considered here, in order that appropriate activities will be chosen for computerization, and to improve the man-machine system.' [7] One rationale for a "perfection seeking" approach to the

---

[2] HIPAA (Health Insurance Portability and Accountability Act).

[3] Last year, VHA (Veterans Health Association) spent about $2.5 billion on medications, and MHS (Military Health System – covering active duty personnel and their dependents) spent about $1.5 billion. Personal conversation, Donald Lees, RPh, 6/01.

Semantic Web is the difficulty of getting the formalizations right, and of maintaining them, and the patient descriptions based on them, in the face of change.

3. A model - semantic definitions for medication active ingredients

If change management were not such a critical issue, already complete approximations of the medication reference model shown in ***Figure 1*** could be used by Semantic Web developers to test proposed representations. Carter, et al. [8] describe how about 1,000 active ingredients were given "Aristotelian" definitions represented in Description Logic and "published" in XML. One result of this effort was a focus on the emerging importance of "The New Drug Transaction" as a necessary conjunct to expansion of the model to cover all important active ingredients, and to trial deployments.

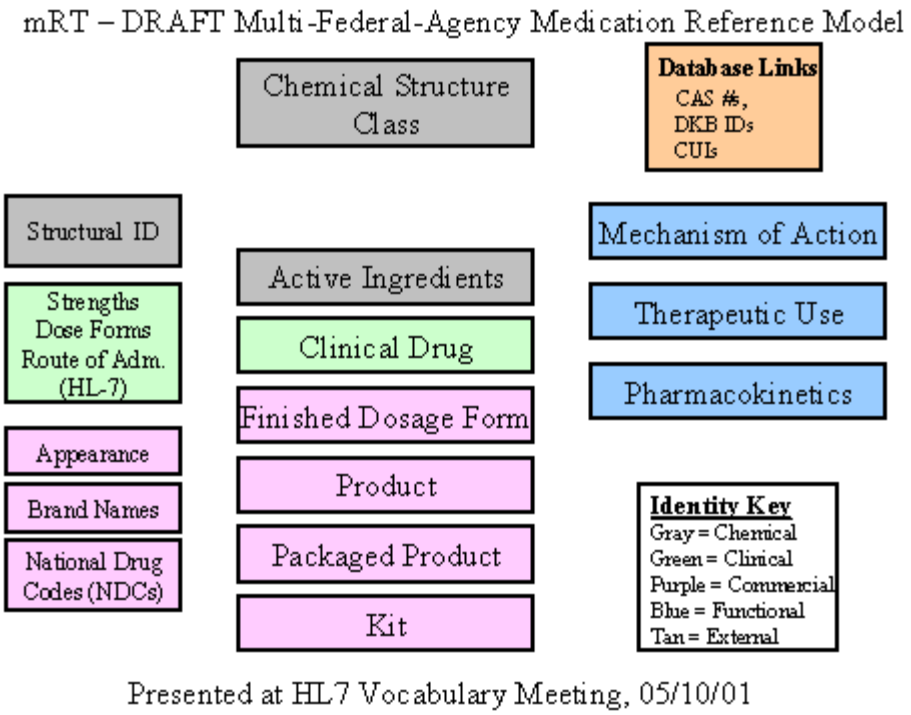

Figure 1 – DRAFT formal model of medications for potential use by three Federal Agencies: Active ingredients have "Aristotelian" definitions represented using Description Logic; these definitions will place each Active Ingredient in an IS_A hierarchy of Chemical Structure Classes, and describe each Active Ingredient using named relationships into reference taxonomies for, respectively, Mechanism of Action, Therapeutic Use, and Pharmacokinetics. Each Active Ingredient (molecule) will also have a machine-processible three-dimensional structural description (identifier). Not shown are inactive ingredients and other necessary details.

