Libraries Faculty and Staff Presentations | Purdue Libraries

2019

Integrating Foundational Data Management Course into STEM

Wei Zakharov
liu31@purdue.edu

Sarah Huber
huber47@purdue.edu

Michael Fosmire
fosmrie@purdue.edu

Follow this and additional works at: https://docs.lib.purdue.edu/lib_fspres
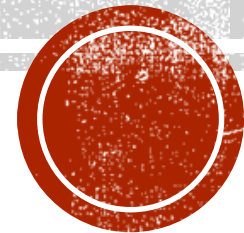
Part of the Library and Information Science Commons

ILS 295 DATA MANAGEMENT

Integrating Foundational Data Management Course into STEM

Wei Zakharov, Assistant Professor
Sarah Huber, Assistant Professor
Michael Fosmire, Professor

PURDUE UNIVERSITY.

Libraries and School
of Information Studies

AGENDA

1 Background

2 Course Setting

3 Take the pulse

4 Discussion

BACKDROP OF OUR PROJECT

Data Science

Interdisciplinary in nature

Integrative Data Science Initiative

The "Embedding Data Science into Domain Curricula" program will serve as a new model to implement a comprehensive educational Data Science initiative. (https://www.purdue.edu/data-science)

Purdue Libraries and School of Information Studies

Data literacy education: Identify needs and integrate data literacy into undergraduate curriculum

SELECTED IDSI EDUCATION GRANT PROPOSALS

https://www.purdue.edu/data-science/education/education-proposals.php

Foundations of the Data Mind: An Interlocking Modules Approach

OVERVIEW

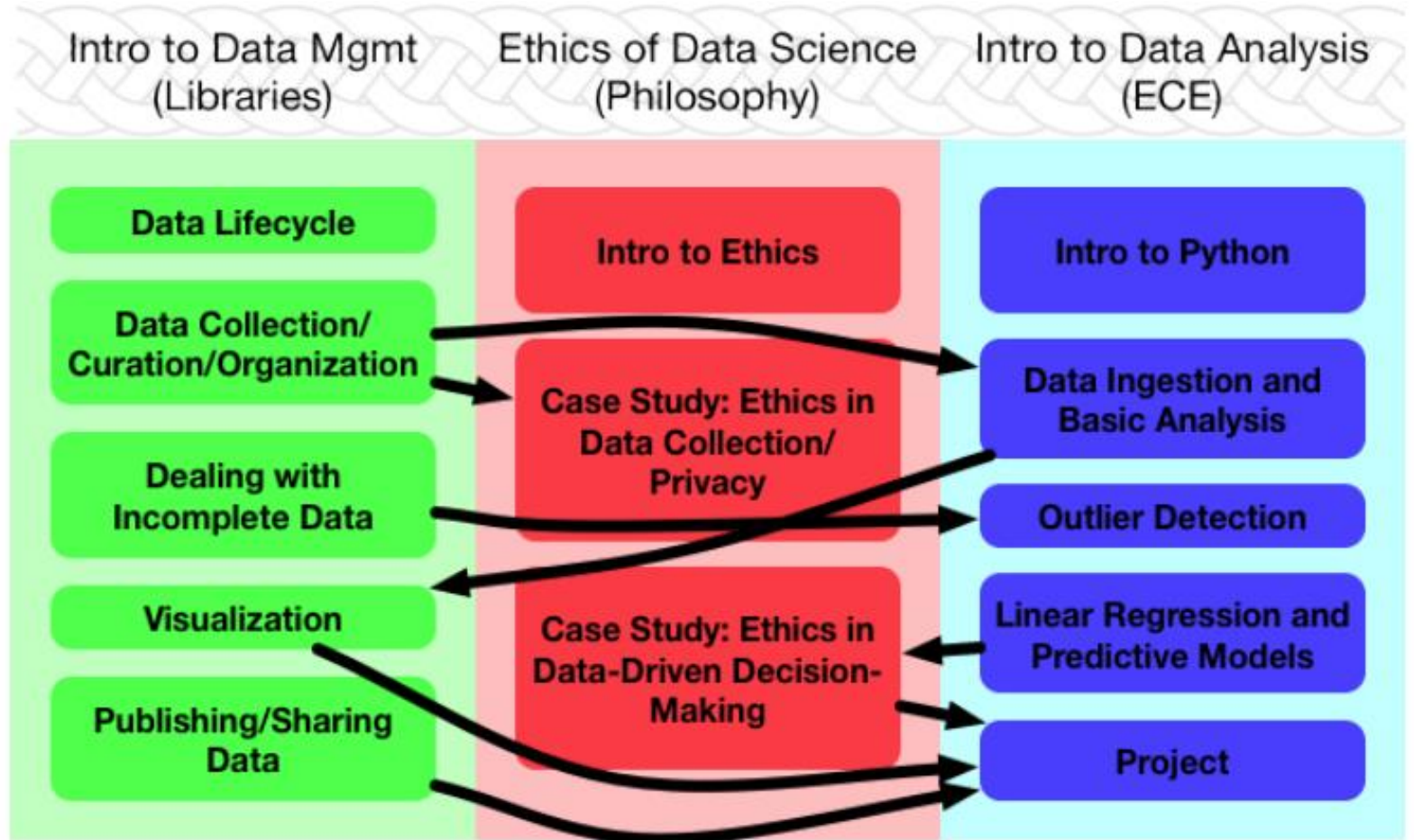*Three one-credit courses* were created to provide an introduction in three key areas of data literacy:

