THE CANCER LETTER

Vol. 37 No. 16
April 22, 2011

PO Box 9905 Washington DC 20016 Telephone 202-362-1809

Bioinformatics:

Cooperative Groups Start To Deploy Medidata Rave, Despite Contract Dispute

By Paul Goldberg

Despite uncertain future of the caBIG bioinformatics venture, NCI appears to be going forward with the deployment of its key component—the Medidata Rave data capture system.

The institute has been working to acquire a robust, commercially available system since 2007, but these efforts bogged down in a contracting dispute.

Initially, the plan was to license the system and procure support from the same vendor in a single $24.3 million procurement.

However, late last year, after losing two administrative appeals waged by a competitor that failed to get the deal, the institute canceled the procurement

(Continued to page 2)

Guest Editorial:

caBIG Has Another Fundamental Problem: It Relies On "Incoherent" Messaging Standard

By Barry Smith

As reported in the March 18 issue of The Cancer Letter, an ad hoc working group of the NCI Board of Scientific Advisors, chaired by Andrea Califano, recently completed a highly critical review of the NCI's Cancer Biomedical Informatics Grid.

The principal recommendation of the Califano working group is that NCI should institute an immediate moratorium on all internal and commercial contractor-based software development projects and that caBIG should "immediately discontinue all strategic efforts to develop and maintain its own brand of software tools." Instead, it should "refocus on its original mission," namely:

(a) helping define standards for interoperability and data exchange,

(b) working with the academic and for-profit communities to facilitate the integration and adoption of these standards into clinical and basic science research software, and

(c) supporting valuable academic software tools that have a proven track record of scientific innovation in cancer research.

The Califano Report is, I believe, an extremely valuable contribution to our understanding of what has gone wrong with caBIG.

However, as one who has followed caBIG's contributions to standards development for some years, I believe that there is a potentially crucial

(Continued to page 5)

Guest Editorial:

V3 Lexicon for Living Subject: "An Organism, Alive or Not"

(Continued from page 1)

factor that is not addressed in this report, which the NCI would be wise to consider in weighing its future actions. This missing factor pertains to the role, in caBIG activities and leadership, of the HL7 organization, a grouping of medical IT specialists which has played an extraordinarily important role in medical information technology standards thus far, and which has long served as the global authority on point-to-point healthcare messaging standards.

HL7's V2 (version 2) set of messaging standards is still used in very many healthcare institutions worldwide. Recently, for example, a number of V2 standards were mandated by the U.S. government as part of the "meaningful use" criteria which will be used in allocating funds for support of Electronic Health Record technology within the "Obamacare" framework. These include, for example, V2 standards for the submission of surveillance information and lab results to public health agencies and to immunization registries.

The problem is that HL7's V2 artifacts, and the point-to-point messaging scenarios with which they are associated, belong to an outmoded technology for healthcare information exchange. As is already clear in bioinformatics research, and as is becoming increasingly clear in the more technologically advanced areas of clinical diagnosis, data in the future will be exchanged not in a walkie-talkie style between pairs of transmitters and receivers, but rather by being deposited on the Web, where it can be accessed by an open-ended set of suitably authorized users.

To support genuinely meaningful use, including effective integration and meta-analysis, such data must be posted to the Web in such a way as to conform to standards which ensure interoperability. But as HL7 itself has recognized, the V2 approach falls short when measured against this requirement.

HL7's response to this problem was its Version 3 set of standards, first released in a 'Normative Edition' in 2005. The underlying strategy of V3 remains the same—it is to produce message specifications.

But a new dimension is added, the so-called "Reference Information Model" or RIM, to which all HL7 artifacts in the future will be required to conform. By this means, HL7 claims, semantic interoperability—interoperability on the level of meaning—will be ensured, and HL7 will be able to retain its place as a leading international healthcare standards body.

The first problem which arises here turns on the false assumption that systems in labs or hospitals exchanging data had no models of the data of their own.

Thus HL7 had to impose such a model from the outside. The reality today, however, is that systems and organizations have sophisticated and often widely shared models of content and workflow (sometimes derived from HLV V2). And if they do not, they aren't interested in messages as a way to define common content—hence the remarkable failure of uptake of HL7 V3 wherever it has been tried, and even in the face of centrally imposed government mandates. The second problem is that the RIM itself, and the V3 standards built on its basis, seem to have been badly and idiosyncratically designed.

Consider the V3 treatment of its own basic terms, definitions of which are provided in HL7's current Normative Ballots. Under "RIM Reference Information Model," for example, we find the following definition of **LivingSubject** as:

"An organism, alive or not."

There is also something called "a **NonPersonLivingSubject**." It's defined as:

"A subtype of LivingSubject that includes all living things except the species homo sapiens."

