# Research Computing Advisory Committee

**Minutes February 13, 2012 (taken by Erik Deumens)**

**Present:** Paul Avery, Reed Beaman, Mike Conlon, Erik Deumens, Lauren McIntyre, Susan Sinnott

## Reports

### NIH RCMRC proposal

Lauren McIntyre and Mike Conlon, who were both involved in the NIH proposal preparation, with Art Edison as PI, responding to a solicitation to create a Regional Comprehensive Metabolomics Resource Core (RCMRC), reported on the great value that is provided by the Research Computing infrastructure that has been put in place in the past year under the governance of this committee. Because of this infrastructure, such as the existence of the computational cluster running the Galaxy software, the proposal could focus on the primary mission and include relatively minor cost of the order of $100,000 in the total budget of $12.5M to provide for computational needs.

### Research Computing surveys

The survey to gather information about Research Computing services and staff members performance has been sent out with a closing date of Tuesday February 21.

The work on gathering data from the ticket tracking system has started and will also be used as part of the annual performance evaluation of staff members in March.

The system to track grants that are connected to use of Research Computing is also being put in place. A first round of data will be collected from that system for the annual review with the Provost and the VP for Research coming up in April.

## Discussion

### Big data and e-science

General discussion is held about the impact of big data and the needs at UF and other universities to support e-science. The role of the libraries and of Research Computing needs to become clear; it is not clear yet, although there are some good ideas and examples.

One clear role of the libraries is to provide training on how to manage data and do research of literature that used to be in books and journals and now has increasingly moved to the Internet. The way such research is done is very different and requires updated training.

There is a need for the researchers to have access to experts to help manage data and help process it and make sense of it, more as a partner than as a service provider. These experts could be members of the libraries, there are examples at other

universities of such a structure and the UF Libraries have some people working in such a role.

Especially in the area of medical data there is a huge need for much improved infrastructure to deal with data and to increase the meaning of data as opposed to just keep data in inscrutable formats for the sake of keeping records.

Some discussion within the libraries is taking place with regard to open access publishing. The libraries at UF are providing funds for researcher to publish in these journals. There seems to be a notion that this will save precious funds for the library by eventually allowing cancellation of some journals that are not open access. However, the time frame for this may be distant in the future as UF researchers still need access to the journals to read what others publish, even if every UF researcher only publishes in an open access journal.

**Discussion with Computational Biology Committee 2/17/12**

Some members of this committee (Conlon, Beaman, Deumens) had a discussion within the Computational Biology Committee (Jed Keesling, Pramod Khargonekar, Alberto Riva, Marco Salemi) on big data. The concern was expressed that there is a acceptance that performing a sophisticated statistical analysis as long as it is on big data is considered science; the need for developing models and mechanisms to explain what is observed in the data is downplayed. The group considers this a major problem.

The discussion then proceeded to identify that a major need for all researchers working with big data, including those at UF, have a need for expertise in the form of human resources to assist and train in the analysis and processing of big data. This is in addition to the physical resources of storage and computing systems.

What form should be used to keep data so that future research can use it and build on it? Currently large quantities of data are kept, but they are in formats that are not accessible by automated tools to learn anything from the data. Often a significant amount of human intervention is needed to make sense of the data or to make it useful for further research. To turn data into knowledge requires the development of data standards, that may be and should be specific to scientific and scholarly domains.

**Cloud expertise**

The committee discussed the issue of cloud services becoming more powerful and more available in many areas. This leads to more researchers and scholars seeing possible opportunities to use cloud services. The committee supports the idea the UF Research Computing should develop expertise in the form of a staff member who can advise and assist researchers at UF to help find the best cloud services available for certain projects and then to assist in properly and efficiently using such services.

Next meeting will be of the HPC subcommittee on Monday February 27 at 1:30 pm in NPB 2205.