

Use Case 9 - Bioinformatics

In these examples, underlined elements are based either on the original interview data or on feedback sent by the informants in response to draft versions of the use cases. Key *activity types* are highlighted in *italics* and **relevant ICTs** in **bold**.

Narrative

1. Rana is an experimental biologist at a research university specializing in modelling human diseases by analysing mutated sequences of genes in humans and animals. She leads an academic research project that attempts to find mouse models for human Charcot-Marie-Tooth disease by comparing abnormalities in the human disease with those that appear in mice. While she enjoys a morning coffee in her lab, she *searches* **PubMed**, a key online reference tool she uses almost daily. A few days ago a famous US molecular biology professor mentioned an article and she would like to read it.

2. *Using a combination of author and keyword search*, Rana finds the mentioned paper in **Nucleic Acids Research**, an online journal she is less familiar with. The paper is a collaboration between the US professor's team and Alex, a UK-based bioinformatician, who has performed in-depth statistical analysis of mutated mouse and human gene sequence data. Rana *prints out and reads the paper*, which reassures her that her project would benefit from a similar collaboration with a bioinformatician. While reasonably IT literate, Rana knows that neither her nor the other members of her team would be able to perform this kind of specialised analysis, and therefore she decides to contact Alex.

3. Two weeks later Rana and Alex *meet to discuss the terms of their collaboration*. Alex agrees to get involved by performing advanced statistical analysis of the levels of similarity between gene sequences, and to *co-author a journal article* with Rana's team. Over the following weeks Alex designs a series of data analysis sequences using **Taverna**, a workflow management system based at the University of Manchester. The workflows *identify the data* in the **Ensembl online genome database**, *specify the types of comparison required, define the statistical analysis to be performed*, and the format in which the results are displayed. Alex *saves the designed workflows* in **myExperiment**, a workflow repository system, then *starts running the analysis*. After two days she *emails Rana an outline of the first set of results*, and *uploads the full raw data* on her institution's secure web location, from where she and her collaborators can safely download it.

4. On receiving Alex's email Rana *sets up a project meeting to discuss the first results* of the analysis. Alex is also invited, but she plans to attend another meeting on that day, so she decides to join the conversation via **Skype**. At the conclusion of the meeting Rana *sends Alex a series of questions for her to address in a subsequent analysis cycle*. Alex *redesigns the initial workflow* and runs it in **Taverna** a few more times, with further communication between Alex and Rana's team via **Skype**. After two months they are reasonably happy with the progress of the analysis, and decide to *publish the results* in **Molecular medicine**. In addition to contributing to the main article describing the experiment, Alex also *references the **myExperiment** workflows in the paper*.

5. A few days later Alex is notified via RSS that her latest **myExperiment** workflow has been rated three stars by a user. As she finds out by logging in to the service, the user is Gavin, a clinical

microbiologist at a Russell Group university, who is involved in designing a workflow management system that uses **National Grid Service** computational power to analyse experimental data. Although the volume of data Gavin needs to compile in his research is much larger than that involved in Alex's analysis, he notices that the general structure of the workflow is quite similar to what he needs. As he adapts Alex's workflow and runs a test analysis, Gavin realises that the **Taverna** software is open source, so he can modify the code to make it more usable. He downloads the software from **SourceForge**, modifies the relevant area of the code and creates a patch, which he then contributes back to the repository.

6. Having run a few more tests, Gavin decides to use the **National Grid Service** to analyse a larger batch of data. The first results are very encouraging. After a few more successful iterations using different parameters, Gavin phones a colleague who works on a similar project. They decide to organize a Microbiology High Power Computing panel at the next year's **UK e-Science All Hands Meeting**.

Relevant ICTs

ICT	Comments
Nucleic Acids Research ¹	Nucleic Acids Research is an online journal that publishes the results of leading edge research into physical, chemical, biochemical and biological aspects of nucleic acids and proteins involved in nucleic acid metabolism and/or interactions. In 2005 NAR joined the Open Access initiative, which means that it is no longer necessary to hold a subscription in order to read current NAR content online.
Molecular Medicine ²	Molecular Medicine is an online journal that publishes work in the format of original research articles, review articles, mini-reviews, editorials, and commentaries on emerging concepts in the interdisciplinary field of molecular medicine. The journal strives to understand normal body functioning and disease pathogenesis at the molecular level that allow researchers and physician-scientists to use that knowledge in the design of specific molecular tools for disease diagnosis, treatment, prognosis, and prevention. All content is published free online, ahead of print. Following online acceptance, papers are quickly posted to PubMed, with an e-publication ahead of print citation.
PubMed ³	UK PubMed Central is a freely accessible, UK-based archive of biomedical and health research findings providing a stable, permanent, and free-to-access online digital archive of full-text, peer-reviewed research publications. The resource is based on PubMed, a service of the U.S. National Library of Medicine that includes over 18 million citations from MEDLINE and other life science journals for biomedical articles back to 1948.
Ensembl ⁴	Ensembl is a joint scientific project launched in response to the

