

BUILDING A DATA SCIENCE EDUCATION ECOSYSTEM RESOURCE COLLECTION

Presenter:

Gladys K. Andino, PhD

Senior Computational Scientist

Research Computing

Purdue University

BACKGROUND

- Purdue University has invested heavily in the data sciences education with programs intended to influence ALL programs.
- The Data Science Education Ecosystem has focused on for credit courses or workshops in disciplines ranging from agriculture to digital humanities.
- Large cohorts (hundreds) of students participate programs such as The Data Mine or Data Mind .
- Infrastructure such as the instructional computer labs and clusters must grow and adapt to meet a wide range of needs and skills.
- Purdue University Libraries has a role both in providing resources but also expertise related to managing data.

PROJECT GOALS

1. Create a collection of instructional materials to support the Data Science Education Ecosystem.
2. Improve and expand existing infrastructure to support high-impact instructional material such as Jupyter Notebooks and Shiny learnR tutorials.
3. Engage The Data Mine students to test, improve and create instructional material relevant to their discipline.
4. Engage instructors in the Data Science Education Ecosystem to understand their needs and solicit material for the collection.

RESOURCES AND TOOLS

- Scholar (teaching HPC resources)
- Data Depot (storage linked to HPC resources)
- Purdue GitHub
- Libraries
- Purdue University Research Repository (PURR)

CURRENT PROJECTS

- Interactive R tutorials using learnR package targeted to
- Identifying data sets in PURR that can be used to teach data science.
- Revising Jupyter Notebooks from a graduate-level bioinformatics course sponsored by the NIH.
- Building portal website at Libraries to facilitate sharing and discover of materials.
- Planning stage for videos on how to use Scholar, Thinlinc, Jupyter Notebooks, etc. with an emphasis on how to use these at Purdue.

THE TEAM

Gladys K. Andino, co-PI

Senior Computational Scientist
Bioinformatics Specialist
Co-founder of WHPC at Purdue
Research Computing
Purdue University

Michael Witt, co-PI

Associate Professor
Purdue University Libraries and School
of Information Studies
Data Repositories and Data
Management

Pete E. Pascuzzi, PI

Assistant Professor
Purdue University Libraries and School
of Information Studies
Bioinformatics and Data Science

Mark Ward, co-PI

Professor
Purdue University Department of
Statistics
Statistics and Data Science