Preparing Students for Managing Large-Scale Scientific Data

Dr. Angela P. Murillo[a]

[a]Indiana University, Indianapolis, USAA

apmurill@iu.edu

ABSTRACT

This poster presents preliminary observations and feedback from the pilot year of a student internship program that was created to prepare and train students in managing large-scale scientific data. This student internship program is part of the CI Compass project, which is a National Science Foundation (NSF) funded project to provide support and enhance the data lifecycles of NSF Major Facilities (MFs). MFs are the largest-scale scientific efforts that the NSF supports and can take years to build, cost hundreds of millions of dollars to construct, are highly diverse, have heterogeneous data, and a wide range of cyberinfrastructure for capturing, processing, archiving, and disseminating data. MFs span science domains, including astronomy, climate, ecology, natural hazard, ocean science, physics, and seismology.

Due to the complexity of the cyberinfrastructure and data that supports MFs, it is critical that we create educational opportunities for library and information science students interested in pursuing a career in this specialized large-scale scientific data management. Particularly having conceptual understanding and skills in the full data lifecycle is critical, including data workflows, capture, processing, storage, preservation, and analytics.

During Spring 2022, we commenced with the inaugural student internship program, which included a technical curriculum program in data lifecycle-related skills (i.e., FAIR data, cloud computing) and research skills program (i.e., research design, research ethics). During the Summer 2022, students have the option of an embedded project-based learning experience at selected MFs for additional training. This poster describes the technical curriculum, research program, project-based learning experiences, and future goals.

ALISE RESEARCH TAXONOMY TOPICS

Data Management; Data Curation; Data Science; Education of Information Professionals; Big Data.

AUTHOR KEYWORDS

Scientific data management; data lifecycle; cyberinfrastructure; student internships; program development.