Designing undergraduate biology curricula to teach skills and knowledge for data-intensive environmental research

Sam Donovan[1], Anna K. Monfils[2], Alison Hale[3], & Hayley Orndorf[1]

(1) Department of Biological Sciences and QUBES, University of Pittsburgh (2) Department of Biology and Institute for Great Lakes Research, Central Michigan University (3) Carnegie Museum of Natural History

What are the skills and knowledge necessary for data-intensive research?

This poster addresses the challenges faced by undergraduate biology educators in preparing students for participation in data-intensive environmental research. The information is organized into three broad areas: identifying the skills and knowledge required; addressing where and how these skills might be taught; and descriptions of existing resources that could be useful in solving these problems.

Data management and processing

Data has a life cycle and their utility is tied to the quality of their metadata and efforts to make them accessible.

Software skills for science

Using scripting and software workflows facilitates working with large datasets.

Analysis

Analyses may take many forms, but they should all be documented in a way that makes them reproducible.

Visualization

Looking at data in multiple ways is an important part of understanding them and generating questions.

Communication

Interdisciplinary communication plays an important role within a research team, and the impact of research depends on effective external communication.

Figure 1. A high level taxonomy of skills for data-intensive research and example "habits of mind" that might be the target of undergraduate education. Modified from Hampton et al., (2017).

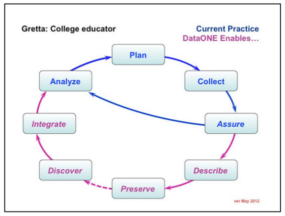


Figure 2. The data life cycle annotated to show a potential use case scenario for a college educator persona. Adapted from DataONE [www.dataone.org/user-personas].

The vast majority of the education and training efforts for participation in data-intensive science have focused at the graduate and professional levels. For example, recent work by Hampton et al. (2016) has identified a set of five capstone skills that provide a "roadmap" for the training required to fully participate in environmental science using modern data and analysis resources (see Figure 1). Professional research communities like DataONE have also played an important role in training scientists about the nature of the data life cycle (see Figure 2)The library and information science community has also attempted to map the data literacy landscape in anticipation of providing resources for teachers and learners (Carlson, et al., 2015, and see Figure 3). Policy guidance for science education curricula have come primarily from the Next Generation Science Standards (NRC, 2013) and the Biology Advanced Placement curriculum guidelines (College Board, 2011). The most widely cited undergraduate biology curriculum policy document, Vision and Change (AAAS, 2011) has only very broad descriptions of how students should engage with data. Additional information to identify specific learning outcomes and the likely developmental relationships among these ideas can be drawn from the computer science, statistics, and mathematics education communities (Finzer, 2013).

- Data Processing and Analysis
- Data Management and Organization
- Data Preservation
- Databases and Data Formats
- Ethics and Attribution
- Data Quality and Documentation
- Data Curation and Reuse
- Data Conversion and Interoperability
- Data Visualization and Representation
- Discovery and Acquisition
- Metadata and Data Description
- Cultures of Practice

Figure 3. The 12 compents of data information literacy identified by Carlson, et al., (2011).

Acknowledgements

This work was funded by National Science Foundation grants DBI 1539900, DUE 1446269, DUE 1446258, and DUE 1446284. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

How might these skills be incorporated within an uncergraduate curriculum?

The challenges involved in determining how the skills and knowledge associated with data intensive research might be supported in undergraduate curricula are myriad. We organize our analysis around three broad issues: mapping foundational skills and learning pathways to desired learning outcomes; articulating pedagogical commitments that support the development of data skills, and exploring instructional strategies that engage students in important features of data intensive scientific practices.

