Engaging students in an ocean data analysis based Virtual REU

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

Josh Kohut
Marine and Coastal Sciences
Rutgers University
New Brunswick, NJ, USA
kohut@marine.rutgers.edu

Christine Bean
Marine and Coastal Sciences
Rutgers University
New Brunswick, NJ, USA
bean@marine.rutgers.edu

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

Abstract*— In the summer of 2020, many Research Experience for Undergraduate (REU) programs around the country were forced to cancel their on-site programs due to the Covid-19 pandemic. In the midst of this crisis, the OOI Ocean Data Labs project at Rutgers University was asked to develop an alternative "virtual REU," relying on our existing network of faculty and virtual community professional development experience. In the end, 15 faculty members were recruited to support 16 undergraduate students, as part of a combined program with Rutgers Research Internship in Ocean Sciences (RIOS). With only a month of planning, the Data Labs-RIOS team put together an 8-week research experience for students, including a 2-week professional development, and a 6-week research project, where students worked virtually with a faculty mentor, in keeping with typical REU summer programs. In this paper, we review the elements of the virtual program enacted in 2020, and how the various components were scaffolded to provide a coherent and positive research experience for students. The program produced a number of positive components worthy of consideration for future virtual REU or other internship programs. Virtual or hybrid programs might be a suitable alternative for non-traditional students or others who cannot participate in traditional programs, but would still benefit from the opportunity to engage in research.

Keywords—research training, career development, undergraduate education, data analysis

I. Introduction

The National Science Foundation (NSF) has long supported hands-on research opportunities for undergraduate students through its Research Experience for Undergraduate (REU) program. The Foundation supports several hundred program "sites" covering all of its research areas, including around 40 in Ocean Sciences alone [1]. Each site typically supports around 8-12 students each year, often as part of an 8-10 week in-person summer research experience including lab and field opportunities. In the spring of 2020, as the scope of the pandemic became clear, states and universities began to shut down in-person activities, enforcing social distancing and stay-at-home orders. As a result, many REU programs were forced to cancel, leaving students who had applied months before with few options.

However, with the support of NSF, several programs pivoted to provide a virtual research experience, and the GEO REU program at NCAR brought together program leaders in a collaborative network to share resources and best practices in mentorship, with a particular emphasis on how to work remotely [2]. As part of this effort, the OOI Data Labs team was asked if it could help recruit faculty mentors from its network to provide additional research opportunities, especially for rising seniors who would not otherwise have an opportunity to participate while still undergraduates.

The Data Labs project started in 2018, with a goal of establishing a "community of practice" of faculty interested using data from the Ocean Observatories Initiative (OOI) with their students. A key part of the project was a series of development workshops, during which faculty worked together to create data-focused activities, called "Data Explorations," that used authentic data from the OOI [3]. The development of these activities required more time than the workshop allowed, in order to design the interactive widgets and assemble appropriate background information and student questions. Thus, the Data Labs team facilitated virtual workgroups following each workshop. In hindsight, this experience with virtual collaboration proved valuable to the Data Labs team and faculty participants, as we all were forced to shift to 100% virtual teaching and collaboration with the onset of the pandemic. Having a ready community of faculty with some virtual collaborative experience was a key reason why the Data Labs project was able to develop a virtual research experience in short order.

The pandemic struck the United States in early March, with many states shutting down by March 13th. By the end of the month, it became clear to many REU program organizers that it would not be possible to run their summer programs as initially planned. As a result, in early April efforts began on reaching out to existing facility networks to identify possible alternatives for the summer. When we reached out to the Data Labs community, we had a large number of faculty express their willingness to help. Over the course of the next month, available mentors were actively recruited, and arrangements were made for supplemental funds to cover alternative program costs.

