

Data Life Cycle subcommittee (RCAC-DLC)

Minutes January 30, 2012 (taken by Erik Deumens)

Present: Sophia Accord, Erik Deumens, Rolando Milian, Laurie Taylor

Reports

Storage

The HPC Center is working with several groups on campus to provide long-term storage for research data. This is in addition to the 230 TB of fast scratch storage for volatile data used during computational projects. That storage is called scratch storage and is considered of a fleeting nature.

Two Nexenta network attached storage appliances have been acquired and taken into production in January 2012, each with a capacity of 100 TB. Several research groups have bought storage capacity with the Research Computing Matching program that ran July-September 2011. The second unit will provide the data replication, which is one of the options for buying long-term storage and ensure a backup exists.

The HPC Center is also working with the computational biology group to find the specification for storage system to be part of the biocluster. Initial research and data gathering has been completed and a recommendation has been sent to the group for discussion and feedback.

e-Science Institute

Research Computing is a partner of the e-Science Institute. This institute is organized by the Association of Research Libraries (ARL) and the Digital Libraries Federation (DLF). Each participating institution has three representatives to explore the status of e-Science in the institution, then assess the needs, and finally develop a strategic agenda for the institution with regards to e-Science and "Big Data". The three representatives for UF are Cecilia Botero, associate dean of the Health Science Libraries, Diane Bruxvoort, associate dean of the Marsten Science Libraries, and Erik Deumens, director of Research Computing. The team attended a workshop Jan 25-27 in Dallas, TX to draft the strategic agenda.

Research Computing Day April 25

The main topic of Research Computing Day this semester will be a workshop on the basic of data management. It will be organized by Research Computing and the Libraries. This workshop will help the discussion and training on data management, including helping faculty develop and write good data management plans for their proposals. Details will be on the website <http://www.it.ufl.edu/community/events/rcday>. This site now shows the details of the event in the fall of 2011.

Digital Humanities Day April 26

The Center for Humanities and Public Space is organizing a Digital Humanities Day on April 26, the day after Research Computing Day. There will also be a field trip the day after, including an opportunity to visit the HPC Center facilities in Larsen 121. Contact Sophia Acord skacord@ufl.edu for details.

Discussion

OSTP RFI

A response to the request for input from OSTP was formulated and discussed by the committee. See Minutes of Nov 21 meeting for the request. The discussion was very lively and brought some important conflicting requirements to the fore: namely the need to make data available and useable by others takes effort that is not generally funded and conflicts with the desire and need to have time for a research group to work on the data they obtain or generate and produce meaningful results. After the discussion, no response was submitted, but the document produced is useful and will be refined further.

NIH has a similar RFI: <http://grants.nih.gov/grants/guide/notice-files/NOT-OD-12-032.html> with the response due on March 12, 2012.

e-Science Institute strategic agenda

The strategic agenda from the e-Science Institute will be presented as soon as it becomes available. The focus will be a coordinated effort by the Libraries and Research Computing to gather information about data management practices and then document them in best practices and turn these into training materials and workshops for graduate students, researchers, and faculty at UF.

A full strategy for managing data cannot be developed at this time. The subject is too much in flux still. However, some support can be given now and is available for research groups on campus. The Libraries have a service for long-term data preservation and Research Computing offers storage for research data on a more intermediate time scale.

Next meeting will be of the RCAC committee on Monday February 13 at 1:30 pm in NPB 2205.