(i) data organization and management

(ii) ethics of data collection and use, and

(iii) methods of data analysis

The courses are aimed at a sophomore level student from any major (i.e., no prerequisites).

VISUAL
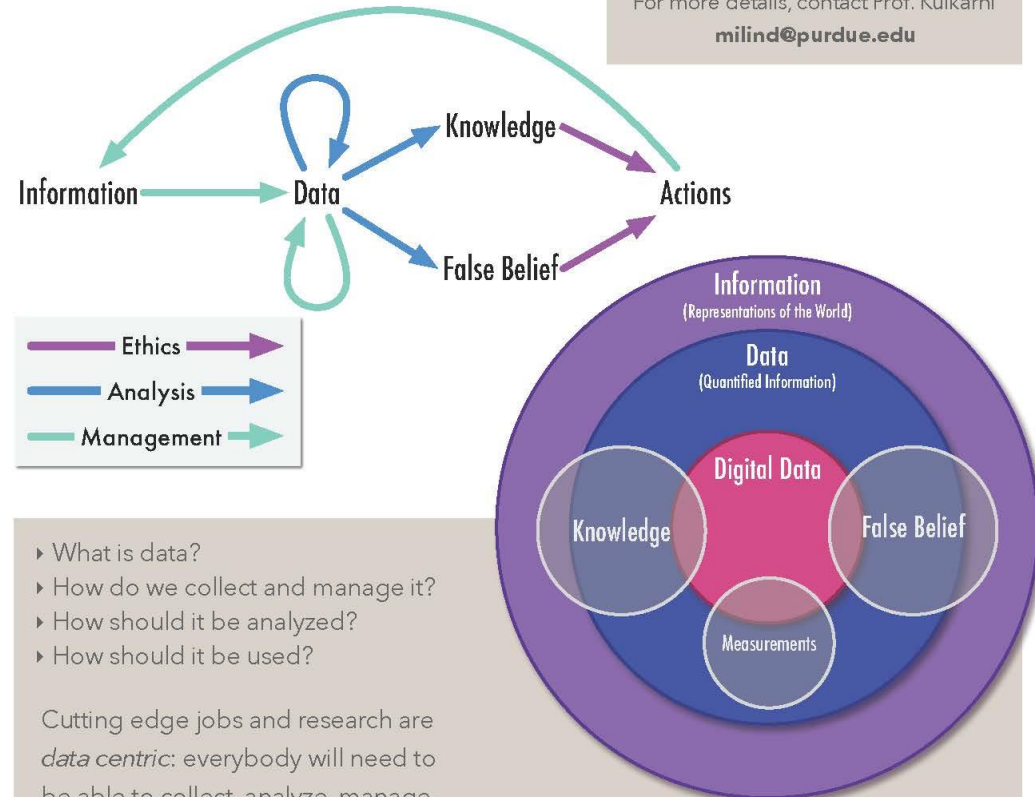
- Courses Flyer
- Courses Structure

VISUAL

- Course flyer we developed as a group

Foundations of Data Science
ECE 295-PHIL 293-ILS 295

For more details, contact Prof. Kulkarni
milind@purdue.edu

Information → Data → Knowledge → Actions
Data → False Belief → Actions

Ethics
Analysis
Management

- What is data?
- How do we collect and manage it?
- How should it be analyzed?
- How should it be used?

Cutting edge jobs and research are *data centric*: everybody will need to be able to collect, analyze, manage, and use data. But what does that mean? How do you know what questions to ask? How do you present that data to others? What are the right and wrong ways to navigate the world of data?

Information (Representations of the World)
Data (Quantified Information)
Digital Data
Knowledge
False Belief
Measurements

Spring 2019
- **Mon** ILS 295, Intro to Data Management
- **Wed** PHIL 293, Ethics of Data Science
- **Fri** ECE 295, Intro to Data Science

3:30-4:20
Take all three for the full picture!

