Program Development Exercise: Life Sciences

Background

You are the data management specialist at The Higher Ed University. After providing an introductory 1-hour session to the biology department graduate students on data management, you are approached by several members of a neuroscience research team to work with them as they try to improve their data management practices. The team has a mix of masters and PhD students in Professor Jones' lab, working on individual or collaborative projects. The team meets once every two weeks to review progress on projects, share new information, and discuss/solve any team-wide issues. After attending one of the team meetings, Professor Jones asks you to work with the team and encourages all team members to participate, with only minimal limits on time/access (1-2 hrs per person for interviews, and several hours for implementing solutions/training).

Potential areas of need

Based on the team meeting you attended and initial conversations with Jones, you believe that the team has a large array of files that are generated for each study, but lacks organization of the data for cross-project review and analysis. As students finish their research and/or leave the institution, they hand off data sets to the professor but no system exists to keep these organized and described in such a way as to enable search and reuse. Data sets up to now have not been shared when dissertations are deposited with the library, but multiple publications are typically generated from a PhD project. Data management plan requirements are making data sharing a topic of discussion, but where and how and what to share is unclear.

The research team

The research team consists of one primary faculty member, one postdoc, four PhD students, two MS students, and a research technician. Many collaborations occur with other research teams on topics such as nutrition and parasitic impacts on neurological health.

Research summary

The team uses a variety of (non-human) model organisms, such as fruit flies and mice, the team investigates regulation of the structure, function and survival of neurons and neural circuits, using electrophysiology, biophysics, genetics, genomics, behavioral analysis and other approaches. Research is funded primarily by NIH funding but also NSF grants. The team generally seems very supportive of open access, and is exploring options for an open notebook system for collaboration within the lab.

The types of files that this team generates

Genetic sequence data as text/ascii files; SPSS or R for statistical analyses; video (proprietary formats) and still images (.jpg or .tif) of organisms and cells; and data representing measurements of cell growth, neurons, expression of hormones, cell culture histories, and behavioral metrics, all typically recorded in or transcribed to spreadsheets.

Jake Carlson – PI (Purdue), Camille Andrews (Cornell), Marianne Bracke (Purdue), Michael Fosmire (Purdue), Jon Jeffryes (Univ. of MN), Lisa Johnston (Univ. of MN), Megan Sapp Nelson (Purdue), Dean Walton, (Univ. of OR), Brian Westra (Univ. of OR), Sara Wright (Cornell)