This model, developed over the last few years, has proven remarkably robust in the face of multi-disciplinary and multi-institutional inspection, and sample instantiations. Its next test will be to represent portions of various order-entry formularies used by the public and private sectors. A typical formulary covers about 10,000 – 100,000 "orderables" and the goal will be to produce "useful" definitions of the active ingredients contained in these orderables using early versions of the reference taxonomies for Chemical Structure, Mechanism of Action, Therapeutic Use, and Pharmacokinetics. This test will also allow us to gain experience assembling and formalizing medication information obtained from multiple authorities and disciplines that is used for related but still different purposes. For example, there will be at least three different kinds of "Therapeutic Use," also called "indications" – "FDA approved", "VA approved"(generally a superset of FDA approved), and "Described in the Literature".[4] The whole notion of "orderables" will also force clarification of the boundary between the so-called "terminology model" (categories and hierarchies) and the "instance" or "database" model (the orderables themselves, along with all their attributes). Everyone agrees that that the former is a good way to organize the latter, and that there should be a boundary between the two models – that is, the two models are similar and related but not the same, but few agree on where the implementation boundary should be, especially in light of emerging interoperation requirements based on re-usable objects. This dilemma should resonate with those working on the Semantic Web.

4. A process – embracing change and making it explicit

The model presented in Figure 1 is little more than an academic exercise without accompanying productive change management. Currently, excepting MeSH and MEDLINE (described above), change management in authoritative, deployed biomedical terminologies is at best primitive. [9] [10] As a result, there are few "warehouses" of patient descriptions that can be searched over time, that is across changes in the terminologies used to formalize the descriptions. Of the few patient description repositories that support such "time travel" no two do so in the same way, and none use existing or proposed standards. An explicit goal of the mRT project is to begin to overcome this shortfall at least in the context of medications.

The view of change management presented here is a synthesis of current and emerging practices in healthcare terminology, e.g., the use of Description Logic, earlier and current work on the handling of time-oriented data in database system models, e.g., POSTGRES [11] and T-SQL [12] [13], and our current understanding of the Semantic Web. This synthesis can be summed up by the conclusion that "Process is more important than representation."

4.1 A "new drug" transaction

The first step in making formal terminology changes into a terminology/ontology "thing," or unit, is to create a unit of change that has the same general properties as any

[4] An important side-effect of this effort will be an authoritative collection of so-called "off-label" uses of medications; such uses represent legal, but not FDA-approved, medication indications.

other "thing-ness" unit. For example, given the appropriate reference taxonomies, used to (in the Description Logic sense) "classify" medications, one can create the desired reference terminology – mRT – by "adding" the (Aristotelian) definitions of each drug, one drug at a time. But, of course, this ignores, among many other things, the fact that the reference taxonomies need to be changed, too. Frequently, new drugs come with new mechanisms of action and new indications (therapeutic objectives), and thus the corresponding "new drug transaction" may need to update the reference taxonomies before adding the definition of the new drug. These latter cases will be covered in "Other transactions" below.

To make the simple case more tangible, here is one potential near term future of the kind of "New Drug Transaction" that does not require updating the reference taxonomies:

1) The FDA will approve a new drug and "publish," as XML, a newly "structured" version of the traditional package insert, or "label," designed to "explain" that drug to both humans and computers. (One can think of this document as a "contract" between the FDA and the drug manufacturer.) The data that will appear in the new drug transaction is the result of processes now in place at the FDA; regulations are pending that will increase the degree of machine-processibility and formality of this data. [14]
2) The NLM will further process and enhance the parts of the label that can be processed usefully by computers, and then "publish" it, once again in XML. The "enhancements" may include connections to the biomedical literature, related terminology and foreign language names.
3) Applications or servers electing to process the new drug transaction will see that the XML indicates that it is an "add," the simplest kind of transaction to process. That is, the transaction will add a new concept – the new drug, the appropriate relationships to other concepts in the various reference taxonomies, and attributes of the new drug. (In every formulary or medication reference terminology known to the authors this is done manually, at present.)