II. Student and Mentor Selection

In a normal year, a traditional REU summer program will solicit undergraduate applicants during the fall semester, and

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

select and notify them early in the Spring semester. 2020 was anything but normal. Because our REU project only came together in late May, it was not possible to recruit students through a normal process. Instead, our project included 10 students who were initially accepted into other REU sites that had been cancelled, along with 6 students initially accepted in the Rutgers RIOS program, who were rising seniors and willing to continue with a virtual program. This was all possible due to the diligent efforts of everyone involved in matching students with faculty mentors available through the virtual programs.

A traditional REU site will usually recruit faculty mentors from the host institution who have active summer research projects. Indeed, this was still the case this year for the RIOS faculty mentors. However, this physical constraint was not an issue for mentors for the Data Labs component. Instead, we reached out to our existing network to find faculty who were willing to mentor students on such short notice.

Fig. 1 shows a map of the home institutions of the students and mentors who participated in the 2020 Virtual REU, which stretched to the corners of the territory of the U.S.A. Given the nature of the pandemic, this map does not necessarily represent where participants called in from. Indeed, several were in other states at home or with family, and some even changed locations throughout the 8-week program, as personal priorities dictated. However, the virtual nature of the program meant students and faculty could continue to participate, regardless of their summer travels and personal requirements.

III. Ocean Observing Systems Focus

Because face-to-face training was not possible this year, students were not able gain hands-on experience with laboratory or field instruments as they would have with a traditional REU. However, there are a growing number of NSF-sponsored large research programs that collect significant volumes of data, such as the Ocean Observing Initiative (OOI) and the Long-term Ecological Research Program (LTER). These geoscience programs are using advanced technologies, including cabled systems, autonomous gliders, and sophisticated instrument and buoy sensor systems to provide sustained ocean measurements that will aid the study of several science questions. For example, for the OOI this includes climate variability, ocean circulation, ecosystem dynamics, air-sea exchange, seafloor processes, and plate-scale geodynamics over the coming decades [4].

Building data literacy and critical thinking skills utilizing large observatory-based datasets are important for developing the next generation of scientists, but these skills can be challenging for students and faculty to develop [5]. For students studying oceanography, ocean observing systems, like the OOI, provide a wealth of data that can be used to explore scientific processes and investigate actual events [6]. Utilizing observatory data for research requires shifting a mindset from investigating familiar small-scale to perhaps somewhat more unfamiliar larger-scale processes. It also requires going from hands-on experiences to utilizing data collected by others where metadata is needed to gain context about observations, and from simple to complex lines of reasoning. In addition, large observatory datasets also often require utilizing more complex data analysis tools and software code.

Fig 1. A map showing the home institutions of students (blue) and faculty mentors (red) who participated in the 2020 Data Labs-RIOS Virtual REU.

Because there is so much data already collected and ready to explore from the OOI and other observing systems, this became the primary research focus of our virtual REU this year. Students were able to use these datasets to build their programming, data analysis and visualization skills, which are essential and complementary to the traditional lab and field techniques that would be developed in many regular REU programs.

IV. Virtual REU Goals

The "Data Labs Virtual REU," as it came to be known, started with a 2-week professional development workshop (co-facilitated with Rutgers RIOS), where students were introduced to scientific research and the challenge of working with observatory datasets. This was followed by a 6-week intensive individual research experience, during which students worked remotely with the guidance of a faculty mentor on a specific dataset of interest. This format followed the structure of many REU programs, including RIOS.

By the end of the virtual REU, students were expected to:

- Develop and apply their data analysis skills using Python notebooks to access, analyze, and present ocean data.
- Learn about the variety of ocean data collection methodologies and datasets available to oceanographers.
- Participate in a variety of professional development sessions, including scientific question development, science communication, the graduate school process, and Diversity, Inclusion, and Research Ethics.
- Have the opportunity to participate in Career and Graduate Student Panels.
- Develop, carry-out, and summarize a research experience using an online dataset, under the guidance of a faculty mentor. (8-week participants only)

