A design process for developing Ocean Data Explorations

C. Sage Lichtenwalner
Marine and Coastal Sciences
Rutgers University
New Brunswick, NJ, USA
sage@marine.rutgers.edu

Kristin Hunter-Thomson
Dataspire Education & Evaluation LLC
Princeton, NJ, USA
kristin@dataspire.org

Janice McDonnell
Marine and Coastal Sciences
Rutgers University
New Brunswick, NJ, USA
mcdonnel@marine.rutgers.edu

Abstract*—Over the past 5 years, the Ocean Data Labs project has developed a series of online interactive data visualizations that incorporate cutting-edge research datasets available from the Ocean Observatories Initiative (OOI). These activities, which we call "Data Explorations," are designed to support science concepts commonly taught in introductory oceanography courses. Each exploration features an interactive data widget that allows students to explore a curated and canned dataset, but with enough interactivity to support their own process of inquiry. Additional content on each page includes background information and a number of orientation and interpretation questions, which ultimately guides students towards answering a higher-level research question. We have identified a 6-step process that has helped inform our development efforts, especially as we have integrated faculty instructors into the design process. The process includes a) identifying scientific concepts and skills, b) matching educational goals with available instrumentation or research results, c) finding available instrument data, d) cleaning up the dataset, e) building interactive visualizations, and f) adding context and educational surrounds. This paper presents some of the lessons and recommendations we have learned while developing thirty-one activities, which we hope will help others interested in developing similar activities.

***Keywords*—education, data visualization, data analysis, design process, ocean observing systems**

I. Introduction

Engaging students in authentic data analysis experiences can be a helpful way to introduce them to the real-world process of science. All too often, introductory courses in oceanography include limited opportunities for students to analyze or interact with datasets collected in the ocean. In addition, many textbooks present figures that are smoothed or idealized versions of scientific processes, which are often incongruous with actual datasets. And for both science and non-science majors in these courses, this is a lost opportunity to engage them in activities that can develop their data and graph literacy skills, while also reinforcing the scientific concepts being taught. The availability of data from large observing systems, like the Ocean Observatories Initiative (OOI), presents an intriguing opportunity to connect the cutting-edge science and real-time authentic data collected by these systems to introductory classrooms where it can be used to augment existing lessons [1]. The key trick is to overcome the measurement, data system, and natural system complexities inherent in the analysis of data in order to create activities that are advantageous for student learning.

A. The OOI Dataset

The OOI was commissioned in 2015, as a cutting-edge research facility to study multi-scale, high-resolution processes in the ocean over both short and long-term time periods. The OOI includes over 760 deployed instruments, spread across 7 arrays in the western Atlantic and eastern Pacific Oceans, recording over 200 different data streams [2]. Most of the moored instruments collect data at a minimum of every hour, and many of the cabled and profiling instruments sample at the minute scale or less. This dataset, now over 5 years in length, provides a rich new resource that undergraduate educators can use to augment science topics taught in the classroom, or to provide students with novel research experiences [3]. As originally conceived, the OOI was designed to support both research and educational goals. Over the past several years, our efforts have focused on translating the OOI dataset into resources that can be used and adapted by educators.

B. Early Pilot Projects

Shortly after the OOI first came online, in 2016-2017 the *Ocean Data Labs* team at Rutgers University created a collection of prototype "Data Explorations" to explore the potential for using OOI datasets in the classroom [4]. The goal was to figure out how to overcome the numerous barriers to using the OOI resource with students, such as accessing data from the OOI's nascent data portal, finding appropriate datasets for a given science topic, finding available data for a desired instrument and/or location, or making sense of the measured dataset while accounting for quality issues and parsing out complex interacting natural phenomena. In the end, we decided to develop a series of activities, each comprising a "canned" interactive data visualization, using a pre-processed static dataset, focused on a specific science topic.