It is not hard to imagine that most applications, e.g., drug order-entry systems, would be tolerant of such an insertion and subsequently "do the right thing." However, the problem with this simple form of the new drug transaction is that, as described by domain experts, most new drugs represent "changes in understanding," and it is not at all clear how existing applications can deal with such changes in understanding automatically, or know when they need help from humans. An extreme instance of such new understanding would be a drug that triggered a reorganization of the Aristotelian classification of existing drugs. (Changes in understanding due to pharmacogenetics may cause these kinds of "re-organizing" updates.)

4.2 Other transactions

In this context, "changes in understanding" are represented by changes in the reference taxonomies, e.g., for chemical structure, mechanism of action, pharmacokinetics, and therapeutic use. That is, a typical "new drug transaction" will need to include one or more changes to the reference taxonomies along with the (simple) "add" described above,

and these changes will represent "changes in understanding." It can be assumed that changes to the reference taxonomies will "break" existing applications, e.g., the decision support that operates in conjunction with order entry. The authors claim that to the degree that we can overcome this problem in the context of medication terminology maintenance that we are solving a problem that will be faced by the Semantic Web.

As presently planned our solution will be built on two foundations: First, mRT will not "overwrite" information; that is, per POSTGRES [15] any "garbage collection" or "archiving" will be handled asynchronously with new drug transactions, the practical effect being that an explicit, time-oriented, history of mRT is available to applications at all times. Second, appropriate use of Description Logic permits consistency-preserving updates; for example, if prior to execution of the new drug transaction an off-line, an updated copy of mRT is "reclassified" successfully (in the DL sense), then, in principle, mechanisms exist that can correctly update a run-time database (terminology server) "incrementally" (and thus quickly). Thus, such updates represent one useful repertoire of units of change.

Per earlier work of Stonebraker, et al. and more recent work of Snodgrass, et al., one can view a "database" as a time-oriented accumulation of changes. Thus the current "state" of the database is acquired through a computation, or "view," on the transactions accumulated over time. (See [13], for an enumeration of the many subtleties implicit here.) Part of the desired functionality is implemented, currently, in the MEME (Metathesaurus Enhancement and Maintenance Environment) deployed at the NLM. [16] The Metathesaurus (http://www.nlm.nih.gov/pubs/factsheets/umlsmeta.html) is a gigabyte+ synthesis of most authoritative biomedical terminologies, now released multiple times per year. Increasingly, a "release" is becoming a report on a time-oriented database.[5] Gradually, the whole notion of a "release" will become less important, and, instead, the Metathesaurus will be seen as a time-oriented record – a no-information-loss history – of authoritative terminologies. Of interest to those trying to deploy solutions on the Semantic Web, in run-time systems, use of incremental "write-once / read-many" databases make locking and error recovery significantly simpler.

We expect that the simple new drug transaction will be the easiest formal unit of change to specify. Quantitatively, the most important unit of change will be a transaction that introduces a change to the definition of an existing medication. For practical reasons the latter transactions are both the most important to accommodate in existing medication order entry systems and the most difficult. Frequently, they affect how drugs are to be used, e.g., the change may be a new contraindication.

Complicating this approach are the presence of larger changes-in-understanding – so-called "lumps" and "splits" - that, seemingly, violate the "axioms" implicit or explicit in DL tools. "Splits" occur when a concept is "split" into two or more concepts, typically because of the emergence of new knowledge. The latter may be new concepts or existing concepts. And the original concept that is split may be retired, or it may be retained as a subsumer of the "split" concepts. Splits are most often prompted by new information system needs, related to the emergence of new knowledge. Similarly, "lumps" – the merging of two previously distinct concepts – is usually prompted by the detection of clerical errors, or by the discovery that two things we thought were different proved not to be. As will be appreciated by those who favor the use of Description Logic (DL) in

[5] Brian Carlsen, personal conversation.

these contexts, a feature of DL, namely its support for formal definitions, helps to decrease the number of inadvertent "missed synonyms". Alternatively, some less mature domains, e.g., Bioinformatics, avoid the problem by using terminologies in which terms are freely lumped or split as needs dictate.