The full schedule, including workshop presentations, example coding notebooks, and other resources, as well as students' final presentations are available on the program website at https://datalab.marine.rutgers.edu/2020-virtual-reu/

V. Two Week Virtual Workshop

The initial 2-week professional development workshop included daily 3-hour Zoom sessions, where students were introduced to the basic components of research and processing data using Python. They also had a chance to meet and interact with graduate students and a career panel. Students then worked in groups on a mini data analysis project as a culminating activity. Workshop sessions were held between 3:00pm and 6:30pm Eastern Time, with a 30-minute break. Ideally, it would have been nice to have a longer (mental) break between sessions, but because participants were calling in from Puerto Rico, Guam and across the continental US, this was really the only viable option. An outline of the workshop sessions is presented in Table 1. The workshop included 3 topical tracks, and a kickoff session, which we describe in the following sections.

A. Workshop Kickoff

The first day of the workshop was dedicated to introductions, community building, logistics, and setting expectations. Following a welcome, students were presented with a series of activities and discussions. The goal was for participants to get to know their cohort, the organizers and maybe a little more about themselves. Activities included:

- *Icebreaker*: We used a simpler version of a "True Colors" activity to allow participants to introduce themselves, sharing 2 personality traits they identify with, and 1 they do not.
- *Mentoring*: A discussion on mentoring relationships allowed students and mentors to reflect on the aspects that make mentoring successful, and strategies for both to gain from the experience.
- *Collaboration*: A discussion on virtual collaboration was essential in setting the stage for the workshop and subsequent research project experience. As this workshop was relatively early on in the pandemic, most participants had yet to really find their footing in the virtual world.
- *Innovation*: Finally, each participant received the results of a FourSight survey [7], which allowed them to reflect on the aspects of innovative and creative thinking they most identify with.

B. Research & Career Training

A key goal of any REU program, apart from the actual research experience, is to provide undergraduates with additional opportunities to learn about careers in research. We relied on the experience and expertise of the Rutgers RIOS team to organize facilitate these sessions, which included:

- *How to ask a testable question*: Students were introduced to the Question Formulation Technique (QFT) from the Right Question Institute [8] as a way to brainstorm and select appropriate testable research questions.
- *The Wonderful World of Data*: Thanks to observing systems like OOI and LTER, students have access to a wide variety of data just sitting on a shelf waiting for them to explore. This session introduced some of the instruments, datasets and research projects possible. Students were also introduced to the ERDDAP interface, which is used to provide data access to a number of popular oceanographic datasets.
- *Science Communication*: This session introduced students to the many aspects and challenges of effective science communication, a necessary skill for researchers who need to consider ways to broaden the impact of their work.
- *Virtual Career Panel*: One of the advantages of a virtual program was the ability to engage panelists from across the country, which resulted in a more diverse panel than might have been possible locally and in-person. We also innovated on the panel format, and asked panelists to record video introductions, which students were able to watch ahead of time to select those they were most interested in meeting. We used the Zoom breakout feature to allow students to visit with their selected panelists in groups of 2-3. For more on this approach, see [9].
- *Getting into Grad School*: This session introduced the graduate school application process, and allowed students to ask detailed questions of faculty advisers.
- *Graduate Student Panel*: This session, presented without faculty present, allowed undergraduate students to ask current graduate students questions on the application process and life in grad school.
- *How (not) to give an effective science presentation*: Our final career building session of the week featured Dr. Oscar Schofield doing his best impression of a bad scientific talk. Students were asked to reflect on the

TABLE I. Virtual Workshop Schedule

Day 1	Day 2	Day 3	Day 4	Day 5
Introductions & Ice Breakers	How to ask a testable question	Data Analysis I – Intro to Python & Accessing Data	Data Analysis II – Data Visualization Formats	Career Panel
Program Overview	The Wonderful World of Data	Science Communication Introduction	Group Project Starter	Data Analysis III – Analysis Techniques
Day 6	**Day 7**	**Day 8**	**Day 9**	**Day 10**
Mid-Workshop Review	Getting into Grad School	Diversity, Equity and Inclusion	Group working session to finalize presentations	Group Presentations
Group work with faculty mentors	Graduate Student Panel	How (not) to give an effective science presentation		Workshop Wrap-up & Graduation

presentation, identifying ways they might be more effective in presenting their work. This exercise was also used as helpful guidance for the presentations students would give throughout the program.