Eighteen initial prototype explorations were developed, supporting topics like primary production, properties of seawater and tectonic plate boundaries. We hosted three workshops over 2 years to engage faculty from around the country in piloting these initial explorations, while collecting

This material is based upon work supported by the National Science Foundation under Grant No. OCE-1831625.

their feedback on the usability and utility of the prototype explorations [5].

C. The OOI Ocean Data Labs Project

In 2018, we received funding to expand this effort and launched the "OOI Ocean Data Labs" project, with the goal of establishing a "Community of Practice" by engaging faculty from around the country interested in using the OOI dataset with their students [6]. As part of this project, we held a series of four "development" workshops, during which faculty participants helped design and develop over a dozen new Data Explorations. These later explorations include comprehensive instructors' guides written by the faculty, and because of their input, they are directly connected to undergraduate learning outcomes.

Through these efforts, we have learned a lot about the challenges of translating complex datasets from the OOI into useful activities for the classroom. By engaging the faculty community in the development process, we have learned first-hand about their teaching requirements and classroom needs. This has also informed us about how much of the development process faculty are able to handle on their own, and the amount of support needed at each step. From this, we have defined and refined our development process into one we hope is transferable to future collaborative partnerships between educators and those with the technical skills to develop data visualization tools.

II. Data Exploration Design

Each Data Exploration includes a data visualization tool, affectionately called a "widget," and an associated activity or series of questions to guide student analysis and interpretation. Fig.1 shows an example Data Exploration on the tectonics of the Juan de Fuca plate. The full collection of Data Explorations can be found on the OOI Ocean Data Labs project site [7].

Each widget is designed to be 1) simple, intuitive & easy to use, unlike the complex interfaces found in many data portals, 2) focused on a specific learning goal, such that the data visualization is learning outcome-driven, and not driven by the fanciness of the dataset, 3) interactive, to support deeper exploration and hopefully understanding, and 4) embeddable, so that the widget can be placed within an appropriate educational context, e.g., on an online course page with helpful background information and questions.

Each Data Exploration page typically includes the following elements:

- *Your Objective* – This is essentially the "challenge question" posed to students, i.e., the scientific question they need to address in this activity.
- *Data Widget* – An interactive data visualization that allows students to explore a canned dataset in a way that supports the learning goal.
- *Data Tips* – Guidance to students on how to use the widget, and specific things to look for.
- *Questions for Thought* – Specific questions students answer while interacting with the widget that guide them towards addressing the challenge question. These are generally divided into "orientation" questions that ask students to read and understand the graph type shown, and "interpretation" questions that ask students to think more deeply about

