User Study Report IV
Yuwei Lin, Meik Poschen and Rob Procter
06-06-2008

1. Overview

With the growth of the myExperiment user community, we have shifted the focus of the study away from staged evaluations and surveys to interviews with users. We are currently in the second round of interviews. Subjects are selected on the basis of (a) having been interviewed before or (b) being new to myExperiment. This allows us to track the possibly changing views of experienced myExperiment users ('friends and family') and to gauge initial experiences and expectations of less familiar users. The latter are being recruited via myExperiment and being selected on the basis of their visible activity.

Respondents whom we have interviewed previously include one user from the Netherlands, one from Germany and two from the UK. New respondents include one user from the Netherlands and one from the UK. We have also interviewed an important power user and key member of the myExperiment team from the Netherlands. Most of these interviews lasted for 20-30 minutes, except for the one with the power user and the one with two users at the same time which both took 45 minutes. The table below shows a list of our 2nd phase interviewees and their profiles, all are from the academic sector so far (in the first round we interviewed eight users including the four 2nd time interviewees in the table).

Interviewee	Location	Disciplines	Sector	Joined	Workflows (credited)	Friends	Groups: admin	Groups: member
User 1 (2nd time)	Germany	Bioinformatics	academic	13/08/07	1 (1x)	6	2	0
User 2 (2nd time)	Netherlands	Bioinformatics	academic	25/02/08	7 (6x)	5	1	5
User 3 (2nd time)	UK	Bioinformatics, Bioanalytical Sciences	academic	13/11/07	0 (0x)	0	0	0
User 4 (2nd time)	UK	Bioinformatics, Bioanalytical Sciences	academic	13/11/07	4 (5x)	14	1	4
User 5	Netherlands	Informatics, Medicine	academic	04/03/08	1 (1x)	2	0	1
User 6	UK	Bioinformatics	academic	04/01/08	1 (1x)	4	2	1
User 7	Netherlands	Biology, Bioinformatics	academic	21/07/07	26 (27x)	39	9	5

In this second round of interviews, we have focused on users' experience with workflow sharing, workflow reuse, and how myExperiment can help them. We also asked them about their experiences with other social networking sites and how myExperiment differs from them.

2. Summary of Findings

The interviews suggest that there is an emergent and basic division in the user community between workflow 'end-users' and workflow 'developers'. These categories emerge from discussions between interviewer and interviewee about the types of workflows the interviewees have uploaded:

Q: When you upload a workflow, do you also upload smaller workflows?
A: I am trying to do both. I upload the complete thing and I also upload its parts. The idea I have is that workflow creators will more look for components, therefore in my case the nested workflows. As the end user you probably look for the larger workflows. I guess. In this sense I am a workflow creator
Q: Or the part of the creator is bigger than the part of the workflow user at the moment in your case.
A: Yes. Maybe there is another thing which leads to the specificity: A larger workflow might be too specific for my needs, it might be more worthwhile to look for the smaller parts, to adjust them to my needs. Things in bioinformatics are often so specific, it can be difficult to find the right thing, smaller things are easier to evaluate.

When asked of their experience of browsing and downloading workflows, two users who are both workflow developers seemed to suggest that there is a distinct kind of workflow that a workflow developer favours and nested workflows seemed to be more difficult to be recycled for them:

Q: Have you found something interesting to download?
A1: No – not for me. Not in system biology.
A2: I found a couple, but it's quite a bit effort to see what they really did or didn't do. In the end, I think it was quite close to what I wanted, but not really useful. They're workflows written for a specific purpose. The exact task is not identical.
A1: Some of the interesting workflows are large and complex and because they are large it's difficult to reuse. It's very hard to understand. It needs to get rid of syntactic problems.
A2: Two or three workflows that I was interested in were actually nested ones. You can't get into what was exactly going on. It's a bit frustrating because it looked like a good solution to a problem. But it wasn't available because it's in a nested workflow. I can see what it's doing but couldn't use it. I did notice that specific example. You couldn't directly reuse it.

The emergence of the categories of workflow developer/creator and workflow user may be important because they seem to favour different kinds of workflows: end-users prefer larger workflows ready to be down-loaded and executed while developers prefer modularised workflows (small components) which can be assembled/customised. (I.e., there are users who see myExperiment as a workflow 'supermarket' and others who see it as a 'toolbox'. How are their needs and expectations of myExperiment different?)