Our student participants were also invited to participate in weekly webinars presented by the GEO REU program at NCAR, which included a series of workshops on Diversity, Equity and Inclusion, career training, harassment, and research ethics. These workshops also hoped to bring together participants from several virtual REU programs into a larger community [2].

C. Python Skill Building

When we initially designed the Virtual REU workshop, we decided that a major focus would be on building students' programming and data analysis skills. Our goal was to utilize a common programming language and example dataset to focus our teaching efforts and allow groups to have a common working environment. Python was chosen as the programming language, because of its rapid rise in many oceanographic research fields and the number of libraries now available that make accessing and visualizing data easier. More importantly, we also relied on Google Collaboratory, a free browser-based tool for executing Python code, which allowed students to quickly start coding without having to install software on their own machine.

Our goal for the workshop was not to teach Python programming, but rather to introduce students to data analysis using Python. After all, we only had a few hours. Thus, it was important for students to be able to understand the teaching examples. We choose to develop examples using data from the National Data Buoy Center (NDBC). This dataset was chosen for two reasons: 1) it comprises instruments that measure weather, tides, waves, and ocean temperature that are more relatable to students with limited oceanographic backgrounds, and 2) the data is readily available in common formats on a THREDDS server, making it easier to access with code.

The workshop included 3 sessions dedicated to data analysis with Python. For each session, a Python notebook was developed that included a variety of example code "recipes" for students to run and learn from. The notebooks included explanatory text to describe each code block. The first session was largely presented as a lecture, with students following along and running the notebook themselves and asking questions as they got stuck. This was a challenge in the virtual environment. In a classroom, it is easier for helpers to walk around the room to help students who are stuck, or for students to turn to their neighbor. It is also easier to quickly look over someone's shoulder to see what their problem might be. In a virtual room, it is essential to encourage students to speak up when they get lost or are stuck, as it is much harder to read the room. For later sessions, we decided to let students run through the notebooks on their own and ask questions, allowing them to drive the session. We also broke out into groups, enabling the smaller groups to help each other.

Dealing with varying skill levels is always a challenge in any programming course, and this was no different. While a majority of students had limited to no programming experience, a few had a substantial amount. Our strategy was to assemble groups with varying Python programming skills. That way there was at least one person with some familiarity in each group who could help mentor the others and keep the group on track.

The three programming sessions included 1) an introduction to Python programming and accessing data using the *xarray* and *pandas* libraries, 2) a quick overview of data visualization using the *matplotlib* library, especially focusing on time series graphs, and 3) a review of potential data analysis strategies for time series data, including bulk statistics, histograms, and hourly/daily/monthly means.

In the end, even the faculty mentors appreciated this part of the workshop, as it was a new skill for many of them as well.

D. Group Mini Data Projects

The 2-week workshop culminated with the presentation of group mini-projects. These projects allowed students to apply the research skills and Python data analysis techniques they learned during the first half of the workshop. Several sessions during the workshop were dedicated to group work, which included support from faculty mentors and programming experts. The division of tasks and meetings were largely left to each group. Several groups scheduled additional meetings on their own, though this was a challenge in the short timeframe. Groups were given the task of exploring their own questions using the atmospheric and oceanic data available from NDBC. In the end, all groups wished they had more time, but the amount of analysis they were able to accomplish in one week was actually quite impressive, especially given the circumstances. And each of the 5 groups chose a unique topic, which demonstrated the breadth of research questions possible with the NDBC dataset.

