Reviewer's report

Title: The CAP cancer protocols - A case study of caCORE based data standards implementation to integrate with the Cancer Biomedical Informatics Grid

Version: 2 Date: 11 April 2006

Reviewer: John Sinard

Reviewer's report:

I have to begin by admitting that much of this work was beyond my ability to truly appreciate. I do not question, in the least, the amount of work that clearly went into this project, and I think others could benefit from learning about this work, but I am concerned that this paper does not adequately communicate it to the reader. It is both very long and the content is very dense, using esoteric terminology which is going to discourage, I fear, the vast majority of readers. My biggest problem with the manuscript, and honestly why it has taken me so long to review it, is that I have read it multiple times, trying to understand what the “message” is. This manuscript is does not propose any hypothesis which is subsequently tested or evaluated, but rather describes the arduous process of adapting a CAP cancer protocol into a caBIG framework. Although separate Methods, Results, and Discussion sections are provided, most of the quite lengthy manuscript is actually methods. This is not necessarily a bad thing in the Informatics domain, since “process” constitutes a lot of what informatics is, but it does leave me wondering why one is bothering to do this if it is so difficult unless it provides some advantages. Possible uses and advantages to this approach are suggested in the discussion, but actual use of the resulting data model is never described, nor is there any data supporting the claim that this method of modeling the protocols has any real advantages over the existing protocols, or some other data-element persistent implementation which is not burdened by the requirement that every component be an element in the NCI Thesaurus. In addition, the manuscript seems to have had some trouble defining its audience. The introduction is well written and introduces the caBIG project, and some of its components, to the uninitiated, making it appropriate for even non-informatics people. The manuscript then dives pretty deep into ontologies, vocabularies, and modeling, and is sufficiently esoteric to be appreciated only by a rather different audience. Personally, I feel that this work would benefit from selecting one of the two audiences. If the former, more diagrams and a toned down description of the process using vernacular terminology would be better received. If the latter, then much of the introductory material could be eliminated (and I would definitely not be an appropriate reviewer!)

I feel bad that I cannot provide a more critical review of this work… it is simply too far outside my area(s) of expertise for me to comment intelligently, and I fear I will have to defer to alternate reviewers. However, there are a few concrete things I can point out which need to be addressed.

caCORE is never really defined. I think it is one of those acronyms that doesn’t actually stand for anything, but still should be defined.

The figures have some problems. The legend for Figure 2 seems to refer to what is actually figure 3, and the legend for figure 3 seems to refer to what is in figure 2. The text in page 26 refers to figure 3 two times, but the first reference seems to be talking about something which is not in figure 3 at all, or any other figure – most likely a figure which has been subsequently deleted. Same for the reference on page 30. Finally, figure 3, as well as figure 2, is labeled “Fig 2”.

Many typos. Examples: Page 11 “Layer II – information mModels”; Page 12 “…it has NSC Number is ‘007’”; page 42 “…Cerner/DHTI to …”; several others. Page 28 refers to five CDEs being created for
CutaneousMelanomaNegativeSurgicalMargin, but only two are listed in the figure.

**What next?**: Unable to decide on acceptance or rejection until the authors have responded to the major compulsory revisions

**Level of interest**: An article whose findings are important to those with closely related research interests

**Quality of written English**: Acceptable

**Statistical review**: No

**Declaration of competing interests**: I declare that I have no competing interests.