Findings from the DIL Interviews: Metadata and Data Description

Skills in this competency may include:

- Understands the rationale for metadata and proficiently annotates and describes data so it can be understood and used by self and others.
- Develops the ability to read and interpret metadata from external disciplinary sources.
- Understands the structure and purpose of ontologies in facilitating better sharing of data.

Additional skills mentioned by an interviewee:

- Individuals who publish research must be ready at any point to answer questions from others regarding the data set.

Average Ranking of Importance (5=essential): Faculty = 4.57, Students = 3.88

Faculty responses:

Students were considered barely proficient or worse, and most faculty feel that this competency is an area that needs improvement. Nearly every faculty interviewed reported that the data was not described well enough for someone outside of the lab to understand and make use of the data; some reported that even they had trouble understanding the data because the students did not document it thoroughly. None use a metadata standard to describe their data.

- One of the interviewed faculty felt that this competency was of primary importance and that much could be gained by addressing the need; he expressed personal interest in learning more because he himself was unsure of the meaning of the term and feels that a lack of knowledge in this area could be damaging.
- Currently, researchers “spend more time doing the work than explaining the work [they] are doing.”
- For ongoing projects in one of the labs, in which code is passed from one group of students to another each semester, current documentation is “definitely” not enough for someone outside of the lab to understand and make use of the data. This is a major issue during project transition between semesters.

Student responses:

Many students claim to have an understanding of metadata, though most have not received any formalized training, and some did not actually provide an accurate definition when pressed to explain what they thought metadata is (two confused it with meta-analysis). Student knowledge of metadata has largely evolved from past projects, trial and error, and even past work in industry for one graduate student. For example, the natural resources graduate student’s methods for describing data have been learned through a “personal coping strategy,” or, in essence, through trial-and-error over time. One graduate student familiar with metadata noted that the metadata she creates is often not detailed because the student “doesn’t have enough time.” Several students reported no trouble understanding the metadata that accompanied the external data. None of the students reported using a metadata standard.