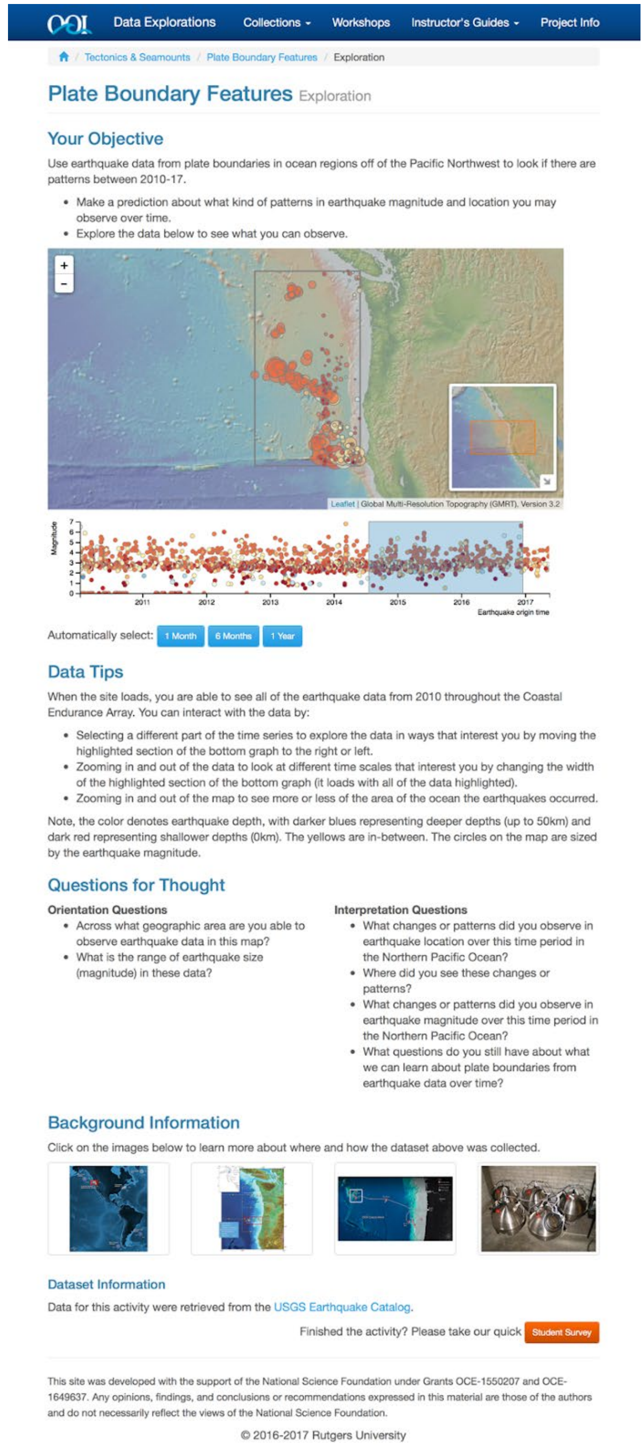

Fig. 1 A screenshot of the web page for the Data Exploration on "Plate Boundary Features."

the dataset and the science topic. Reference [8] includes a description of this questioning strategy.

- *Background Information* – Each page also includes a collection of images to show students where the data were collected from, the instruments used, as well as other relevant figures to support student understanding. Information on the dataset's provenance is also included for those interested in accessing the raw dataset for further exploration.

Ultimately, the Data Explorations are designed to be used within a 15-30 minute period of time in a lecture or lab class, without requiring students to go through additional training that is often required when working with a comprehensive and complex data portal. That is, the Data Explorations are designed to be opened quickly, allowing students to immediately begin exploring the dataset. This is not to say that using data portals and manipulating raw data using data analysis tools (like Excel and Python) are not important for students pursuing science careers. Our goal is to lower the barrier so that all students can engage in data analysis using authentic data before having developed analysis skills with specific tools, perhaps providing more motivation to do so later.

III. Design Process

Fig. 2 illustrates the iterative design process we use to develop a Data Exploration. We have found this process helpful whether we are working as a small internal team or with a group of faculty instructors.

While this process seems linear, and the steps are all important, given the challenges of finding viable datasets, the complexities of working with observatory datasets, and matching what is available to educational goals, we have found that the workflow is often nonlinear. Ensuring the final activity meets the educational design requirements, including the desired learning outcomes, ultimately requires a lot of iteration and refinement.

Here we describe the key steps in our design process, along with some of the challenges and recommendations for each that we hope will be useful for teams developing their own data visualization-based activities.

A. *Identify Target Scientific Concepts & Skills*

The first step in designing a Data Exploration, is to identify the scientific concepts and skills you wish students to learn and develop as they go through the activity. Our team has found it useful to use the "Backwards Design" approach when designing curricula and activities. By defining the end goal up front, it serves as a signpost from which other decisions can be based. Decisions on the number and types of instruments to include, how a dataset should be structured, how students will interact with the visualization tool, and what additional educational materials are needed to support the activity are all guided by the end goal. Thus, it is important to engage faculty up front in the design process, pulling from their knowledge and experience on the most relevant science concepts and data analysis skills to focus on. They are the best advocates for determining what will work for their students, and what is needed to support their curricula goals.

