Couture Curricula - BD2K Data Science Tailored to Your Needs

William Hersh[1], David A. Dorr[1,2], Melissa Haendel[1,3], Nicole Vasilevsky[1,3], Bjorn Pederson[1], Jackie Wirz[1,3], Shannon McWeeney[1]

[1]Department of Medical Informatics and Clinical Epidemiology, [2]Department of Medicine, [3]Ontology Development Group, Library
Oregon Health & Science University, Portland, OR
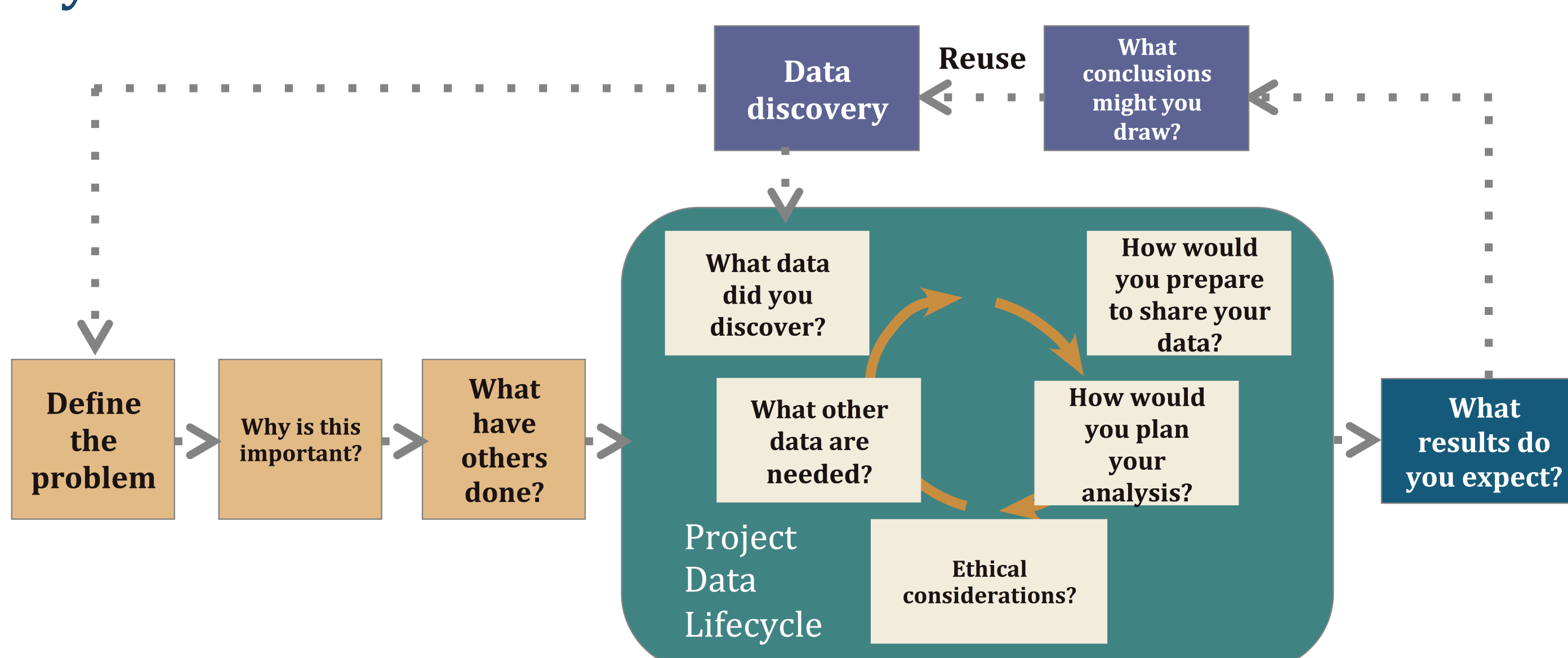
Why?

- Major challenge: how to manage, analyze and interpret vast amounts of data being generated in biomedical research
- One goal of NIH Big Data to Knowledge (BD2K) initiative: provide training for students and researchers to address this
- Research team in the Department of Medical Informatics and Clinical Epidemiology (DMICE) is developing skills courses

Approach

1. Develop Open Educational Resources (OERS)
2. Teach Skills Courses

OERs and courses connect the dots that help researchers understand how to apply data science techniques in the context of their whole research life cycle

- Skills course and OER topics are aimed to fill specific gaps
- Teaching students how to ask the question and follow through

Skills Course Topics

Defining The Problem

- Problems amenable to analytics
- Importance of question
- Team definitions
- Scope
- When we do this wrong: methods don't match

Wrangling Data

- Exploratory Data Analysis
- Data Dictionary
- As you touch data, what can go wrong?

Data Identification And Resources

- Finding the right data
- Search methods
- Use of metadata
- Data management

Methods, Tools And Analysis

- Visualization
- Matching algorithms to problems

Scientific Communication

- Reporting Findings and Limitations
- Giving "Elevator Speech" on ideas of how to approach problem
- Critique of related problem

Course Offerings

Intro Course	• Week long course in Summer 2015 • Offered to interns and undergraduates • Taught basics of data science in the context of the research life cycle
Data After Dark	• Two-evening course in January 2016 • Offered to OHSU students, staff and faculty • Taught basics of data science
Advanced Course	• Four-evening course to be offered in May 2016 • Offered to OHSU students, staff and faculty • Teach advanced topics in of data science

OER Modules

1 | Biomedical Big Data Science
2 | Introduction to Big Data in Biology and Medicine
3 | Ethical Issues in Use of Big Data
4 | Terminology of Biomedical, Clinical, and Translational Research
5 | Computing Concepts for Big Data
6 | Clinical Data and Standards Related to Big Data
7 | Basic Research Data Standards
8 | Public Health and Big Data
9 | Team Science
10 | Secondary Use (Reuse) of Clinical Data
11 | Publication and Peer Review
12 | Information Retrieval
13 | Version control and identifiers
14 | Data annotation and curation
15 | Data and tools landscape
16 | Ontologies 101
17 | Data modeling
18 | Data metadata and provenance
19 | Semantic data interoperability
20 | Semantic Web data
21 | Context-based selection of data
22 | Translating the Question
23 | Implications of Provenance and Pre-processing
24 | Data tells a story
25 | Choice of Algorithms and Algorithm Dynamics
26 | Statistical Significance, P-hacking and Multiple-testing
27 | Displaying Confidence and Uncertainty
28 | Visualization and Interpretation
29 | Replication, Validation and the spectrum of Reproducibility
30 | Regulatory Issues in Big Data for Genomics and Health
31 | Hosting data dissemination and data stewardship workshops
32 | Guidelines for reporting, publications, and data sharing

Available at:
http://skynet.ohsu.edu/bd2k

Challenges

Scope
How to scope generic curricula for different levels of users

Style
How to translate diverse teaching styles into general materials

Images
How to incorporate images and other copyrighted materials into open

Dissemination
How to maximize dissemination while protecting intellectual property

Acknowledgements