Ultimately, we found that the mini-projects and 2-week workshop proved essential in enabling students to get a jumpstart on their own research projects with their mentors in the following weeks.

VI. Student Research Projects

Following the workshop, students embarked on a 6-week research experience under the guidance of their mentor. This part of the program was similar to regular REUs, in that a positive mentor-mentee relationship was key to each students' success. The major differences included the shift to virtual collaboration (e.g. Zoom meetings instead of in-person), and an emphasis on data analysis and programming instead of more hands-on lab and/or field techniques (because of the limitations of the virtual program). Prior to the program, each student was matched with a mentor whose research interests most aligned with their own, ensuring they would be interested in their ultimate research project.

Eight of the students were interested in working on datasets from the OOI, with mentors from the Data Labs community. For these students, we dedicated an additional day to introduce them to the OOI, its instrumentation, and science goals. We also reviewed a series of example Python notebooks that demonstrated how to access and visualize data from the OOI. This was crucial to setting students, and their mentors

(many of whom were exploring data from the OOI in depth for the first time) off on the right foot for their projects.

Students were expected to check in with their mentors on a daily basis, and meet with them every 2-3 days at a minimum. Many mentors involved students in their lab meetings, giving them a chance to meet other students and postdocs throughout the experience. This allowed them to expand their research network even further.

We also scheduled 3 group meetings of the entire cohort (students and mentors) during the research period, so that each student could share their progress and raise any challenges they were facing. This provided an opportunity for everyone to share ideas and socially catch up.

The research portion culminated in a 2-day "Virtual Symposium," where each student presented their work. We decided to make the symposium 2 days to limit Zoom fatigue. Each student was asked to create a poster and pre-record a 5-minute presentation. We initially asked if they would prefer live talks, but the consensus was to record. Recording their presentations gave students the ability to see and edit their own work so they were more confident in the result. Posters were made available the day before for everyone to review, and on the day of the symposium, the recordings were presented. Following the recording, 10-minutes were available for the full group to ask questions of each presenter. Unlike a traditional talk, where only 2-3 minutes might be allotted for questions, this gave preference to a discussion of the work. In the end, we found this approach to be less intimidating and also more collaborative and encouraging.

VII. Discussion

Thanks to the support of the NSF, in the spring of 2020, we were able to quickly organize an 8-week virtual research experience for 16 students across the country, many of whom were rising seniors that would have missed out on an important opportunity to experience first-hand research without this alternative program. Planning a last-minute virtual REU was not ideal, nor was it the experience that students had hoped for when they initially applied to their traditional REU programs the previous winter. But because of this program, students were still able to have an opportunity to virtually meet with faculty mentors, and build their capacity to work with large online and openly accessible datasets.

In summary, we discovered that there were positive impacts of a virtual REU.

- Both students and mentors could participate from anywhere, regardless of family priorities or health concerns.
- Students could set their own schedule, outside of workshop sessions. With participants across the country, standard working hours are hard to define. The virtual format allowed students and mentors to find the times that worked best for them to collaborate.
- Breakout rooms worked very well to facilitate small group work. It was also easy for mentors and technical staff from around the country (each with their own expertise) to move from room to room to aid students in their work.
- Throughout the program we utilized the Slack platform for communication. Slack allowed us to set up different channels for each workshop mini-group, mentor/mentee team, a few larger research groups, the staff, and a few other ad hoc teams through the entire project. This allowed everyone to stay engaged, with one central place for communications.
- Many professional society meetings in 2020 also went virtual, which greatly lowered the cost of participating. As a result, several of our participants were able to present their work at the Fall AGU Meeting as well, gaining yet more experience presenting their research.

We also identified several considerations for future virtual programs.

