

Research Computing Advisory Committee

Minutes Jul 18, 2016 (taken by Erik Deumens)

Present: Many people could not attend. Peter Barnes, Erik Deumens, Lauren McIntyre

Updates

- HiPerGator 2.0 is now in full production and users are in the final stages of migrating their data to the upgraded part of the storage system. Then the old /scratch/lfs will be reformatted and merged into /ufrc to give it the designed size of 3 PB.
- ResVault build has been completed. We are updating documentation and training materials to reflect the final configuration. The security documentation will be written in the next two months to certify that it is FISMA moderate NIST 800-53 and CUI NIST 800-171 compliant.
- All universities in the State have identified the need for cheap archival storage at a cost around \$25 per TB per year. FLR has created a draft of the ITN (Invitation to negotiate) to send to vendors.

Discussion

Archival storage pilot

Lauren McIntyre is working with Plato Smith in the libraries on a pilot project to manage a special kind of research data. The raw data is stored in national repositories. The data supporting a publication is attached to the publication as supplemental data, but this is usually small. There is some data of large size that is produced during an intermediate step. It has value to other researchers, but it is too big (1 TB) for publishers to accept as supplemental material.

The business model needs to be developed, but should work as a deposit fee for storing data and allow it to be retrieved for some time in the future. With an initial period, say 2 years, where retrieval is immediate because the data is live, and after that time, say another 3 to 5 years, there may be a delay and a fee to retrieve the data from tape or some other slower storage system. After 5 to 7 years the intermediate data is no longer pertinent, as new algorithms would provide better analysis of the raw data.

The need for storing data that is intermediate between the raw measurement data and curated publication data is growing in all fields, because of the pressure to increase scientific reproducibility.

B2K going to NLM

NIH accounted that it will move its BD2K (Big Data to Knowledge) initiative to the National Library of Medicine, since that is a more natural place to maintain and extract knowledge from data.

Replication vs. backup

The website needs to be updated with an explanation of what is replication and what is backup. There is confusion about that among many.

Organization

Next meeting will be on Monday September 12, 2016 in the smaller conference room in NPB 2304 from 1:30 – 2:30 pm. Sep. 5th is a holiday.