B. *Match Educational Goals with Instrumentation*

Once you have identified the concepts you wish to design for, the next major step is to narrow the universe of available data to those that will best support your goal. The easiest approach is to review available literature to find previously identified and vetted datasets that can be repurposed for educational activities. Alternatively, if your goal is to use data from a specific observing system, you may wish to have faculty learn about and explore the available instrumentation to find a good fit. A hands-on tutorial of a data portal or other informational websites could also be helpful.

In our OOI Data Labs development workshops, we spent time introducing faculty participants to the OOI, including a brief introduction to its arrays, science themes, and instrument types. Then we asked faculty to identify potential instruments and locations that could work for their activity, making sure their selections matched with the topics and concepts they specified earlier.

C. *Find Available Instrument Data*

The next, and perhaps most challenging step, is to find available data for each desired instrument. If you are running a workshop, and expect instructors to identify viable datasets, we have found that it is essential to have observatory scientists or technicians on hand to answer questions about instrumentation, data availability, data quality and data interpretation. Questions often come up regarding instrument gaps, biofouling, quality control, data formats, and variable/parameter names. In addition, some users may also struggle to use the complex user interface of some data portals. Non-experts may also be unfamiliar with the natural processes depicted by the measured data. Figuring out what is "real" vs. an artifact of sampling, instrument operation or another process, can be a big challenge. Having experts who can help faculty choose specific instruments and time periods to use, can save a lot of time and frustration.

As an example of this discovery process, Fig. 3 shows a Data Exploration widget where the first 3 steps were relatively straight-forward. The faculty team initially decided they wanted their students to explore a dataset that showed a coupled

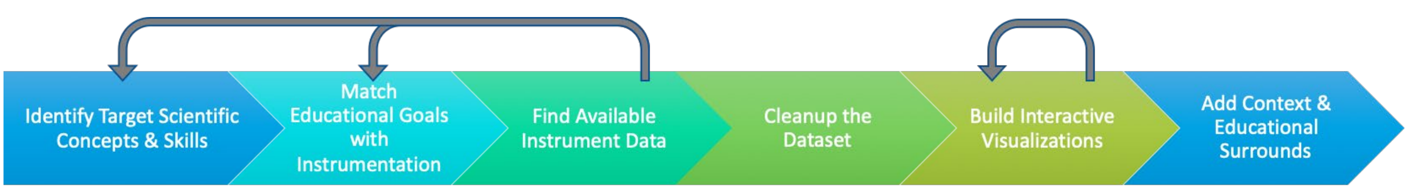

Fig. 2 The Data Exploration iterative design process

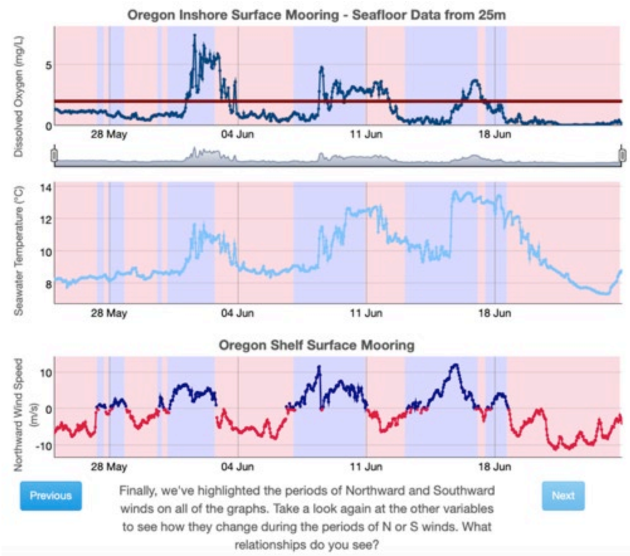

Fig. 3 The data widget for the "Anoxic Events" Data Exploration, showing one month of dissolved oxygen (top), seawater temperature (middle) and northward wind speed (bottom), from May 25 to June 24, 2017. The red and blue highlights denote the periods of southward (upwelling) and northward (downwelling) winds respectively.

biophysical process. They then reviewed an article in the 2018 OOI Special Issue on the Endurance Array, and noticed a figure that showed a time series of dissolved oxygen, temperature and winds off the coast of Oregon that aligned with their goals [9]. After exploring the OOI data portal, they decided to use data from the same instruments but from a different month that captured several upwelling and downwelling events. This new dataset would allow students to explore these processes in more detail.

