Findings from the DIL Interviews: Data Processing and Analysis

Skills in this competency may include:

- Is familiar with the basic data processing and analysis tools and techniques of the discipline or research area.
- Understands the effect that these tools may have on the data.
- Uses appropriate workflow management tools to automate repetitive analysis of data.

Average Ranking of Importance (5=essential): Faculty= 4.63, Students = 4.35

Faculty responses:

This area is generally viewed as a more direct component of doing science in most disciplines, and therefore receives a high ranking of importance by faculty. Overall, this competency was seen as critical for students, not just to avoid mistakes in evaluating data, but to gain efficiency in their work. Several faculty mentioned students being unfamiliar with process and analysis tools in the lab as well as within their discipline.

Faculty estimated that their students' skill levels in this competency ranged from "not systematic" and inefficient, to highly experienced upon entering the programs. One faculty described students as good in this area, but not necessarily efficient, meaning that its takes students longer than it should to perform their tasks. Some workshops and classes were mentioned as potential resources for graduate students, but peer-to-peer contact is still noted as being the most influential component for learning. Another faculty responded that he doesn't typically teach these skills because students typically absorb the material better by engaging with it themselves even though they may fail repeated.

As with many of the competencies, the nature of training is dependent on local and disciplinary practices and culture. Broad trends are difficult to state, but there was an emphasis on developing skills and critical thinking through personal engagement with the data and tools. Some of the pathways to skill acquisition included: peer-to-peer and advisor contacts; formal courses, such as in statistics; and the self-taught/trial-and-error path.

Student responses:

Students are asked to perform a wide variety of tasks in processing and analyzing data. Several students reported being self-taught on the tools they use to perform these tasks. Statistical programs dominate the list of tools that students described (R, SPSS, SAS), as does Excel. In addition, a variety of other programs and tools for collecting and transforming data specific to the particular research domain and project were described. They include ArcGIS, Data loggers, ENVI for analyzing landstat images, MATLAB, various coding languages such as Python and C++, as well as pencil and paper.

As with faculty, students recognize these skills are generally at the core of scientific practice in their domains. One student commented: "That's one of the – I think – biggest mistakes that people make in our field is improperly analyzing data."

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)