Hosts: Monash eResearch Centre and MessageLab

Seminar :The Long tail Scientist

Presenter: Prof Carole Goble, Computer Science, University of Manchester

Venue: Seminar Room 135, Building 26 Clayton

Time and Date: Wed 3 August 2011, 5-6pm

Abstract

Big science with big, coordinated and collaborative programmes – the Large Hadron Collider, the Sloan Sky Survey, the Human Genome and its successor the 1000 Genomes project – hogs headlines and fascinates funders. But this big science makes up a small fraction of research being done. Whilst the big journals – Nature, Science - are often the first to publish breakthrough research, work in a vast array of smaller journals still contributes to scientific knowledge. Every day, PhD students and post-docs are slaving away in small labs building up the bulk of scientific data. In disciplines like chemistry, biology and astronomy they are taking advantage of the multitude of public datasets and analytical tools to make their own investigations.

Jim Downing at the Unilever Centre for Molecular Informatics was one of the first to coin the term of "Long Tail Science" – that large numbers of small researcher units is an important concept. We are not just standing on the shoulders of a few giants but standing on the shoulders of a multitude of the average sized. Ten years ago I started up the myGrid e-

Science project (http://www.mygrid.org.uk ) specifically to help the long tail bioinformatician, and later other long tail scientists from other disciplines. And it turns out that the software, services and methods we develop and deploy (the Taverna workflow system, myExperiment, BioCatalogue, SysMO-SEEK, MethodBox) apply just as well to the big science projects.

Are the needs of the long-tail scientist being met by the e-Science community and its funding bodies? What are those needs and do they differ from the big projects? How do we develop and deploy e-Science application and e-infrastructure to the tail that makes an impact and helps the scientists make an impact? What should that e-infrastructure be? Should we encourage a long tail or seek to shorten it? What's the future of the long tail? I'll examine these questions using my personal experience of "e-Science for hypo-science" gathered over the past decade.

Biography Carole Goble is a full professor in the School of Computer Science at the University of Manchester. She researches semantic technologies, distributed systems, data integration and social computing to solve information management problems for life scientists and other scientific disciplines. She directs the myGrid e-Science consortium

(http://www.mygrid.org.uk ) which focuses on automated workflow-based scientific pipelines and e-laboratories for research and researchers. She is well known for turning research into production software and services such as: the Taverna scientific workflow system; the myExperiment crowd-sourced workflow sharing platform; the BioCatalogue

socially curated catalogue of web services; the SEEK data/models sharing platform for pan-European Systems Biology; and the MethodBox for socially sharing statistical methods and survey data. A partner in the UK's Open Middleware Infrastructure Institute and Software Sustainability Institute, she is currently building infrastructure for several EU scientific programmes in Systems Biology, Biodiversity, Astro-Physics, Astronomy and Digital Library preservation. In 2008 she received the inaugural Microsoft Jim Gray award for outstanding contributions to e-Science and in 2010 she was elected a Fellow of the Royal Academy of Engineering.

Contact: Rob Gray 0411022041 or rob.gray@monash.edu