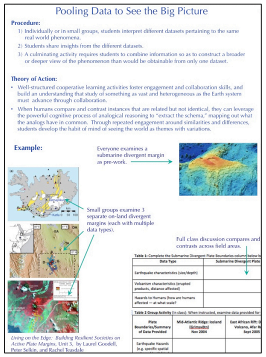
A. NGSS Science and Engineering Practices

1. Asking questions (for science) and defining problems (for engineering)
2. Developing and using models
3. Planning and carrying out investigations
4. Analyzing and interpreting data
5. Using mathematics and computational thinking
6. Constructing explanations (for science) and designing solutions (for engineering)
7. Engaging in argument from evidence
8. Obtaining, evaluating, and communicating information

B. Occupational Profile Tasks

1. Defines the Problem
    - Determines stakeholders' needs
    - Articulates the question
    - Translates question into a research plan
    - Designs the experiment
    - Develops deep domain knowledge of data source
2. Wrangles Data
    - Collects data
3. Manages Data Resources
4. Develops Methods and Tools
5. Analyses Data
    - Develops analysis plan
    - Applies methods and tools
    - Evaluates results of the analysis
    - Answers the question
6. Communicates Findings
    - Compiles report
    - Describes the problem, method, and analysis
    - Contrasts alternative approaches and past result
    - Tells "data story" to convey insight
7. Engages in Professional Development

Figure 4. Comparison of the Science and Engineering Practices (SEP) used to organize Performance Expectations (PE) articulated in the Next Generation Science Standards (Column A) and the mapping of the PEs to various task categories determined for a big data occupational profile (Column B). Of the 71 PEs at the high school level 44% (31/71) could reasonably construed to involve data and 56% (40/71) did not involve working with data. Modified from Kastins (2015).



Pooling Data to See the Big Picture

Procedure:

1) Individually or in small groups, students interpret different datasets pertaining to the same real world phenomena.
2) Students share insights from the different datasets.
3) A culminating activity requires students to combine information so as to construct a broader or deeper view of the phenomenon than would be obtainable from only one dataset.

Theory of Action:

- Well-structured cooperative learning activities foster engagement and collaboration skills, and build an understanding that study of something as vast and heterogeneous as the Earth system must advance through collaboration.
- When humans compare and contrast instances that are related but not identical, they can leverage the powerful cognitive process of analogical reasoning to "extract the schema," mapping out what the analogs have in common. Through repeated engagement around similarities and differences, students develop the habit of mind of seeing the world as themes with variations.

Example:

Everyone examines a submarine divergent margin as pre-work.

Small groups examine 3 separate on-land divergent margins (each with multiple data types).

Full class discussion compares and contrasts across field areas.

Table 1: Complete the Submarine Divergent Plate Boundaries column below b

Data Type	Submarine Divergent Plate
Earthquake characteristics (size/depth)	
Volcanism characteristics (erupted products, distance affected)	
Hazards to Humans (how are humans affected — at what scale?	

Table 2 Group Activity (In class): When instructed, examine data provided for

Plate Boundaries/Summary of Data Provided	Mid-Atlantic Ridge: Iceland (Grimsvötn) Nov 2004	East African Rift: D Volcano, Afar R Sept 2005
Earthquake Hazards (e.g. specific spatial		

Living on the Edge: Building Resilient Societies on Active Plate Margins, Unit 3, by Laurel Goodell, Peter Selkin, and Rachel Teasdale

Figure 5. Description of the design pattern for data-rich curricula, "Pooling data to see the big picture" from Karstens, et al. (2015).

Principles of data informed learning

New ways of using data must build on students' prior experiences.

Learning to use data should occur at the same time as learning about a disciplinary subject.

Learning should result in students becoming aware of new ways of using data as well as developing new understandings of the subject being studied.

Figure 6. Data informed learning emphasizes the development of data literacy in the context of using data in disciplinary contexts (modified from Maybee and Zillinski, 2015).

One of the primarly pedagogical commitments when teaching data skills involves contextualizing those skills within disciplinary problems (Figure 6). This requires a commitment to engaging students with realistic data tasks. Kastens has described a set of design patterns for data-rich curricula including Data puzzles; Pooling data to see the big picture; Make a decision or recommendation; Predict-observe-explain; Nested data sets; and, Deriving a new data type (Figure 5).

Were are the existing models, resources, and projects addressing data skills?

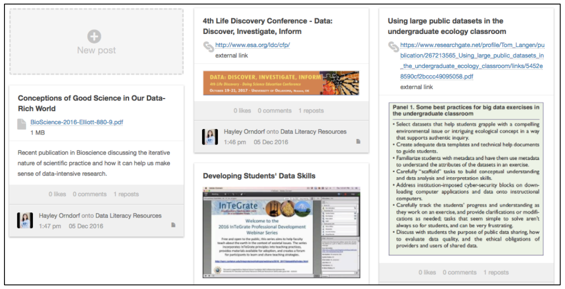
There are a variety of complementary efforts underway to articluate, design, and implement curricula and activities that will support the development of the skills and knowledge required to participate in data-intensive science. You can find a collection of these resources on QUBESHub (Figure 7) where it is possible to share your own resources and comment on posted resources.

