

Data Life Cycle subcommittee (RCAC-DLC)

Minutes April 16, 2012 (taken by Erik Deumens)

Present: Paul Avery, Erik Deumens, Hannah Norton, Laurie Taylor

Discussion

Data centric solicitations

NSF has two outstanding solicitations that have data and data management as a strong focus.

The first is a multi-agency push for work on data as a “fourth pillar of science” (after experiment, theory, and computation). It was announced on March 29 by President Obama under the name BIGDATA and multiple agencies have solicitations for it. The NSF solicitation can be found at <http://www.nsf.gov/pubs/2012/nsf12499/nsf12499.pdf>.

The second is the solicitation from NSF for Campus Cyberinfrastructure – Network infrastructure and Engineering (CC-NIE) proposals. The solicitation can be found at <http://www.nsf.gov/pubs/2012/nsf12499/nsf12499.pdf>. UF Research Computing is looking into a proposal to fund the upgrade of our connection to the Florida Lambda Rail from 20 Gigabit per second to 100 Gbps, either within the context of UF, or within the context of SSERCA, the sunshine state education and research computing alliance. This proposal crucially needs to show the need for such infrastructure by engaging research projects that are data centric and that need to move data for collaboration or for public sharing across the link to FLR.

The Research Computing Advisory Committee and UF Research Computing are interested in supporting, connecting, and enabling investigators who are working on BIGDATA projects or are planning to submit proposals for such work to help and make their proposals better.

The iDigBio (Integrated Digital Biocollections) project is one such project. It will hold a workshop on developing robust object to image to data workflows on May 30-31 at UF in building 105. <https://www.idigbio.org/content/developing-robust-object-image-data-workflows-workshop>

Digital Libraries Institute

Laurie Taylor presents a brief overview of the activities of UF Libraries as a digital library. The UF digital collections are available at <http://ufdc.ufl.edu/>. The site is powered by open source content management (CM) software called SobekCM (<http://ufdc.ufl.edu/sobekcm>) that is developed locally at UF and used by other sites. The site has about 14 TB available online with about 250 TB of archival material as backup stored on the CNS tape system. The site receives about 4 million hits from

humans and 20 million hits from net-robots. Scholars can upload and process data on the site after authentication through GatorLink.

Next meeting will be of the RCAC committee on Monday May 7 at 1:30 pm in NPB 2205.