4.3 An ontology of change

If we view a formal terminology or ontology as a corpus of "facts," or assertions, collected over time, then one can contemplate an ontology of such facts, or changes. This much is straightforward. The difficulty is defining and implementing the semantics to be attached to each type of "change unit." One step toward such semantics is the simple expedient of tagging each terminologic unit – concept, term, relationship, and attribute - with a "Start Date" and (any) "End Date"; then, in principle, an application can know the state of the terminology at any point in time. More disciplined and complete forms of such semantics are what are needed to preserve the longitudinal functionality of systems that use the ontology, and what will be needed to transfer knowledge gained from a successful test of the new drug transaction to the Semantic Web.

In the MEDLINE - "rewriting history" - example described above, semi-automated methods accommodate the effects of new concepts, retired concepts, split concepts and lumped concepts in MeSH, as best as can be done each year. Thus, one "blunt instrument" approach to the analogous problem in the Semantic Web is for every repository of historical information to have a companion "warehouse" that is consistent with the current relevant ontologies. The semantics of change are then implemented in the potentially frequent re-computation of this warehouse, as appropriate. The companion argument here is that so-called Clinical Data Repositories (CDRs) and some biomedical research databases are being implemented as "write-only" databases because they represent the authoritative archive of record. Any so-called "data-healing" is done outside the CDR in adjacent data warehouses that are built from queries that run against the authoritative archive. Such pragmatics may evolve into functional requirements for the Semantic Web.

Regardless, the challenge posed by "ontologizing" these units of change is to represent what, for example, should be inherited or shared by other units. Thus, the new drug transaction is a specialized version of the change transaction and thus should inherit any properties of the former. At present, it is not clear how "split" and "lump" should be handled, formally.

4.4 "Perfection Seeking"

While the notion of "perfection seeking" has been very helpful in that it helps those in an inter-disciplinary project "satisfice" in particular domains so as to make progress toward the over-all goal, it has not yet been formalized, e.g., in the form of a metric. At present, terminology and terminology process are bereft of quality metrics. One exception is some work by Campbell, et al., that measured the degree to which lexically implied subsumption (one term appearing within, or sharing sub-strings with, another term) had been represented logically, i.e., in DL, in a large healthcare terminology. [17] While the metric was aimed at measuring the yield of "lexically suggested logical

closure" it also revealed the degree to which the lexical suggestions were not converted to logical relationships, e.g., because of linguistic ambiguity.

A related hypothesis was that expressed by Blois, namely that conceptualization and naming was more stable and predictable at "lower" biological levels, e.g., for molecules. [18] Thus, we would expect fewer synonyms and fewer changes to the Chemical Structure portion of the formal definitions of ingredients.

The fact remains, however, that we've yet to "formalize" (or even measure) perfection-seeking to any useful degree. It is still an entirely human process. However, there is some evidence that tools can aid formalization and while doing so improve conceptualization. [19] [20] Specifically, when a user, in this case a physician, is given the task of entering a formal term for a patient "problem," an interface that displays potentially related formal terms in response to the input of a casual term can help the user better conceptualize the concept being entered. Thus, even when the user interface returns an exact equivalent for the casual term, users may choose a "better" formal term from the displayed semantic neighborhood. The simple explanation for this phenomenon is that humans are better at recognition than recall. Those developing ontologies will be familiar with the phenomenon; once domain experts can "see" a domain model they can almost always make it better.

4.5 Architecture, tools and the local enhancement problem

Implicit in much that has been written here is the architectural notion of vocabulary servers, or in this context, formal terminology or ontology servers. That is, such servers "normalize" terminology functions for enterprises, some at Web scale. [See for example the National Cancer Institute's EVS (Enterprise Vocabulary Server) http://ncievs.nci.nih.gov/NCI-Metaphrase.html] We believe that such servers will be essential to the support of the Semantic Web, and as usual on the Web, the challenge will be how to maintain them in loose synchrony as appropriate.