If you are looking to create activities based on an observatory with a large number of instruments, another approach is to pre-identify selected datasets that faculty can choose from, as we tried with our OOI Data Nuggets project [10]. You could also list potential instruments and available time periods, or pre-generate quick-look graphs for easy reference to help faculty learn about the kinds of datasets that are available.

This step is one of the most iterative in nature. If faculty cannot find a viable instrument with available data, they will need to choose a different instrument, or location, or ultimately a different science concept to pursue.

D. Cleanup the Dataset

Once you have identified the instruments, locations, and time ranges you wish to use, the next step is to clean and process the dataset. Cleaning and wrangling data also can be a highly non-linear process [11].

In the case of the OOI, we typically had to make multiple data download requests for each instrument type and location needed. For example, the widget shown in Fig. 4 uses chlorophyll-a, dissolved oxygen, sea surface temperature and salinity data from 2 different locations, which required downloading 8 different data files. We used Python scripts to load the downloaded datasets, subset the desired parameters and date ranges, remove any outliers, average the datasets when needed (e.g., down-sampling to hourly or daily data), and merge all of the remaining data into one output file. If you are creating JavaScript based interactive tools, you will likely want to generate a comma separated variable (CSV) file. If you expect students to analyze data using another tool (like R, Python or Ocean Data View), other formats may be appropriate, like NetCDF.

We shared the resulting Python notebooks as extension materials for those faculty or students interested in accessing the raw or processed datasets themselves [12]. This supported students who wished to do their own analyses (e.g., with different instruments or time periods), or learn more about using Python to explore the dataset.

E. Build Interactive Visualizations

Once you have your datafile you can develop your desired data visualization or interactive tool. For traditional data activities, this has meant creating a static graph, or having students use a tool like Excel to plot the data themselves. For the Data Labs project, we prototyped developing bespoke JavaScript-based widgets that allow students to interactively explore and visualize the dataset in a structured way. To do this, we relied on several open-source data visualization libraries, including dygraphs (for many of our time series widgets), D3.js (for more complex widgets, including profiles) and Leaflet (for seismic maps).

This step is also very iterative, as designing tools requires a lot of back and forth with faculty to make sure the data are visualized with the proper interactive features to support the intended learning goals, while neither limiting nor overwhelming the student.

For example, Fig. 5 shows the final screen of the "Dynamic Air-Sea Interactions" widget. Before using this widget, students explore time series graphs of a number of variables that were recorded as a "bomb cyclone" passed over the Pioneer Array. In

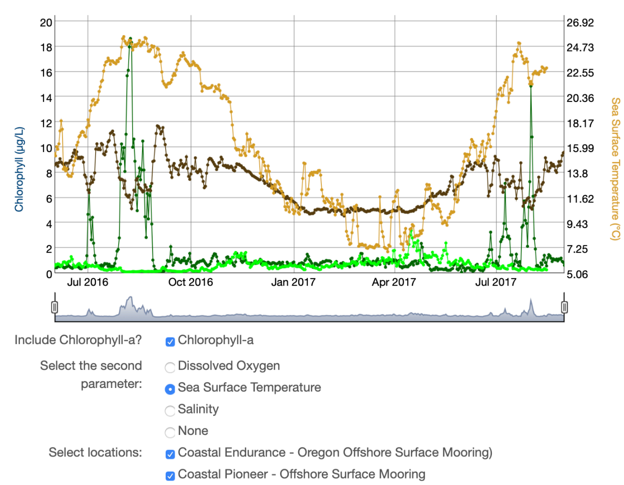

Fig. 4 The data widget for the "Chlorophyll-a in Upwelling and Stratified Temperate Regions" Data Exploration, showing a year of sea surface temperature and chlorophyll-a from offshore the coasts of both Oregon and the Mid-Atlantic.