¹ <http://nar.oxfordjournals.org/>

² <http://www.molmed.org/>

³ <http://www.ncbi.nlm.nih.gov/pubmed/>

	<p>completion of the Human Genome Project. Its aim is to provide a centralised resource for geneticists, molecular biologists and other researchers studying the genomes of vertebrates. All the data and code produced by the Ensembl project is available for download, and there is also a publicly accessible database server allowing remote access.</p>
Taverna ⁵	<p>The Taverna workbench is a free software tool for designing and executing workflows, created by the myGrid project, and funded through OMII-UK. Taverna allows users to integrate different software tools, including web services from many different domains, such as chemistry, music and social sciences. Taverna provides a desktop authoring environment and enactment engine for scientific workflows. The myExperiment social web site supports finding and sharing of workflows and has special support for Taverna workflows. The Taverna workbench, myExperiment and associated components are developed and maintained by the myGrid team, in collaboration with the open source community.</p>
myExperiment ⁶	<p>myExperiment is a Virtual Research Environment that enables researchers to share digital items associated with their research, in particular to share and execute scientific workflows. Users can browse myExperiment to find publicly shared workflows, but for further access and the ability to upload and share workflows one needs to sign up. Users can download the software and run myExperiment instances on their computers. The open source code is maintained on RubyForge and is available under the BSD licence.</p>
National Grid Service ⁷	<p>The National Grid Service (NGS) provides free grid computing resources and additional services for UK academics. It aims to enable coherent electronic access for UK researchers to all computational and data based resources and facilities required to carry out their research, independent of resource or researcher location. The NGS is funded by two governmental bodies, Engineering and Physical Sciences Research Council and the Joint Information Systems Committee. Part of the NGS remit is to encourage the use of the e-Science Grid particularly by those academics who ordinarily would not have access to Grid type resources.</p>
Skype ⁸	<p>Skype is a software application that allows users to make voice calls over the internet. Calls to other users of the service, and in some countries to free-of-charge numbers, are free, while calls to other landlines and mobile phones can be made for a fee. Additional features include instant messaging, file transfer and</p>

⁴ <http://www.ensembl.org/index.html>

⁵ <http://taverna.sourceforge.net/>

⁶ <http://www.myexperiment.org/>

⁷ <http://www.grid-support.ac.uk>

⁸ <http://www.skype.com>

	video conferencing.
SourceForge ⁹	SourceForge is one of the largest open source software development web sites. The site provides free services that help users build software and share it with a global audience.
UK e-Science All Hands Meeting ¹⁰	The UK e-Science All Hands Meeting provides a meeting place between computing science researchers, researchers from a wide range of discipline domains, and those who support the combined research. The conference is interdisciplinary and almost uniquely provides a forum for meetings across boundaries that would otherwise not take place.

Commentary

A large proportion of this use case is based on information derived from the main interview account. The PubMed search for relevant online journal literature, the research collaboration between an experimental biologist and a bioinformatician, the comparative statistical analysis of mutated mouse and human gene sequence data, and the joint publication of research results linking to the workflows employed to perform the analysis, are all based on interview-related information. However the situation described in the final section of the narrative introduces a semi-fictional element to the scenario. While the informant had not encountered this specific instance of a workflow published by a bioinformatician and reused by a clinical microbiologist, the example was nevertheless perceived as a plausible way of identifying and modifying a workflow to fit one's research needs.

Comments by informants

The informant's comments mainly referred to what in the draft version was perceived as a too wide gap between unrelated research areas. Thus Gavin's adaptation of Alex's workflow to fit his astronomy research was seen as unlikely to happen in real life due to the different methodologies for analysing data associated with the two disciplines. A research area closer to biochemistry was suggested, and as a result in the final version Gavin's speciality was changed to microbiology. Similarly in the first section of the narrative the field of the research paper consulted by Rana was changed to molecular biology. The informant also suggested enhancing the anonymity of the research team by changing the original name of the disease mentioned in the interview.

Other Editorial Considerations

Element	Usage
Links to direct quotes?	No
Year?	No
Month?	No
Time of day?	Yes
Location given?	Yes
Real institutions named?	Yes
Real journals named?	Yes

⁹ <http://sourceforge.net/>

¹⁰ <http://www.allhands.org.uk>

Real conferences named?	Yes
-------------------------	-----