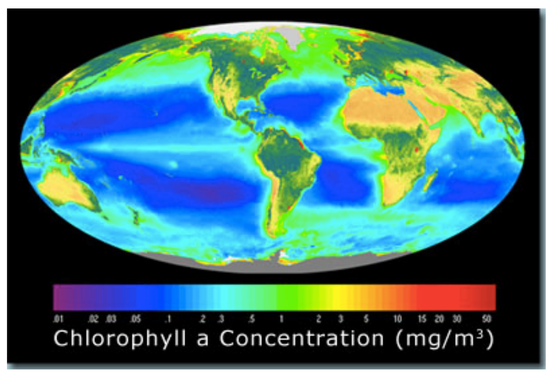
Adoption and Adaptation of Data Science in Oceanography

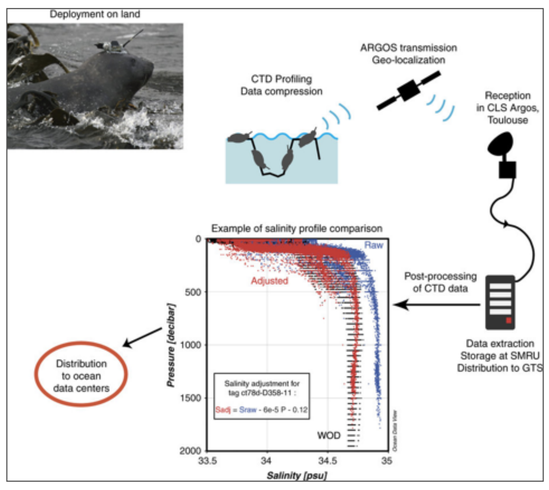
Katerena Kuksenok
University of Washington
kuksenok@cs.uw.edu

ADOPTION: LEARNING NEW SKILLS

What, if anything, is qualitatively new about the suite of tools and skills under the broad "data science" designation as it is currently popularized and understood?



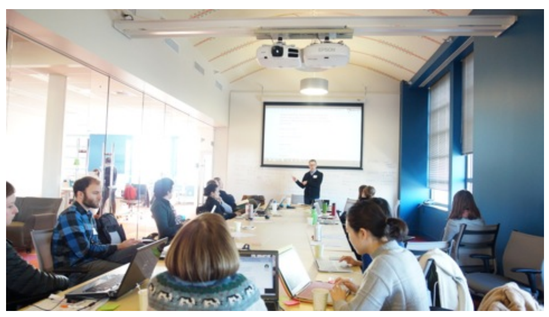
oceancolor.gsfc.nasa.gov/SeaWiFS/



www.nature.com/articles/sdata201428

*"**Traditional methods** for determining spatial distributions of planktonic taxa involve net, pump, and bottle collections followed by the **tedious and time-consuming task** of plankton sample analysis. Thus, plankton ecologists often **require months or even years to process samples from a single study**. In this paper, we present a method that allows **rapid visualization** of the distribution of planktonic taxa while at sea. … We describe the techniques used in imaging the plankton, analyzing the video, and visualizing the data. We present an example of at-sea data analysis conducted aboard [research vessel at specific place on a specific month] and visualizations of the 3-dimensional distribution of selected planktonic taxa in a 2 2 km 90 m volume of seawater. A video of the image processing and visualization is included on the CD-ROM accompanying this volume and is **an essential part of this paper.**"*
-'96 oceanography abstract

Although the novelty of tools or practices can be debated, data science workshops are increasingly popular. What is their role and impact?



It's been on my to-do list, to learn Python, for a long time, but I was having trouble sitting down with the raw manuals (laughs) so SWC was like a way of getting into it, and I had heard about git but I haven't used it.
- natsci post-doc & workshop participant

Working Insight
Workshop model is popular for allowing trying something with minimal investment against backdrop of anxiety. Self-efficacy is the goal; success is the capability to look for technical solutions.

Background & Motivation

Changing practices of software creation and use in the natural sciences transform scientific practice; intra-disciplinary collaboration; inter-disciplinary collaboration

Oceanography is unstandardized, exacerbates challenges in collaboration over software and data & building shared software and hardware infrastructure

Research Questions and Goals

What are social and emotional dynamics of learning new skills for creating and using scientific software for data collection, management, analysis, presentation, and sharing?

What are the challenges of integrating new skills into existing computation practice?