- Virtual collaboration and especially community building is tough, but it is possible with proper facilitation and expectation setting. We were surprised at how our students took the time to work together, even outside of scheduled sessions. We found it helpful to dedicate time for both regular communication and social events.
- Regular check-ins are essential. This is no different from in-person internships. At a minimum, mentors should meet with their students every other day if not daily, to make sure they are on track. The full group should meet weekly to share their experiences and help build community.
- While we were successful with a focus on data analysis and programming, it was crucial to have staff on hand to help students with their code. This was true for both the group working sessions during the workshop, and on an ad hoc basis during the 6-week research portion. Some students are often shy about asking questions, especially virtually, so it is incumbent on staff and mentors to reach out to make sure students are properly supported.
- Virtual or hybrid REU may be more amenable to students with family commitments. Summer REU programs generally expect participants to spend 40 hours a week on their research, but does it have to be 9-5 for 10 weeks straight? An alternative approach might allow students to set their own hours, including evenings if necessary.
- A hybrid REU could include a weekend or week-long field experience, followed by virtual mentoring and research work. This approach might be more advantageous to non-traditional students, especially those with children who cannot spend 6-10 weeks at a remote research institute. It would give them time to actually conduct research in the field or laboratory, and meet with other students and mentors in their cohort for community building and support. The subsequent

weeks could then focus on data analysis and presentation building, which are more easily done virtually.

- We focused our program on data analysis and utilizing existing observatory datasets, because they were easier to integrate into a virtual program. There may be creative ways for remote students to conduct lab or field experiences (e.g. by sending them equipment to use in their local area). Otherwise, when designing remote or virtual programs, it is probably best to continue a focus on data analysis and programming skill development.

At the end of both the 2-week workshop and the full program, we conducted a formative evaluation with students to learn more about their experience. In general, the program met most students' expectations, although many were not sure what to expect given the irregular nature of the year. Most students were also very appreciative of the opportunity, especially given the circumstances. Eight students took the opportunity to present their work at the AGU Fall Meeting, and several have continued to work with their mentors on research papers.

In the end, the Data Labs-RIOS Virtual REU brought students and mentors together from across the US, including Alaska, Puerto Rico and Guam, and demonstrated that a virtual research experience is possible when well-designed and supported. As a result, the lessons we learned through this experience may be helpful for future virtual REUs or internship programs looking to provide an alternative to a traditional 8 to 10-week on-site experience that may not be accessible for all students. Indeed, the NSF has already indicated this may be a viable option for future REU programs in their latest request for proposals. We hope that the learned experiences we took from this irregular year will help in planning future virtual programs. Hopefully, they will have more time to prepare and develop even richer virtual experiences.

ACKNOWLEDGMENT

We would like to thank our fellow facilitators Daphne Munroe, Anthony Vastano, Janine Barr, and Silke Severmann, for their help in coordinating this virtual REU. We especially want to thank all of the faculty mentors who agreed to participate at short notice, and especially the students who fully engaged themselves in the program, even though it was not what they had initially applied for. We also want to acknowledge the support and efforts of Lisa Rom and Rennie Meyers at NSF, without whom this alternative experience for these students would not have happened.

REFERENCES

[1] National Science Foundation, "NSF REU Sites: Ocean Sciences," accessed July 22, 2021. https://www.nsf.gov/crssprgm/reu/list_result.jsp?unitid=5053

[2] V. Sloan, R. Haacker, R. Batchelor, and C. Garza, "How COVID-19 Is Affecting Undergraduate Research Experiences," Eos (Washington. DC)., vol. 101, Jun. 2020.

[3] 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.

[4] L. Smith et al., "The Ocean Observatories Initiative," Oceanography, vol. 31, no. 1, pp. 16–35, Mar. 2018.

[5] 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.

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

[7] FourSight, accessed July 22, 2021. https://foursightonline.com

[8] Right Question Institute, "What is the QFT?," accessed July 22, 2021. https://rightquestion.org/what-is-the-qft/

[9] J. Barr, C. Bean, and J. McDonnell, "Strategies for Running a Successful Virtual Career Panel," Oceanography, vol. 33, no. 2, Jun. 2020.