

Longitudinal Sustainability

Workshop

Gwen Jacobs, University of Hawaii
Bill Michener, University of New Mexico
Tue, 11/07/2017 - 2:00 pm

This workshop explores the benefits and challenges of managing and sharing research data generated through EPSCoR Track 1 research programs. Examples of successful efforts from several jurisdictions are showcased and the panel discusses best practices and how they can be adopted in new and existing programs.

Summary of Panel:

Gwen Jacobs: “A view from the trenches, the science of data management”. Jacobs discusses common themes: data access, integration, and community response; workforce need – CI professionals; and software sustainability. Hawaii’s Track 1 Project is a cyberinfrastructure program, developed to manage, share, and sustain Hawaii’s research data. The program is based off the Agave Platform, modeled after designsafe-CI. Success comes from making software engineers a part of the research team, managing expectations with users, setting standards for data quality and data annotation, and listening to users. Be flexible, adaptable and willing to start over.

Pips Veazey: “Preparing for Successful Team Data Management”. Team Science has fluid boundaries, interdependence, and goal attainment. Alaska is co-creating a data management plan with shared understanding. They start by visualizing their data into larger “buckets” and then into more specific groups and subgroups. This is done using a “collaborative visual environment”. Alaska has what they term as the “decision theater” (dtn.alaska.edu) where co-creation of proposals, grants, data management plans, etc. can occur. Their large goal is to match a proposal with a data plan, and with the strategic plan as well as annual reports.

Ewers: University of Wyoming’s approach to dealing with the challenges of technical issues with data types and reticence to data use is discussed.

- petaLibrary hardware – data in database is increased in orders of magnitude after making it a requirement for students to submit their data for continued funding.
- Automate the process of DOI generation and curating with the UW’s libraries.
- DataCorral is used to manage data internally and externally. Bar codes from initial sample locations to sequences to final figures create many ways to access and view data.
- Data Science Center (DSC) allows an entire jurisdiction to help all entities do better data science.

Michener: “Tools and Approaches for Preserving and Sharing EPSCoR Data”. The general NSF policy on data is, “Data are viewed as products of the research enterprise, supported by tax dollars, and need to be accessible after completion of a project.” New Mexico promotes the discovery of data and works to preserve it through:

- DataONE through the Data Management Planning Tool (DMPTool.org)
- Data portals and user-friendly databases

- Training post-docs and graduate students across the state with “Software Carpentry” to sustain data