Program Report: Exploiting and providing research data: finding strategies to help researchers

Miriam Blake

Follow this and additional works at: http://digitalrepository.unm.edu/ulls_fsp

Recommended Citation

This Technical Report is brought to you for free and open access by the Scholarly Communication - Departments at UNM Digital Repository. It has been accepted for inclusion in University Libraries & Learning Sciences Faculty Publications by an authorized administrator of UNM Digital Repository. For more information, please contact disc@unm.edu.
On Friday, September 18, 2009, the UNM Libraries hosted a lecture at UNM’s Centennial Science and Engineering Library. The talk brought Malcolm Atkinson, the official “e-Science envoy,” funded by the Engineering & Physical Sciences Research Council (EPSRC) in the UK, and David De Roure, a computer science professor at the University of Southampton, UK and director of the “myExperiment” project. Together the two are visiting a number of institutions across the country, including MIT, Harvard, UNM, CalTech, and Microsoft Research. Bill Michener, who is leading NSF’s DataOne project at UNM, hosted them at the campus.

Atkinson and De Roure are looking at the e-Science landscape as the revolutionary set of policies and practices that are emerging around digital-data and the research lifecycle. They articulated this as a triangular interplay, connecting the 3 areas of research, computer technologies, and infrastructure. Their mission for this trip (and the purpose of the e-Science Envoy, as stated on Atkinson’s project website [http://wikis.nesc.ac.uk/escienvoy/Main_Page](http://wikis.nesc.ac.uk/escienvoy/Main_Page) ) is to champion e-Science across various research communities, to assess needs for further training and education, to review strategies and compare visions, and to stimulate dialog. They are looking at the broad, multidisciplinary range of research and talking to both experts in the field of research data and with users of research data with the goal of refining their own understanding and fostering collaboration.

Atkinson called the current state of the digital-data environment “revolutionary” because of the multiple dimensions, scale, complexity, heterogeneity of systems and data, and the tremendous challenges we face globally. Transformative technologies and exponential amounts of data being produced are presenting researchers with opportunities to revolutionize our communal and personal experiences. He discussed, as an example, the potential for advances in healthcare using mobile sensors that could tell immediately if someone was having a health event, pinpoint their location, and send targeted appropriate help instantaneously, possibly also being delivered though advanced technology-based mechanisms. The sheer volume of very specific data being produced is allowing researchers to see things in the data they could never see before – a good example is the ability for astronomers to see more deeply and profoundly into the universe because telescopes are more precise and powerful than ever before. The digital-data revolution is still in its very early years, and we are just beginning to learn how to understand it and explore it.

David De Roure continued the talk with a discussion of the evolutionary “ramps” necessary for humans to learn. Just like learning to ride a bicycle, the
transformation of science and research data use requires ramping up through easier workflows and basic understanding. He said there is a co-evolution of science and technology, and it is imperative to understand the drivers and niches behind scientific communities and how they work. He believes more and more researchers could benefit from the adroit use of data, and that the “scholarly infrastructure” (education, libraries, IT systems, etc.) should be committed to helping develop, deliver and maintain services (ramps) that support data usefulness.

To this end, De Roure described the arena of scientific workflows, which are a core component of today’s e-Science. He called workflows the “new rock and roll,” saying they are the machinery for coordinating the execution of scientific services and linking together resources, making repetitive work easier. The idea of workflow packages, such as Taverna (http://taverna.sourceforge.net/), is to give researchers tools and methods they can reuse, recycle, and repurpose so that they can focus on doing science instead of writing code and managing data. This is in service to what he called “the Final Mile;” the idea that it is ultimately the endpoint data analysis that really produces new knowledge, and that building ramps, such as workflow tools, enables researchers to reach the final mile more quickly and effectively.

De Roure ended with a demonstration of a web 2.0 tool that has gained significant traction in the research community called “myExperiment” (http://www.myexperiment.org/). This collaborative webspace (which users have resisted labeling as “facebook for scientists”) includes a repository for research workflows/methods, social networking, content aggregation, federation and a virtual research environment. It is all built with open-source and semantic web technologies, and has attracted a community of over 2300 users from around the world. De Roure emphasized that myExperiment is a testament to the idea that “scientists DO share” if the software is built around empowering them rather than forcing them to change and comply.

This, in fact, was take-home message emphasized by both speakers at the end: “survival in the digital-data revolution depends on the speed and appropriateness of the adaptation.” Building it so they will come means making sure you know what they want and how they will use it. I also heard another clear message – there is a revolution going on, and the data-intensive science is at its core; it is just beginning to change the way we work and live, so we would do well to participate in the conversation and learn how to build ramps in our areas of expertise and specialization.

Miriam Blake  
Research Library Director  
Los Alamos National Laboratory  
meblake@lanl.gov