We will continue to explore whether this is a significant distinction between types of myExperiment users and consider its implications for myExperiment. For example, it may mean that workflows written for the developer sub-community may not easily cross over into the end-user community and vice-versa. There may, for example, be a need of workflow brokers (e.g., tutors, advisors) who are able to detect what users need and provide suggestions to them, or assemble and customise workflows for them, like how the market of Linux distributions functions. One question to address in the future will be if the community can be further segmented into meaningful sub-groups.

All the people we have interviewed are 'workflow developers' (although one interviewee remarked that he thought he has been changing from a user to a developer). They have been using the site mainly (if not solely) for publishing and disseminating their workflows. Overall, they report that they get few responses from users on myExperiment. For example, a user who has uploaded 27 workflows received one response regarding the workflow only (the other two were students looking for career opportunities). This query was about how to use the workflow (how to use it, the input and the output).

But this query potentially could have been avoided because the workflow creator, the interviewee, did put examples in his descriptions, but the descriptions were hidden in Taverna and/or did not show those examples easily. Another workflow creator thought that probably nobody has used the workflow she uploaded yet. That workflow "was sort of an initial workflow, which I offered to one person, who in the end already got the scripts that I had written previously on his machine, so you can imagine, why should he bother?" Many interviewees also said that they didn't get in touch with the authors when finding an interesting workflow which unfortunately was not reusable. And the reason for this is mainly lack of time and low priority.

Respondents referred to difficulties in reusing workflows because 1) most of the workflows are designed for a specific task within a specific field and this makes it difficult to repurpose a workflow 2) there are no clear descriptions about what the workflow is for and what are the inputs and outputs 3) the workflows on myExperiment are mainly about bioinformatics; no workflows available in other fields (e.g., in Physics). This suggests that workflows may not easily migrate between research domains.

In terms of usability, most respondents are happy with myExperiment although they do think some features can be improved. For instance, there is no delete button for the message inbox (and there is no sent item folder). The site does not support interaction with people beyond myExperiment (e.g., you can't invite people to become a member of myExperiment). Few respondents have used blogs or forums. Some are confused about the term 'research objects' used on myExperiment. It has been suggested to have a FAQs populated from the messages on the (open) mailing list. Our respondents would also like to see myExperiment connecting to other workflow systems and services.

The privacy issue has largely improved, but the features created to protect privacy and data security sometimes cause confusion because some workflows are not downloadable (only available to group members) but still visible. Perhaps this can be solved by adding a note saying 'This workflow is not publicly available'.

Respondents with experience of other social networking sites such as Facebook thought myExperiment is more purposeful than Facebook (i.e., focused on workflows as objects of scientific exchange). That is, they use Facebook occasionally to be in touch with their family members and friends, but they would use myExperiment for work. Although the interface of Facebook is slicker, they would foresee themselves using myExperiment more often. Other respondents who are not frequent users of commercial social networking sites also thought myExperiment has a user-friendly interface. This suggests that it may not be necessary to migrate all functions of commercial social network sites to the specialised myExperiment in order to get scientists to appropriate Web 2.0.

We are in the process of doing a more detailed comparison of respondents' experiences and evolving attitudes and will report on this shortly.

3. Future Plans
We are continuing to conduct interviews and track changing perceptions of myExperiment. We plan to focus on a number of issues such as understanding the decision making processes behind workflow selection and re-use and the social networking side of myExperiment.

To facilitate this, we need to examine the data on the community membership, social structure and forms of participation to see if we can use it to identify particularly informative respondents for interview. This raises the need to consent new users to the use of myExperiment log data for research and to retrospectively consent existing users.

We plan to profile users in more detail to understand if, for example, requirements are discipline-dependent and to see if we can identify the emergence of different roles within the myExperiment community and how these serve to facilitate the mobility of workflows. For example, we are interested to explore the differences between those who take group coordinating responsibilities and to tease out the complex range of factors affecting relationships between workflow generators, workflow intermediaries (if identified) and workflow end-users. To achieve this, we need to do social network analysis on the log data to look for emergent structures and patterns of activities, interactions and workflow mobility, and their changes over time. To do this requires that we retrospectively consent existing users and get consent from new users for the use of their log data.

We are in the process of developing some ideas for further research on the topic of 'trusting workflows' with Antoon and his collaborators from Gent.

4. **Publications and Dissemination**
Rob gave a talk at the recent Cooperative Systems conference (COOP) on 'affording e-Science' which discussed cooperative approaches to building e-Infrastructure, challenges to the sharing and re-use of research resources and drew on myExperiment as an example of a promising innovation in this area.