Findings from the DIL Interviews: Ethics and Attribution

Skills in this competency may include:
- Develops an understanding of intellectual property, privacy and confidentiality issues, and the ethos of the discipline when it comes to sharing and administering data.
- Acknowledges data from external sources appropriately.
- Avoids misleading or ambiguous representations when presenting data.

Additional skills mentioned by an interviewee:
- Identifying what data not to show for privacy purposes.

Average Ranking of Importance (5=essential): Faculty = 4.38, Students = 4.35

Faculty responses:
Fewer faculty gravitated towards the data representation element of the competency, focusing more on the citation, intellectual property, privacy and confidentiality elements.

Although citing data was ranked as essential to very important, faculty also stated that standards for citing data within the discipline were lacking. However, most felt that students were good enough at citing data. One of the faculty felt that ethics and attribution are consistently discussed in the lab and at the university, and believes students recognize that ethics extend beyond literature and into data sets. Two of the interviewed faculty felt that students cite outside sources sufficiently. One of them noted that students may not know how to cite a data set versus a piece of literature, and he himself didn’t know of a disciplinary standard for citing data.

Several faculty noted that graduate students receive ethics training, either at the university or departmental level. The majority of the faculty noted that the question of who owns the data is “somewhat shaky” or “up in the air”. One of the faculty we interviewed felt that Privacy and IP issues are covered, but more detailed, practical instruction for handling sensitive data is necessary. Another faculty stated that students need to understand the differences between copyrights, trademarks, and patents.

Student responses:
Several students report citing the research paper associated with a data set rather than a data set itself, although a general feeling of being competent at citing data was expressed by nearly every graduate student interviewed. Encouragingly, students reported receiving university training for ethics and IP issues, although they had mixed attitudes about how useful the training was with regards to data.

One of the computer science graduate students mentioned that the lab seeks software code with the GNU or PSD licenses to ensure that they can properly utilize code generated by other individuals, which aligns with the faculty’s assertion that it is very important that these students understand issues with IP and copyright, patents, etc. Of potential concern, one student asserted that since she consulted external code but never used it outright, she didn’t need to cite it.