INTERLOCKING APPROACH

- **ECE 295: Introduction to Data Analytics**

Milind Kulkarni
Electrical and Computer Engineering

Provides a broad introduction to data analysis and modelling for Engineering majors. Using Python, students propose research questions, work with data sets to help answer those questions, and visualize data to communicate the results of the analysis more clearly.

INTERLOCKING APPROACH

▪ **PHIL 293: Ethics of Data Science**

Matthew Kroll, Taylor Davis, Dan Kelly
Philosophy

A framework for identifying ethical issues and thinking clearly about ethics in practical contexts, as well as practice applying this framework in the analysis of case studies. Areas of focus include questions of autonomy and privacy arising from the collection and storage of sensitive data, and questions of justice and equity arising from the use of data in policy and decision-making.

INTERLOCKING APPROACH

- **ILS 295: Intro to Data Management**

Sarah Huber, Wei Zakharov, Michael Fosmire
Libraries and School of Information Studies

Provides a foundation in the concepts of data organization, management, preservation, and publication. Students will develop an ability to locate, access, transform, and evaluate data to answer research questions. They will communicate the results of their data searches, and format the data for sharing.

COURSE WEEKLY SCHEDULE

Week of January 7: Data life cycle, syllabus

January 14: International perspectives

January 21: Different types and uses of data

January 28: Primary and secondary data collection

February 4: Data exploration, Part I

February 11: Archival data

February 18: Data exploration, Part II

February 25: Big data

March 4: Data quality management

March 11: No classes/spring break

March 18: Data quality management

March 25: Ethics of data use (privacy, policy implications)

April 1: Data visualization, Part I

April 8: GIS data visualization

April 15: Data visualization, Part II

April 22: Data sharing & final presentations

April 29: Handling sensitive data & final presentations

GUEST LECTURES

- International perspectives- Olof Olsson, System Developer at Swedish National Data Service (SND)

- Archival data – Carly Dearborn, Digital Preservation and Electronic Records Archivist with Purdue Libraries

- Big data and data management- Yung-Hsiang Lu, Professor from Electrical and Computer Engineering at Purdue University

- GIS data and visualization- Nicole Kong, Associate Professor and GIS Information Specialist with Purdue Libraries

COURSE GRADES

- Participation/Weekly quizzes
- Bi-weekly assignments (7 assignments)
- Final Project, Stage 1 (1,2)
- Final Project, Stage 2 (3, 4,5,6)

FINAL PROJECT

Final project will be on the topic of every day data of your choosing (topics will have to be approved by the instructors) and presented to the class through a Wiki page.

Key deliverables include:

1. An overview of your project
2. An explanation of where you found your data and why you chose it
3. An evaluation of the quality of your data
4. A data management plan analysis (this is not a detailed data plan, just an analysis of what you would do)
5. Data visualization(s)
6. A summary of what you learned from the final project and what would you do differently if you were to do it again.

PRE-CLASS SURVEY

- "Data" is a big buzzword these days, but it's almost never defined. What is your definition of "data"?

- With the increasing amount of data being collected, and all the new ways that data is being used, new ethical implications are arising. What ethical concerns are you aware of when working with data?

- When gathering, collecting, or analyzing data, there are laws and regulations that you have to follow. What is the relationship between following laws and regulations and being ethical?

PRE-CLASS SURVEY

- What is the purpose of analyzing data?

- When you need to work with data, do you know where to find it? If so, where?

- How do you store and manage data that you collect?

- What are you hoping to get from this class?

POST-CLASS SURVEY

- We only changed the last question to:

  - How likely are you to recommend this course to other students?
    (Likert scale from Definitely to Definitely Not)

WHAT'S COMING NEXT

- New Undergraduate Certificate of Applications in Data Science, Starting in Fall 2019

- Integration into undergraduate research programs

REQUIREMENTS: APPLICATIONS OF DATA SCIENCE CERTIFICATE

- A minimum of 16 credit hours

 - Core Courses: a minimum of 10 credit hours
  - 3 credit hours in Statistical Methods
  - 3 credit hours in Computing
  - 3 credit hours in Data Literacy, Management & Analytics
  - A minimum of 1 credit hour in Data Ethics & Digital Citizenship

- Application Focus: Six (6) credit hours in courses concentrated on applications of data science related knowledge

- A minimum of six (6) credit hours must be in coursework outside the student's program.

COLLABORATION OR QUESTIONS?

Wei Zakharov

wzakharov@purdue.edu

Sarah Huber

huber47@purdue.edu

Michael Fosmire

Fosmire@purdue.edu