Integration of OpenBIS - improving metadata management in Galaxy

Aarif Mohamed Nazeer Batcha\textsuperscript{1}, Guokun Zhang\textsuperscript{1}, Ulrich Mansmann\textsuperscript{1,2}

\textsuperscript{1}Department of Medical Informatics, Biometry and Epidemiology (IBE – LMU), Munich, Germany
\textsuperscript{2}German Cancer Consortium (DKTK) partner site Munich, German Center for Cancer Research (DKFZ) Heidelberg, Germany

\textbf{Background}

The NGS-FabLab, an IT infrastructure for handling Next-Generation Sequencing data, was built for researchers at the university hospital and medical faculty of LMU. Galaxy is used as a core platform for biomedical research projects in many different fields, including variant calling procedure, differential gene expression and copy number analyses. Since the data is usually obtained from different study groups, locations etc., management of meta-data becomes highly mandatory. Although data libraries in galaxy are highly useful in accessing data, it has its limitations when it comes to meta-data management and querying.

In order to overcome this issue, the Open Source Biology Information System (OpenBIS) (Bauch et al.) was integrated into our NGS-FabLab which has strong ability to track, annotate and share data throughout our distributed research projects. We have integrated OpenBIS and Galaxy by using a python based webserver frame Web2Py (Massimo Di Pierro) to maintain a smooth communication between two (Figure 1) using Galaxy API and openBIS API. OpenBIS can efficiently complement Galaxy which is shown in Table 1.

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure1}
\caption{Figure 1. Galaxy – Web2Py – OpenBIS}
\end{figure}

\textbf{How To}

The following example shows the main function of our implementation.
- Run Galaxy workflow on OpenBIS integrated Datasets.
- Metadata Query, Project specific access etc.
- Sync OpenBIS and Galaxy

\textbf{Remarks}

\begin{itemize}
\item Why do this?
- Efforts are seen to improve the ability to manage metadata in Galaxy community (F. Morrewe) to integrate Galaxy workflow in an extensible meta-data management environment by using ISA-tools. However, further information about this implementation is not sufficiently available.
- Galaxy – OpenBIS connection discussion was initiated (M.Kohler GCC2012). We implemented similar idea to integrate OpenBIS as data source tool and access meta information via custom Web2Py web application.
- Why Web2Py?
- Python based, easy to develop web applications
- Easy-to-use web interface for administration
- Open-source
- Pros?
- Projects, Experiments, User and Group roles can be managed in a structured manner
- Samples can be filtered with metadata values (e.g. cases between age 40-70, Samples from two different projects, etc)
- Every sample is created with a metaID
- Any further downstream datasets gets updated to the parent samples (working on it)
- Cons?
- Authentication has to be done everytime. Establishing auto-authentication is still an issue.
- Not-so-good interface, ugly and highly unprofessional scripts yet it works!
\end{itemize}

\textbf{Table 1. OpenBIS Complementing Galaxy}

<table>
<thead>
<tr>
<th>Task</th>
<th>Galaxy</th>
<th>openBIS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Main task</td>
<td>Workflow and User management</td>
<td>Biological data management system</td>
</tr>
<tr>
<td></td>
<td>Data management to an extent</td>
<td>User, Group and project management</td>
</tr>
<tr>
<td>User interface</td>
<td>Web based</td>
<td>Web based</td>
</tr>
<tr>
<td></td>
<td>Python language-based remote APIs</td>
<td>Java language-based remote APIs</td>
</tr>
<tr>
<td>Meta data management</td>
<td>Data Libraries in file system style</td>
<td>Organize the meta-data in a logical structure</td>
</tr>
<tr>
<td>Workflow</td>
<td>Running bioinformatics workflow inside of Galaxy</td>
<td>openBIS can be integrated in other bioinformatic workflow (eg KNIME)</td>
</tr>
<tr>
<td>Authorization and data access</td>
<td>User/group based, sharing data within a group of Galaxy users</td>
<td>Users/groups assigned roles with different privileges to access data (eg: space admin)</td>
</tr>
</tbody>
</table>

\textbf{Reference}


https://bitbucket.org/ibe/web2py_connect