A clear result of experience to date shows that terminology development, especially formal terminology development cannot be undertaken for long without non-trivial supporting tools and software. Foremost among the required tools is a scalable terminology editing workstation, one evolutionary sequence of which was begun by Mays, et al. [21] The fact that formal terminologies will almost always be constructed and maintained by geographically separated domain experts implies additional requirements for "configuration management," conflict resolution, and the like. One approach to these problems is described in [22]. Further, experience in both the U.S. and United Kingdom has shown that the rate-limiting factor for large-scale terminology development is workflow management, rather than the editing work itself.

One short-term reality is the need for what we call "local enhancement." In the healthcare domain, enterprises will have some locally produced, i.e. "locally mixed and prepared," medications for the foreseeable future, and academic medical centers will always have new terms and concepts in substantive use locally. For these and other reasons, an authoritative reference terminology will need to be enhanced locally. The so-called "update paradox" is that those who add the greatest quantity of local enhancements incur the greatest maintenance burden as the external terminology authority evolves. This tradeoff is made more complex by external reimbursement and reporting requirements.

5. Additional exemplars - reference terminologies and semantic messages in healthcare and biomedicine

In response to the shortfalls of current authoritative biomedical terminologies a number of efforts are underway focused on the development of so-called "principled" reference terminologies. For the purposes of this paper the "principles" in question are those that are computer-empowering, indeed the whole point of a reference terminology is to empower computers, particularly, as with the Semantic Web, to empower computer-to-computer interoperation. Several examples are represented in Figure 2.