Ways To Get Involved

Add your email to the signup sheet attached to this poster or email sdonovan@pitt.edu to receive updates on these projects.

Visit QUBESHub.org for information about Faculty Mentoring Networks and other resources that can help you engage with teaching data skills.

Track the "*Biodiversity Literacy in Undergraduate Education*" [https://qubeshub.org/groups/blue_data] and "*Data Incubator Group*" [https://qubeshub.org/groups/data_incubator] Research Coordination Network projects.


New post
0 likes 0 comments 1 reposts
external link
DATA: DISCOVER, INVESTIGATE, INFORM
InTeGrate

Conceptions of Good Science in Our Data-Rich World

BioScience-2016-Elliott-880-9.pdf
1 MB

Recent publication in Bioscience discussing the iterative nature of scientific practice and how it can help us make sense of data-intensive research.

Hayley Orndorf onto Data Literacy Resources
1:47 pm 05 Dec 2016

4th Life Discovery Conference - Data: Discover, Investigate, Inform

http://www.esa.org/ldc/cfp/

Hayley Orndorf onto Data Literacy Resources
1:46 pm 05 Dec 2016

Developing Students' Data Skills

Welcome to the 2016 InTeGrate Professional Development Webinar Series

Using large public datasets in the undergraduate ecology classroom

https://www.researchgate.net/profile/Tom_Langen/publication/267213565_Using_large_public_datasets_in_the_undergraduate_ecology_classroom/links/5452e8590cf2bccc49095058.pdf

Panel 1. Some best practices for big data exercises in the undergraduate classroom

- Select datasets that help students grapple with a compelling environmental issue or intriguing ecological concept in a way that supports authentic inquiry.
- Create adequate data templates and technical help documents to guide students.
- Familiarize students with metadata and have them use metadata to understand the attributes of the datasets in an exercise.
- Carefully "scaffold" tasks to build conceptual understanding and data analysis and interpretation skills.
- Address institution-imposed cyber-security blocks on downloading computer applications and data onto instructional computers.
- Carefully track the students' progress and understanding as they work on an exercise, and provide clarifications or modifications as needed; tasks that seem simple to solve aren't always so for students, and can be very frustrating.
- Discuss with students the purpose of public data sharing, how to evaluate data quality, and the ethical obligations of providers and users of shared data.

Figure 7. Screen shot from a collection of data literacy teaching and learning resources. View the complete set at [https://qubeshub.org/groups/blue_data/collections/data-literacy-resources].

References

American Association for the Advancement of Science. (2011). Vision and change in undergraduate biology education: a call to action: a summary of recommendations made at a national conference organized by the American Association for the Advancement of Science, July 15–17, 2009. Washington, DC.

Carlson, J., Fosmire, M., Miller, C. C., & Nelson, M. S. (2011). Determining data information literacy needs: A study of students and research faculty. portal: Libraries and the Academy, 11(2), 629-657.

College Board. (2011). AP Biology Curriculum Framework 2012–2013.

Finzer, W. (2013). The data science education dilemma. Technology Innovations in Statistics Education, 7(2).

Hampton, S. E., Jones, M. B., Wasser, L. A., Schildhauer, M. P., Supp, S. R., Brun, J., ... & Fernández, D. S. (2017). Skills and Knowledge for Data-Intensive Environmental Research. BioScience, 67(6), 546-557.

Kastens, Kim. (2015, May). Data Use in the Next Generation Science Standards (revised edition) [White paper]. Waltham, MA: Oceans of Data Institute, Education Development Center, Inc. Retrieved from http://oceansofdata.edc.org/our-work/data-next-generation-science-standards

Kastens, K., Krumhansl, R., & Baker, I. (2015). THINKING BIG. The Science Teacher, 82(5), 25.

Maybee, C., & Zilinski, L. (2015). Data informed learning: A next phase data literacy framework for higher education. Proceedings of the Association for Information Science and Technology, 52(1), 1-4.

National Research Council. (2013). Next generation science standards: For states, by states.