Reviewer's report

Title: The Cancer Translational Research Informatics Platform

Version: 3 Date: 4 April 2008

Reviewer: Jochen R Moehr

Reviewer's report:

In assessing the software, we strongly encourage you to test it: please let us know if you have problems accessing it. Please consider the following points before writing your review:

1. Does the software address a novel task? Alternatively, if there is already software available that performs this task, does the software outperform it in terms of speed, reliability, efficiency, or breadth of application?
   This is a very innovative complex software package. It is not intended for individual users so all questions relating to the reviewer testing the software are not applicable.

2. Is it easy to use?
   See above.

3. Does it satisfactorily address the task or application the authors intend?
   The paper is description of a complex integrative software environment to support clinical decision making and research in oncology. The description lays out the architecture of a system that has progressed to phase two of several phases and is at the stage where a defined set of users apply it in a limited domain of knowledge. There are no evaluative components in the paper.

4. Is the software freely available for non-commercial use (note that this is a condition of publication)? And is the availability of the software and any restrictions on use clearly stated in the manuscript?
   As far as I can tell that is so.

5. Does the manuscript clearly describe the problem the software is designed to address?
   Yes

6. Does the manuscript clearly describe how the software is implemented?
   Yes

7. Does the manuscript clearly describe how the software performs and its advantages / limitations over existing applications?
   This is only touched on. The advantages are described essentially as design specifications.
8. Does the manuscript state the software's operating requirements?
   Yes

9. Are the discussion and conclusions of the manuscript well balanced and adequately supported by the data?
   This being a descriptive paper of a complex software design, there is no data collection on performance, ease of use, etc. That will and should be part of a later phase.

10. Do the title and abstract of the manuscript accurately convey what has been found?
    Yes

11. Is the writing acceptable?
    The writing is generally exquisitely clear. There are some minor typos and very few aspects that one could improve.

    • Discretionary Revisions (which are recommendations for improvement but which the author can choose to ignore)
      None

    • Minor Essential Revisions (such as missing labels on figures, or the wrong use of a term, which the author can be trusted to correct)
      The authors introduce the meaning of the abbreviations ER and PR but not of Her-2/neu. Inclusion of a few words to this effect, such as “The human epidermal growth factor receptor 2 (HER-2)/neu (c-erbB-2) gene . . .” would be helpful to readers like me who are neither oncologists nor geneticists.

      On p 16 (my count) the following seems uninformative and redundant:

      However, with this comes the flexibility to perform queries not supported by the “Simple Interface”. For example, the user can combine data in more flexible and complex ways.

      The second sentence does not give new information and a practical example of what the user really can do would be helpful.

      There are numerous minor typos. The manuscript does neither have page count nor line count. So I merely include the text lines with the errors below in the sequence in which they occur in the paper, and let you figure out, where in the text they occur. In the future you might want to provide manuscripts with page and line counts.

      (caBIG) project.1 caTRIP allows isers to query across a number of caBIG data services,
and their corresponding outcomes such as survival. Of importance, caTRIP elies on the pathology annotations; (5) caIntegrator http://gforge.nci.nih.gov/projects/caintegrator), a tool for storing queryng, and analysing translational data, including SNP data.

Allow for the search of available tumor tissue for patients with predicting factors of interest. A typical question to be answered by caTRIP would thus be would be

A primary goal forcaTRIP is to leverage existing technologies wherever possible. These

• Index Service for discovering services on the grid

classes (see Figure 4). In order to construct queries that span multiple classes and associations, the Simple Interface leverages a configuration file is used that describes the

The first step in the security workflow is authentication, which is the process by which a trusted organization claims a user is who they claim to be. (In this sentence, I would use something like “asserts” rather than “claims”)

(http://www.oasisopen.org/committees/download.php/14361/sstc-saml-tech-overview-2.0-draft-08.pdf) (in my version, this link is not highlighted)

When a data field is accessed on behalf of the user, an authorization structure is queried to determine if the user has permission to access it. If not, then a security exception is returned to the client. (Here, the use of “Whether” might be preferable to “if”.)

A trust fabric is instantiated between the distributed query engine service and the domain service such that foreign CDEs is returned to the DQE service with the understanding that the delegated user is not authorized to view them but is authorized to join on them. (Here, “are” might be preferable to “is”.)

One challenge that arises from this is deidentifying datasets that are managed by different groups, which especially effects deidentifying foreign CDEs (MRNs) that are used to link across different datasets. (Should probably be “affects”).

• Major Compulsory Revisions (which the author must respond to before a decision on publication can be reached)

None
Level of interest: An article of outstanding merit and interest in its field

Quality of written English: Needs some language corrections before being published

Statistical review: No, the manuscript does not need to be seen by a statistician.

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