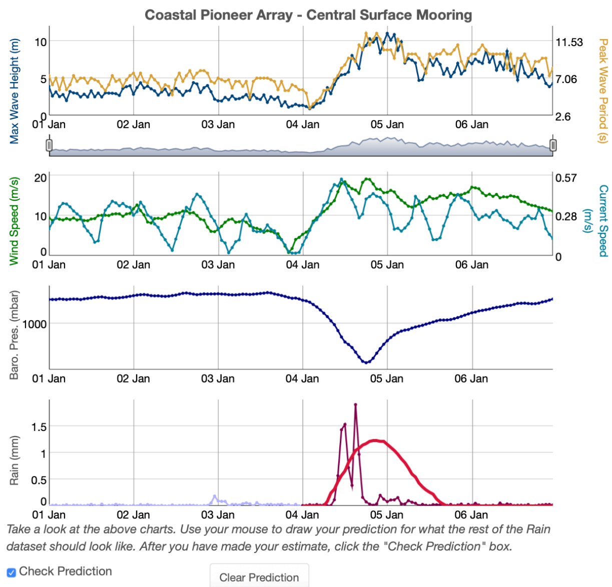

Fig. 5 The data widget for the "Dynamic Air-Sea Interactions" Data Exploration, showing data from a number of instruments as a bomb cyclone passed over the OOI Pioneer Array.

the preceding activities, students are asked to explore relationships, like the correlation between barometric pressure, wind speed and wave height. Then, in this final activity, students are asked to utilize their knowledge of storm processes to predict the rainfall as the storm passes. Using the widget, they can draw a prediction (an example is shown as the red line) before turning on the measured dataset to compare their prediction with the actual data. Other widgets allow students to toggle on/off different variables, adjust the date range of data displayed, or provide other forms of interactive control to support exploration.

F. *Add Context & Educational Surrounds*

The final step in developing an activity is to add context to the lesson as well as educational surrounding materials to support students in their work. If you are following the Data Exploration format, this would include all of the items listed in the *Data Exploration Design* section above. At this point, faculty should revisit their draft challenge question, data tips and questions for thought and refine them once the final widget is developed. Background information is also key to supporting students in their work. The trick is to include enough materials to help them understand the dataset's link to the selected scientific concepts, without providing so much that they won't read it. It is important that students have some agency in exploring the dataset, and thus any and all of this supporting material should not tell them what to interpret from the data. An instructors' guide can also be important for sharing thoughts and resources on how the activity can be integrated into various courses or lectures beyond the course of the faculty members designing the activity.

IV. Discussion

Developing bespoke data visualization activities takes a lot of time and effort, but feedback from faculty and students has been very positive. Simple interactives that "just work" can save precious class and faculty preparation time that would otherwise be spent finding relevant data. When well designed, they can also direct students through questions based on the dataset, leading to higher engagement and better discussions. These advantages have been essential for incorporating these activities into introductory classes, especially by faculty with backgrounds outside of oceanography.

Over the past 5 years, we have followed and refined the 6-step process outlined above as we expanded the collection of Data Explorations. Each new activity and working group presents a new challenge, but the process helps to ensure everyone is on the same page and nothing is missed.

Of all the steps, we have found that step 2 (matching educational goals with available instruments or research results) is often the most challenging, especially with new or complex ocean observing systems like the OOI, that are unfamiliar to many users you might engage. Steps 3 and 5 tend to be the most iterative. As you dive into the data, you often run into data gaps or quality issues that require you to pivot your plan. Thankfully, better data portals can help make this step more efficient. Designing intuitive interactive data visualizations requires a significant amount of effort to meet the goals of the activity, often requiring multiple iterations and rounds of feedback. More than likely, this will require technical expertise to wrangle the data and build the widget. Finally, if you are interested in involving faculty more deeply in the process, steps 1 and 6 are a great place to start, as it is really their expertise that is needed here. As data portals get better, faculty can also help with steps 2 and 3.