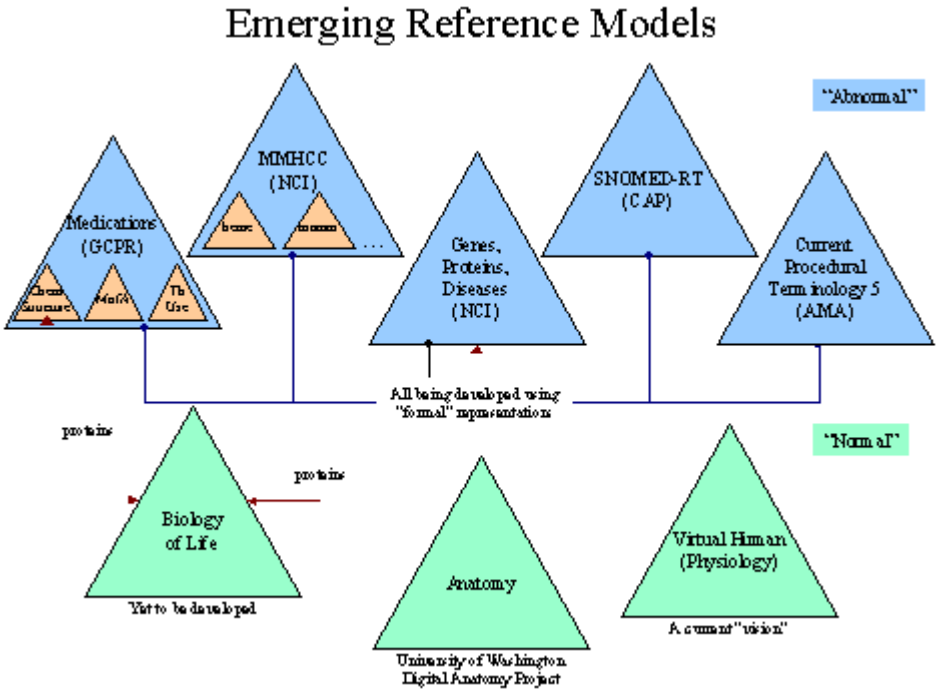

Figure 2 – Emerging Reference Terminologies in Biomedicine: The GCPRMedications reference terminology defined some 1,000 medication active ingredients in terms of Chemical Structure Class, Mechanism of Action, and Therapeutic Use. The NCI MMHCC (Mouse Models of Human Cancer Consortium) is developing detailed diagnostic terminologies for eight organ sites in mice, as a prelude to "certification" of the models as representative of human cancer behavior. NCI is also "modeling" about 250 genes known to be associated with cancer; in particular the association between these genes, the proteins they produce (or do not produce), and diseases is being made explicit. SNOMED-RT is a large (100K+ concept) effort by CAP (College of American Pathologists) and Kaiser Permanente Healthcare to "formalize" SNOMED International (SNOMED = Systematic Nomenclature of Medicine). The AMA (American Medical Association) is formalizing CPT-4 (Current Procedural Terminology). Each of these efforts employs a Description-Logic-based representation. The modular approach implied by this repertoire of reference terminologies in turn creates a need for a reference terminology for Biology that would represent the considerable commonality in, for instance, mice and humans. Similarly, a formal model of human anatomy being developed by Rosse, et al., at the University of Washington may evolve into a reference terminology for vertebrate anatomy as a way to, again, capture inter-species commonality for reuse in other models. A terminology model of Physiology, now being contemplated by some groups, may represent another piece of the "normal" reference model. Not shown is a laboratory testing method terminology being developed by the CDC (Centers for Disease Control and Prevention) .[23]

As recently as a few years ago such a (relative) "explosion" of formal terminology efforts would have been inconceivable. Now such efforts are taking on, in specific domains, the challenge implied by the Semantic Web, namely the development of ontologies for specified domains. Early versions of some of these terminologies are being deployed this year.

HL7, version 3 (http://www.hl7.org/page.cfm?p=524), is a proposed standard for semantic messages in healthcare. It builds on the widely deployed HL7, version 2, standard syntax by using "value sets" taken from external, authoritative, formal terminologies.

6. Summary – healthcare and biomedicine are a rehearsal for the Semantic Web

We are building on our experience with the use of formalization processes for update management in critical working systems. We believe that the challenges we face are specialized equivalents of challenges to be faced by Semantic Web developers as more and more sophisticated systems are deployed and become critical. Among other things these experiences reveal the critical role of process, and that this process needs to be made explicit and intrinsic. We are attempting to fulfill this requirement through the development of an ontology of change, and a recognition that process is more important than representation. If successful, the Semantic Web community may be able to generalize this ontology sufficiently to allow it to be migrated into the "horizontal" Semantic Web infrastructure, and support a "perfection-seeking" Semantic Web.

Acknowledgements

This work was partially supported by Contracts with NLM (National Library of Medicine), NCI (National Cancer Institute), and GCPR (Government Computer-based Patient Record). The GCPR (Government Computer-based Patient Record) medication Reference Terminology Model project team also included Ha Nguyen and Joanne Wong, University of California Berkeley, Munn Maung, University of California Davis, Maung Than, Tun Tun Naing, Apelon, Richard Dixon, MD, FACP and Joe Awad, MD, Vanderbilt. Also contributing were Betsy Humphreys, MLS, William T. Hole, MD, Suresh Srinivasan, PhD, Alexa McCray, PhD, Frank Hartel, PhD, Sherri De Coronado, Robert Cardiff, MD, PhD, Scott Kogan, MD, Mark Erlbaum, MD, David Sperzel, MD, David Sherertz, Brian Carlsen, Cornelius Rosse, MD, DrSc, and Randy Levin, MD, and several others at the FDA (Food and Drug Administration), though as customary they should not be held responsible. Eric Mays, PhD and Bob Dionne from Apelon clarified some of the utility of Description Logic. Important feedback regarding the utility of the model presented in Figure 1 was received from the HL7 Vocabulary Committee and from NCVHS (National Committee on Vital and Health Statistics), including assistance from Jeff Blair, Stan Huff, MD, Christopher Chute, MD, DrPH, James Cimino, MD, Jeff Blair, Ed Hammond, PhD, Michael Fitzmaurice, PhD, Joan Kapusnick-Uner, PharmD, and William Braithwaite, MD. Related material was presented at Pharmacology Grand Rounds Vanderbilt University, Informatics Seminars at the University of Utah, the University of California San Francisco, and the University of California Davis, and also at the Department of Defense, the FDA (Federal Drug Administration), the HL7 (Health Level 7) Vocabulary Standards Committee, and TEPR (Toward Electronic Patient Record) 2001.

References

[1] Berners-Lee, T., "Metadata Architecture," W3C, January 6, 1997.