What are some concepts useful for data science advocates and practitioners that want to enable scientific progress?

From left: Maddie, Emmerline, Katerena

Methods & Research Sites

Ethnographic studies using interviews & observations
Focus on breakdowns, blockers & communication
Focus on how small groups:

- Negotiate data collection needs
- Negotiate software priorities
- Ideate future studies or software
- Understand and navigate technical roadblocks
- Recover from disagreements, tensions, breakdowns

(A) Depth: Oceanography Research Sites

- Research sites: 3 oceanography groups (N=~30, 1 year)
- Activities: meetings, debugging / coding sessions, cruises

(B) Breadth: Intensive Learning Environments

- Astronomy workshop on software integration (N=10, 1 week)
- Data science workshops (eg, Software Carpentry Bootcamp)

Progress

5/12mo fieldwork; ~100 hours observation, some interviews
2/4 workshops observations, 15/30 workshop interviews
4/30 oceanographer interviews conducted
3/15 debugging sessions shadows

Acknowledgements

Advisors: Cecilia Aragon (HCDE), James Fogarty (CSE)
Additionally: Gina Neff, Bill Howe, Brittany Fiore-Silfvast
Master's students: Emmerline Wu, Mei-Jen Chen
Undergraduate students: Ron Piell, Maddie Wood

Funding: NSF GRFP and AT&T Research Fellowship

ADAPTATION: VALUES IN INTERDISCIPLINARY COLLABORATION

Which communities and what bridges?

(1) Different kinds of oceanographers working together with shared code and/or data, or negotiating a mutually-beneficial data collection or analysis

(2) Computer scientists and oceano-graphers collaborating on novel data, software, hardware

(3) Computer scientists teaching domain scientists new tools skills or practices

(4) Different kinds of domain scientists teaching one another

Working Insights
Charts; Visualizations; and Illustrations: core for communicating & negotiating research process
Elegance and beauty as a value that informs task prioritization, differently for different stakeholders

Post-doc is presenting a chart that combines the data from multiple data collection cruises.

Supervisor asks if it has Student's data

Post-doc: no

Student (excitedly): it's ready! just finished!

Post-doc (neut.): tell me where it is, I'll upload it

Student (hesitant): I need to make sure there's no crap in it…

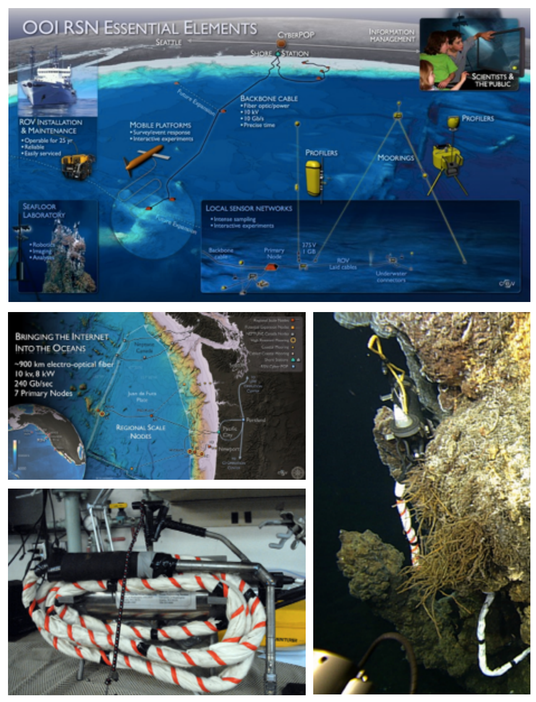
Several group members, including Post-doc and Supervisor talk about what the crap is

Post-doc (agreeing, concluding): well if you know it's crap then I don't want it

-Bio Group Observation Fieldnote

*[On seeing iPython in a demo:] I was really surprised at how simple it seemed to be to get this up and running, as well as to look at **the potential for presenting computations live** in presentations by connecting to iPython Notebook. It really seems like it could be a **very impressive** difference from having a static slide deck to being able to, partly through pop up a browser window and show a Notebook running live on the cluster, to demonstrate some results as you're giving a talk*

–Computational biology PhD student



The Ocean Observatories Initiative is a large hardware and software infrastructure project on underwater volcanoes. See Steinhardt & Jackson CSCW2015 on "anticipation work" carried out by oceanographers in this project.