User Study Report II
29-11-07
Yuwei Lin, Meik Poschen, Rob Procter

1. Overview
This report is based on data collected from the follow-up telephone interviews with survey respondents, a small focus group with two workflow users/creators (but not myExperiment users), a face-to-face interview with an experienced workflow user/creator (also an myExperiment user), and a small user-developer meeting. The meeting transcripts are in the appendix.

2. Trust between users
When searching for workflows, most users start with the people they know. This inclination of working with acquaintances could probably be linked with the concern over privacy and confidentiality expressed by some respondents. Some members do not want their online activities to be openly monitored/watched/screened by everyone online; some do not want their workflows to be shared with all (possibly only after publishing); and still others would like to remain invisible on myExperiment (or any other social networking site). This contrasts with the widespread positive discourse of social networking that sharing is beneficial to all. This leads to the importance of trust between members, and suggests we think about how to build and enhance trust between myExperiment users through design, where this is possible. It has to be kept in mind that these issues are socio-technical and are also influenced by users' domain specific research practices.

2. Data Services
The data services used in a workflow present a decisive factor in why users would like to use a workflow or not. Our respondents have pointed out three main reasons for finding a workflow useful: 1) transparency of the service: users would like to know what data comes in and what results might come out of a workflow 2) IPR: if data services are not publicly accessible this may inhibit re-use (including possible restrictions of publishing results) 3) data formats: researchers need to know which data formats are used and which tools (e.g., BeanShell scripts) can be applied to convert them into compatible formats that can be processed. These three problems can be solved through providing detailed descriptions of workflows.

Most respondents recognised this as a matter to do with good practices in the research community and individual workflow creators. As a respondent said, 'It's a compromise how much time a contributor would like to invest to make sharing workflow feasible. If I have to be honest, my own workflows are not documented properly either. My workflow is used by a small group of people. Once a while when I present my work, that's how they know how my workflow works. My documentations are more limited. Sometimes a workflow is simply too complex - too many services and parameters.'

Even so, it would be good if the developers can come up with something to solve this problem. We suggest two solutions: 1) A Wiki page of data services: using a Wiki to create a list of data services describing the types of data they provide, terms of use, data format and data provenance. In doing so, the Wiki page can be updated frequently and capture the dynamic of bio-medical data services. 2) Linking workflows to other documents/files (e.g., slides, papers) to provide more background information of a workflow.

3. Acknowledgements, Licences and IPR
A number of users expressed their concerns with IPR and licensing issues (of workflows and data). Some stated they would be content if acknowledgements for their work were given. At the moment,

myExperiment only gives credit to people within the myExperiment community, but in the future this should be expanded to people outside of myExperiment. The access rights model in the recent myExperiment version further helps in letting users decide with whom they want to share their workflows.

4. A BeanShell repository
This was mentioned in particular by one of the experienced workflow users/creators. BeanShell scripts are used to link services and transform data formats (e.g., transforming sequences into the right formats to be processed). BeanShell scripts are common elements in workflows, but at the moment they are hard to find and share within myExperiment. Our respondent specifically asked for a BeanShell library/repository (a collection of shim services) on myExperiment, because people could then grab existing BeanShell scripts for building their workflows. For researchers who hardly know anything about programming, it's easier to use than having to write conventional Java. And it encourages sharing and reuse of scripts as well.

5. Other workflow formats
The issue raised by workflows written in different languages (e.g., Perl) or created by different editors is whether they can be used right away or need to be rewritten. Some users expressed the view that if myExperiment starts supporting other workflow editors other than Taverna, it would not make much different to them (because they were not familiar with workflows written in other languages or created by other editors). One respondent said: 'For me it would not make too much difference. I can upload my Perl scripts as well, but they have so much dependency – using them you need to know Perl. I can upload them as references. But I don't think they'll be easily recycled. Taverna is much better.'

6. Too few workflows on myExperiment at the moment
Quite a few respondents mentioned that they haven't found any interesting workflows to (re-)use because there is nothing in their specific fields. The number of workflows is quite small currently. We need to encourage sharing workflows. However, a foreseeable problem with the growing number of workflows is management – how to maintain and manage these workflows efficiently.