Under "Personnel Management Domain", however, we find a different, and logically apparently incompatible definition of **LivingSubject** as:

"The abstract super-class (parent class) of Person and Animal."

We also find this definition of **Person**:

"A subtype of **LivingSubject** representing single human being who, in the context of the Personnel Management domain, must also be uniquely identifiable through one or more legal documents (e.g. Driver's License, Birth Certificate, etc.)."

And here is the definition of **Capability:** "A quantitative, semi-quantitative, or qualitative assessment of the ability of an instance of a Party in a Role to perform and/or participate in a specific activity or task. An a subtype of Credential, a Capability is assigned to an instance of a Party (the instance with the ability) by another instance of Party (the Person or Organization assessing the capability). A Capability is often a specific (e.g. 'can do this task' (Person), or 'has 5 slots available' (Organization)."

This definition has been formulated in exactly this manner for at least the last eight years.

These failures are important, given that the RIM is put forward by HL7 as guaranteeing what is called

'semantic interoperability'—which means consistency at the level of meanings. HL7, it seems, has failed to achieve such consistency even within its own documentation.

As one leading Asian medical IT specialist expressed it: V3 "…is an unnecessarily complex, incoherent and confusing messaging standard, while V2 is simple, workable, elegant and deployed at more than 95 percentof healthcare institutions wherever HL7 is used."

Computer architects working on the development of the U.K. Connecting for Health (CfH) national health IT infrastructure have identified the mandated use of HL7 V3 as one factor in the $12 billion failure of the CfH program.

We know of no countervailing success stories which would speak on behalf the HL7 V3 approach—to the degree that at least one major international HL7 middleware company is severing its ties with HL7 because it has become clear that the HL7 V3 approach to creating the "final solution" for healthcare interoperability is "doomed to fail." Other bodies, too, such as the Clinical Data Interchange Standards Consortium standards, although still closely tied to HL7 V3, have distanced themselves from at least some of the V3 idiosyncrasy.

The conclusion of the Califano Report can be summarized as follows: that while the caBIG program "has not met expectations for providing software systems for use by end users," it has nonetheless "had a positive impact on creating common data models and other standards for clinical data management."

Given that many caBIG efforts have been and are still closely tied to the V3 approach, I believe that this conclusion is insufficiently nuanced. Indeed I believe that the very problems affecting the caBIG architecture and software systems may be laid at the door of the V3 approach to common data models and standards which underlies them. It would be wise, therefore, for the NCI to reconsider also the approach to standards taken in the caBIG program, and to recommend new directions there as well.

Barry Smith is Julian Park Distinguished Professor of Philosophy in the University at Buffalo and Research Scientist in the New York State Center of Excellence in Bioinformatics and Life Sciences. He is Director of the National Center for Ontological Research, a principal scientist of the National Center for Biomedical Ontology and a Coordinating Editor of the Open Biomedical Ontologies Foundry. See http://ontology.buffalo.edu/smith

Professional Societies:

ASCO Conquer Cancer Foundation Awards $6 Million to Researchers

The **ASCO CONQUER CANCER FOUNDATION** will award $6 million in new research grants to 197 recipients at the ASCO Annual Meeting in Chicago this June.

The foundation's Young Investigator Award provides funding to promising investigators who are within the last two years of their final subspecialty training at an academic institution, aiding in their transition from a fellowship program to a faculty appointment. This year's 47 awardees will each receive a one-year grant of $50,000 to fund their investigative studies as they begin their careers.

The first Jane C. Wright, MD, Young Investigator Award will be presented this year. Wright is one of the original founders of ASCO. The award's inaugural recipient is **Jung-min Lee**, of NCI and NIH, studying Sequence Specific DNA Damage with PARP Inhibition and Carboplatin.

The full list of this year's Young Investigator Award recipients and their research projects can be found here: http://bit.ly/idTM8J.

The International Development and Education Award provides opportunities for early-career oncologists in low- and middle-income countries to further their knowledge and careers and build long-term mentoring relationships with established oncology researchers.

The awards enable recipients to attend the annual meeting, visit a leading cancer center in the U.S. or Canada with an appointed mentor, and receive a complimentary three-year ASCO membership. This year, 24 awardees are participating in the IDEA program, including four oncologists focusing on palliative care.

The full list of this year's IDEA recipients can be found here: http://bit.ly/gAg2xs.

The Long-term International Fellowship provides young oncologists in low- and middle-income countries a one-year fellowship with a U.S. or Canadian colleague at the colleague's institution. The LIFe award was developed in 2009 in memory of Lina Cassol, who was killed in a plane crash in 2007 in Sao Paolo, Brazil, before she was able to join her IDEA mentor in the U.S. for her fellowship.

The 2011 recipient is **Parisa Karimi**, of the Tehran University of Medical Sciences.

The foundation will present 100 Merit Awards and two Special Merit Awards to oncology fellows