



# At ISMB, U Manchester Releases Biocatalogue, a Web Services Registry and Social Networking Hub

July 06, 2009

**Newsletter:** [BioInform](#)  
[BioInform - July 6, 2009](#)

By [Vivien Marx](#)

**STOCKHOLM, Sweden** – The University of Manchester today launched [Biocatalogue](#), a registry of curated life science web services intended to help scientists easily connect with teams offering web-based bioinformatics resources.

The university announced the first public release of the catalog here at the joint Intelligent Systems for Molecular Biology conference and European Conference on Computational Biology.

The catalog is the product of a collaboration between Carole Goble, who heads the Manchester team, and Rodrigo Lopez and colleagues at the European Bioinformatics Institute. EBI stores the catalog on its servers and performs back-end operational tasks such as data management and monitoring, while the front end development and the curating is done at Manchester.

Goble told *BioInform* prior to the conference that a team of 11 members curates the services in Biocatalogue, which currently holds over 1,060 services and has nearly 50 users. Developers can register their own services through the website, and services are annotated via tags, user comments, and text descriptions. All annotated services and their components, such as operations, inputs, or outputs, are searchable.

There are other registries for web services, such as BioMoby along with many less mature ones for life sciences and the general-purpose Seekda search engine, but some of the these projects lack extensive metadata, she said.

Life sciences researchers increasingly draw on web services to access data, run compute jobs, and perform in silico experiments, and the catalog is intended to help scientists — and software programs — locate the right service for the analysis they need.

This week's launch of Biocatalogue is the "end of the beginning," Goble said, adding that the intention is to keep the catalog a perpetual beta release. "It's an evolving and feature-

expanding software and open to comment and also open to listening to its user base."

As a central repository, she believes it will help scientists who often don't know how to find web services, or how to use them once they locate them.

"If I know where to go, how do I know how to use it? Because often the metadata is very poor. What do the operations mean? What are the constraints on using that service?" Goble asked.

For example, scientists might not know if the services have been recently versioned, how reliable they are, if local mirrors exist, which metrics have been collected about these services, or how often they have been called in a given time period. Scientists will also want to know who among their peers is using a particular service.

"It's a social gathering point for crowd-sourced information about the services," Goble said. More than a registry, "it is an aggregator."

The effort is a sister project of the [My Experiment](#) repository of scientific workflows. "We're using a lot of the same code base and technology," Goble said.

The two resources are linked architecturally and conceptually, since workflows also use web services. Biocatalogue and MyExperiment each have "programmable interfaces to the other," and can be incorporated into [Taverna](#), a workflow platform developed as part of an [open source middleware project](#) in the United Kingdom.

Goble explained that she is talking with scientists at the National Cancer Institute, which is interested in including its catalog of cancer resources in Biocatalogue, and is having similar discussions with other organizations she did not name.

This project is not related to an older project at EBI called biocatalogue, whose site now points to the new Biocatalogue.

### **More Ruby on Rails**

U Manchester's Jiten Bhagat, a Biocatalogue developer, previously worked on MyExperiment and explained that both resources are Ruby on Rails applications.

"Technical and socio-technical elements" have been transferred from MyExperiment to Biocatalogue, he said, including social networking and sharing capabilities, as well as "borrowed" modules for log in, user activities, and versioning.

As in MyExperiment, commercial firms can take the Biocatalogue source code and pull it behind their firewall.

Bhagat and his colleagues are also following web 2.0 principles of "perpetual beta, constant iteration, constant feedback, and improving" the tool, releasing it as soon as possible and getting people playing with it, he said.

As Bhagat works on the back end, Franck Tanoh, a bioinformatician and curator, keeps the user side in mind. Tanoh said he adds information about the web services from an "expert user" perspective, and works with Bhagat to take in community feedback and plug that back into Biocatalogue functionality and architecture. "It's a life cycle of many things,"

on the user side and the developer side, Tanoh said.

The team also pulls in annotations from partner registries such as Seekda and Embrace, a European project that builds computational services in the life sciences. Goble added that the entire Embrace catalog has been absorbed into Biocatalogue.

"We also source metadata through expert curating," which is a combination of manual and automatic curation as well as monitoring the service and obtaining information from users and the service providers, Goble said. "We distinguish between all of this so you can see who said what on the entries."

Behind the scenes, when a provider wants to submit a Simple Open Access Protocol, or SOAP, service, for example, they provide a URL to the service document. "We try to harvest as much information as possible from there; we suggest tags that are relevant." Bhagat said.

The team has a script that does auto-tagging, he said. "If we know that certain services come from BioMoby.org, we tag them automatically with the tag BioMoby," Bhagat said.

The team has built special mining tools for generating tags and for suggesting tags, which also enables consistency and comprehensiveness, Goble said. The automated tags help the curators, providers, and commenting users bring in "extra value" that couldn't be found automatically.

For providers who register a service, they can see the automatically generated tags and amend those if needed, and add metadata about their service when registering or at a later date. "It's as much effort as they are willing to put in," Goble said. Using Biocatalogue to access services does not require registration on the site, but adding a service and providing annotation does, because that information is attributed to the person making statements.

### **No Rudeness**

Goble added that Biocatalogue also has a section for blog-style discussions about services and the team has put anti-spamming and governance mechanisms in place so "you can't be rude" in the comments section.

Users can see which services are favored by other users, harnessing the community discussion, Bhagat said. When users have 25 Blast services to select from, they will select the one with the best metadata, Goble noted.

Goble said that developers are encouraged to not just contribute to the code base but to build applications on top of it.

Some users might never encounter the Biocatalogue itself, since they might be using a web application that incorporates its functionality but keeps the resource hidden.

The project's main funder is the UK's Biotechnology and Biological Sciences Research Council, with some funds from the Joint Information Systems Committee that funds tool development for scientists. A potential revenue stream could occur through vendors who wish to pay to have their services included, Goble said. "We can imagine a kind of jewel

funding model there."

#### Genomeweb system

These settings are generally managed by the web site so you rarely need to consider them.

**Issue Order: 3**