It is important to note that this process was designed to support the development of interactive JavaScript-based data tools by small teams, comprised of faculty, curriculum experts, and software developers. We believe this process could also help guide teams or even individuals designing other kinds of data activities, including Python notebook-based activities or even Excel-based ones. In this case, we would recommend working through all of the steps, but the goal of step 5 would be to develop a software notebook or other visualization project concept instead.

The goal of the Data Labs project has been to build a community of faculty interested in tapping into the firehose of OOI ocean data to support undergraduate education. To date we have involved over 250 faculty from around the country in this effort, and many have used one or more of these Data Explorations in their classroom, including over 40 faculty who participated in the development process. We hope that the design process we have refined over the past 5 years can provide a blueprint for future collaborations seeking to bring real ocean science research to undergraduate students.

Acknowledgment

We would like to thank the many faculty who have helped us develop and pilot the Data Explorations over the years. We also thank the OOI Data Team at Rutgers and the OOI scientific staff for their help answering many questions on the OOI dataset throughout the project.

References

[1] J. McDonnell, S. Lichtenwalner, S. Glenn, C. Ferraro, K. Hunter-Thomson, and J. Hewlett, “The Challenges and Opportunities of Using Data in Teaching From the Perspective of Undergraduate Oceanography Professors,” *Mar. Technol. Soc. J.*, vol. 49, no. 4, pp. 76–85, Jul. 2015.

[2] L. Smith *et al.*, “The Ocean Observatories Initiative,” *Oceanography*, vol. 31, no. 1, pp. 16–35, Mar. 2018.

[3] C. Greengrove *et al.*, “Using Authentic Data from NSF’s Ocean Observatories Initiative in Undergraduate Teaching: An Invitation,” *Oceanography*, vol. 33, no. 1, Mar. 2020.

[4] J. McDonnell *et al.*, “Education and Public Engagement in OOI: Lessons Learned from the Field,” *Oceanography*, vol. 31, no. 1, pp. 138–146, Mar. 2018.

[5] K. Hunter-Thomson, S. Lichtenwalner, and J. McDonnell, “Incorporating Observatory Data into Oceanography Courses,” *Eos (Washington. DC).*, Dec. 2017.

[6] J. McDonnell and S. Lichtenwalner, “Sidebar: The Data Labs Project: Building a Community of Practice Using OOI Data” in *Ocean Observatories Initiative (OOI) Science Plan: Exciting Science Opportunities using OOI Data*, Jan. 2021.

[7] OOI Ocean Data Labs project, “OOI Data Explorations”, accessed July 22, 2021. https://datalab.marine.rutgers.edu/data-explorations/

[8] L. Hotaling, J. McDonnell, C. Ferraro, K. Florio, and S. Lichtenwalner, “Educating with Data,” in *Exemplary Practices in Marine Science Education*, G. Fauville, D. L. Payne, M. E. Marrero, A. Lantz-Andersson, and F. Crouch, Eds. Cham: Springer International Publishing, 2019, pp. 207–223.

[9] J. Barth *et al.*, “Warm blobs, low-oxygen events, and an eclipse: The Ocean Observatories Initiative Endurance Array captures them all,” *Oceanography*, vol. 31, no. 1, pp. 90–97, Mar. 2018.

[10] OOI Ocean Data Labs project, “OOI Data Nuggets,” accessed July 22, 2021. https://datalab.marine.rutgers.edu/data-nuggets/

[11] S. Kandel *et al.*, “Research directions in data wrangling: Visualizations and transformations for usable and credible data,” *Inf. Vis.*, vol. 10, no. 4, pp. 271–288, Sep. 2011.

[12] OOI Ocean Data Labs project, “OOI Ocean Data Labs workshop Python notebooks,” accessed July 22, 2021. https://github.com/ooi-data-lab/data-lab-workshops