[2] Berners-Lee, T., Hendler, James, Lassila, Dra, "The Semantic Web," Scientific American, May, 2001.

[3] Tuttle, MS, Information technology outside health care: what does it matter to us? J Am Med Inform Assoc. 1999 Sep-Oct;6(5):354-60.

[4] Kuhn, KA, Guise, DA, "Review: From Hospital Information Systems to Health Information Systems – Problems, Challenges, Perspectives," Haux, R, Kulikowski, C., Editors, *Yearbook of Medical Informatics 2001*, Schattuer, pp 63-76.

[5] Brodie, Michael L., M. Stonebraker, *Migrating Legacy Systems: Gateways, Interfaces and the Incremental Approach*, Freeman, 1995.

[6] Kohn, L., J. Corrigan, and M. Donaldson, *To Err is Human: Building a Safer Health System*. 2000, Institute of Medicine Report, Washington, DC: National Academy Press, 2000.

[7] Blois, M.S., ***Information and Medicine: The Nature of Medical Descriptions*** (University of California Press, 1984), p. 284.

[8] Carter, J., et al., "The Creation and Use of a Reference Terminology for Inter-Agency Computer-based Patient Records: The GCPR RTM Demonstration Project," Fall AMIA Symposium, 2001 (to be presented).

[9] Cimino, JJ, An approach to coping with the annual changes in ICD9-CM., Methods Inf Med. 1996 Sep;35(3):220.

[10] Tuttle, MS, Nelson, SJ, A poor precedent. Methods Inf Med. 1996 Sep;35(3):211-7. (A companion article to [9] above.)

[11] Michael Stonebraker. ``The POSTGRES Next-Generation Database Management System", ***CACM***, 34(10)78-93, October 1991. (Available on-line at http://www.acm.org/pubs/articles/journals/cacm/1991-34-10/p78-stonebraker/p78-stonebraker.pdf.)

[12] Snodgrass, RT, ed., *The TSQL2 Temporal Query Language*, Kluwer, Boston, 1995. http://www.wkap.nl/kapis/CGI-BIN/WORLD/book.htm?0-7923-9614-6

[13] Snodgrass, RT, *Developing Time-Oriented Database Applications in SQL*, Morgan Kaufman, 1999.

[14] Randy Levin, MD, personal conversation.

[15] Michael Stonebraker, "The Design of the POSTGRES Storage System," Proceedings of the 13th VLDB Conference, Brighton 1987, pp. 289-300. (Available on-line at http://www.informatik.uni-trier.de/~ley/db/conf/vldb/Stonebraker87.html.)

[16] Suarez-Munist, O, et al., "MEME-II supports the cooperative management of terminology. Proc AMIA Annu Fall Symp. 1996;:84-8.

[17] Campbell, KE, Tuttle, MS, Spackman, KA, A "lexically-suggested logical closure" metric for medical terminology maturity. Proc AMIA Symp 1998;:785-9.

[18] Blois, MS, Medicine and the nature of vertical reasoning. N Engl J Med. 1988 Mar 31;318(13):847-51.

[19] Tuttle, MS, et al., "Metaphrase: an aid to the clinical conceptualization and formalization of patient problems in healthcare enterprises" Methods Inf Med. 1998 Nov;37(4-5):373-83.

[20] Cimino, JJ, Patel, VL, Kushniruk, AW, Studying the human-computer-terminology interface. J Am Med Inform Assoc. 2001 Mar-Apr;8(2):163-73.

[21] Mays, E, et al., Scalable and expressive medical terminologies. Proc AMIA Annu Fall Symp. 1996;:259-63.

[22] Campbell, KE, *Distributed Development of a Logic-Based Controlled Medical Terminology* (CS-TR-97-1596). Dissertation available from elib.Stanford.ledu, Stanford: Stanford University.

[23] Steindel, Steven, Granade, SE, "Monitoring Quality Requires Knowing Similarity: The NICLTS Experience," Fall AMIA Symposium, 2001 (to be presented).