Author's response to reviews

Title: A repository based on a dynamically extensible data model supporting multidisciplinary research in neuroscience

Authors:

Luca Corradi (luca.corradi@unige.it)
Ivan Porro (ivan.porro@unige.it)
Andrea Schenone (andrea.schenone@unige.it)
Parastoo Momeni (parastoo.momeni@ttuhsc.edu)
Raffaele Ferrari (raffaele.ferrari@ttuhsc.edu)
Flavio Nobili (flaviomariano.nobili@hsanmartino.it)
Michela Ferrara (Michela3ferrara@hotmail.it)
Marco Fato (marco.fato@unige.it)
Gabriele Arnulfo (gabriele.arnulfo@unige.it)

Version: 3 Date: 13 April 2012

Author's response to reviews: see over
REVIEWER 1

1. No mention is made about the size of the infrastructure where the tool has been deployed and tested;

The infrastructure size was tailored to the particular experimental projects the repository will support. The project considered involves three centers, only two of them being able to provide computing resources. Therefore, a lightweight infrastructure based on two iRODS nodes (Italy, Texas) and one repository node with the web application (Italy) has been deployed.

2. No analysis is performed on the scaling of the tool with the number of patients and the number of doctors accessing their records; at some point it is said that 15 patients have been stored (a very limited number) but the number of concurrent doctors is not reported; a scale test, even simulated, should be carried out;

The key objective of the paper was describing the data model and the repository architecture as well as the software functions supporting the sharing of data and samples between partners. From a technical perspective, the focus is on the data model used to allow clinical end-users to tailor their own data representations in the repository. Moreover, the platform has been developed to support a specific kind of projects and it is not intended, at the moment, to scale out to very large numbers of concurrent users. However, the infrastructure level has been considered just to provide a distributed data management and is fully based on iRODS that has very sound scalability performances.

3. No mention is done about the support inside iRODS of the SRM standard protocol;

In the revised paper a mention of the integration between iRODS and the SRM protocol is provided.

4. No mention is done about the availability to encrypt/decrypt data and, if yes, by means of which tool; data privacy and confidentiality, which is very important in the domain, is only marginally addressed;

Data are fully anonymized (and also defaced in case of morphological images) and cannot be linked in any way to patient's names. The link between data in the repository and patients is done using unique and anonymous identifiers managed by clinicians. This solution was one of the project requirements and satisfies the directives authors have been asked to follow about data privacy.

5. In the acknowledgements authors say that part of the work has been funded by the DECIDE project; I found on the web that DECIDE uses the gLite middleware which adopts the SRM protocol for the storage elements so the authors should discuss the interoperability between the middleware they used and other middleware (e.g., gLite);

See point 3.

6. No discussion is done about the sustainability of the service and the possible business model for its long term exploitation by health organisations.

The project is not aimed at creating the basis for a sustainable business model. The work is
the result of a particular clinical need that arose from the cooperation between partners. The present collaboration is fully self-sustained by partners. After the first clinical results, authors will evaluate the prosecution of the experiment, thus eventually producing a cost-benefit analysis.
REVIEWER 2

1. While, as indicated above, this report addresses an important topic area, the formulation of the motivating questions underlying this “technical advance” report is not clear. For example, the authors provide little in the way of evidence that there are not existing platforms that could meet the needs of their particular end-user community, or rationale for the specific technologies or approaches they have selected. In addition, they do not clearly articulate the specific information needs they intend to satisfy with their system in a concise and direct manner;

Even though several platforms have been presented in the last years for managing data in neuroscience experiments, there is still room for improvements as regards extensibility and customization. XNAT and HID approaches provide a very sound base for working on these issues and it is far from our intentions to compare our collaborative environment to such powerful and widely adopted platforms. However, for some specific scenarios, more customizable tools are needed to meet users requirements.

Our work is mostly aimed at addressing the needs of small laboratories without any or little computer science expertises. To this goal, the whole data scheme can be created, extended, and modified by non technical users through the same user friendly web interface used to fill in data. Also, particular security and privacy policies should be addressed by highly customizable software environments in regard to the access to proprietary data and sensible clinical data. Moreover, the management of genetic data often requires an integrated access to public databases such as NCBI and Molgen. Finally, especially in experiments including genetics screenings, tools must be provided for managing specimens and samples in local storage fridges. Our platform tries to answer to these needs in order to put non technical researchers from small laboratories in control of data and samples during collaborative experiments.

2. Building upon the prior comment, the authors provide no significant detail concerning how they reviewed the state of the art in the clinical research and translational bioinformatics domains in order to inform the design of their system. This issue is particular evident in the very cursory and incomplete review of the current literature included as part of the report;

The review of the state of the art has been largely extended and completed.

3. Another major issue is that while a driving use case is presented as a means of verifying or validating the potential efficacy and impact of the described system, not rigorous or systematic evaluation of the ability of the platform to address such information needs is presented. Thus, it is hard to conclude that the system is indeed successful in meeting its stated goals;

The platform has been tested from a technical point of view and results have been presented in the Results section. All the functionalities provided to users have been tested successfully. A customized data model has been defined and also modified by clinical end users through the web interface, all data (neuroimaging data, clinical data, genetic data) have been filled in successfully, samples have been managed successfully, complex queries have been performed successfully. A usability validation is in progress to check whether further user requests should be fulfilled. Findings concerning more specifically the
clinical issues faced in the experiment will be presented in a further paper.

4. Finally, as a result of the preceding concerns, it is hard to assume that the conclusions stated by the authors are or are not in fact supported by their experience in implementing and deploying the described system. Furthermore, little in the way of substantive discussion of the limitations of their system or its comparison to other comparable platforms is provided, further limiting the veracity of the stated conclusions.

Actually, our platform is not aimed at covering all the issues arising in managing neuroscience data. It is far from our intention to compare our system to other more general platforms like XNAT and HID. Therefore, we are fully aware of the limitations of our work. However, we tried to explain which specific needs we have addressed and to provide a description of our approach. In the Results section we tried to assess the performances of our system in regard to these specific needs, not considering more general issues. Nevertheless, some specific limitations are addressed in the Discussion section.