SPOTLIGHT

Deep Ocean Early-Career Researchers Dive into Data Management Training

By Stace E. Beaulieu, Leslie M. Smith, and Karen Stocks

INTRODUCTION

Open access to scientific data is increasingly recognized as critical to fostering scientific progress, trustworthy and reproducible science, global information equity, and evidence-based policymaking. It requires scientists to not only share their data, but to share in such a way that the data have high utility for later users. The FAIR data principles define a set of characteristics for making data "findable, accessible, interoperable, and reusable" (Wilkinson et al., 2016). Training scientists, particularly early-career scientists, on these principles can improve the volume and quality of open science data.

The Deep Ocean Observing Strategy (DOOS) is an international network of deep ocean observing, mapping, exploration, and modeling programs endorsed as a United Nations Decade of Ocean Science for Sustainable Development Programme. A central component of DOOS is the professional development of ~140 early-career researchers from ~30 countries called Deep Ocean Early-career Researchers (DOERs). Professional development activities include learning experiences as well as mentoring and networking to build a globally diverse and inclusive deep ocean observing community.

In 2023, DOOS offered an online workshop series for DOERs to learn about and discuss FAIR data principles and open science tools. Here, we provide an overview of the course as well as a set of key lessons learned. The DOERs data training syllabus and slide presentations are available online (Beaulieu et al., 2024), with recordings linked through the DOERs data training webpage within the DOOS website (Deep Ocean Observing Strategy, 2024). While some examples were specific to the deep ocean, the majority of the curriculum is generic and could be adapted to other disciplines.

FIGURE 1. Modules shown here relate to the research data life cycle and FAIR (findable, accessible, interoperable, and reusable) principles. Diagram modified from Stocks et al. (2023), under a CC-BY license.

COURSE DESCRIPTION

The DOERs data training course was designed for our next generation of leaders in deep ocean observing to gain knowledge and skills in research data management. Given the international audience, the course was free, virtual, platform agnostic, and had no prerequisites in order to minimize barriers to entry. It is a response to both the broader call for capacity building in the environmental sciences for data-intensive research skills (Hampton et al., 2017), as well as a specific need within the global deep ocean community (Bell et al., 2023).

The course was organized into four modules: (1) Foundational Practices for FAIR Data, (2) Collaborating in the Research Data Life Cycle, (3) Best Practices in the Ocean Sciences, and (4) The "R" in FAIR Data Life Cycle: Reusable Data. One module was delivered each quarter over the course of a year. Each module had two sessions, separated by a few days, with a "homework" assignment in between to deepen learning and facilitate discussion. All sessions were one hour (with the exception of combining the sessions for Module 2 into a single 1.5-hr session).

The modules step through a research data life cycle, mapping steps to corresponding FAIR principles (Figure 1). Content for the DOERs data training builds on an in-person workshop (Beaulieu et al., 2020), updated with content developed in collaboration with the Earth Science Information Partners (ESIP) Marine Data Cluster (Stocks et al., 2023). Although the examples used in the training are specific to the ocean sciences domain, many of the resources are more generic, and the training can be adopted or modified for use in other domains in the environmental sciences by selecting examples relevant to the audience (Table 1).

Learning objectives for course participants included:
- Apply FAIR data principles in some way(s) to their own research data
- Understand how FAIR data principles are related to many aspects of their own research data life cycle
- Remember concepts of FAIR data, open science, and steps in the research data life cycle

Communications were sent to all DOERs ahead of each session, including links to the slides and a pre-workshop recording of the slide presentation to help those not fluent in English and to be more inclusive of different time zones. During the live online sessions, we prioritized interactive activities such as polls, "pulse reactions" with emojis, a collaborative digital whiteboard, and breakout rooms. We invited data professionals from DOOS and ESIP to join the breakout rooms, aiming to include a mentor in each breakout room to guide and "normalize data discussions" (sensu Lowndes et al., 2019). We also enabled live captions to improve accessibility.

OUTCOMES AND LESSONS LEARNED

Informal pre- and post-course anonymous surveys were conducted in order to gauge participants' growth in understanding and use of FAIR data principles. Additionally, participants were encouraged to
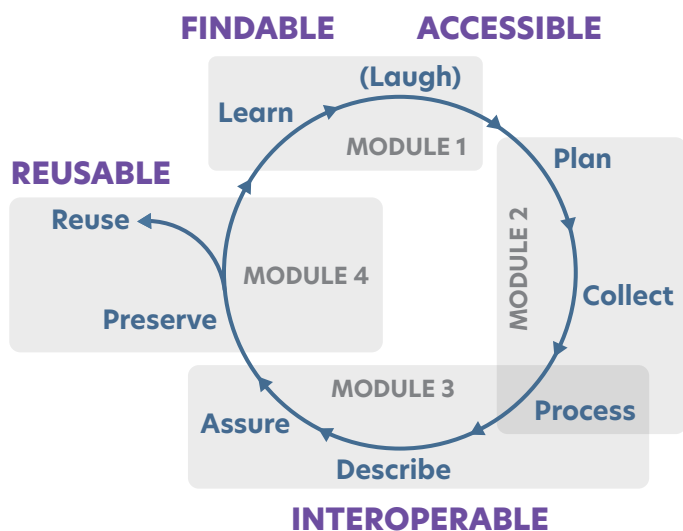
TABLE 1. Selected examples used in the data training course, associated to steps in the data life cycle (Figure 1) and FAIR principles (Wilkinson et al., 2016).

MODULE(S)	EXAMPLE USED IN TRAINING	RELEVANT STEP(S) IN DATA LIFE CYCLE	RELEVANT FAIR PRINCIPLE(S)
1	Find and access data used in a journal article	Reuse*	Findable Accessible
2	Data management plan	Plan	
2, 3	Hosted version control as collaboration tool	Collect Process Assure	
3	Transform data into standard formats common in ocean sciences domain	Process	Interoperable
3, 4	Search for a controlled vocabulary term for a parameter (and keyword)	Describe Preserve	Interoperable (Findable)
4	Pull data from domain repository directly into software environment using web Application Programming Interface (API)	Reuse	Reusable

* Module 1 is Learn (and Laugh) on Figure 1, but the selected example used in training starts to develop the concept of the Reuse step.

provide feedback during the course through an online anonymous form. Comparing pre- and post-surveys, participants expressed an increase in confidence in their ability to make their own data FAIR. In the pre-survey, about half of the participants were neutral in their level of confidence, 40% did not feel confident, and only 10% expressed confidence. In the post-survey, nearly 80% felt confident, with the remaining feeling neutral. Overall, participants enjoyed the course, noting the real-world examples, interactivity, breakout rooms, and flexibility of materials as highlights.

Lesson 1: Select high value, low lift, real-world examples. It was important to address both knowledge and skills in the limited total of eight hours for the DOERs course, as such Learning by DOing (i.e., working with real-world examples) was an important element. However, selecting these activities had to be done with care, as participants came in with a wide variety of backgrounds and skill levels, and examples had to be quickly absorbed (i.e., low ramp up time). For specific examples matching concepts in Table 1, we used publicly accessible resources with open licenses for live demonstrations.

Lesson 2: Strategically plan breakout rooms. Breakout rooms were a critical element of the course, not only to help participants learn the material but also to provide an opportunity for mentoring and networking. We offered a range of breakout topics to appeal to a variety of disciplines, and we deliberately set the tone for a welcoming inclusive discussion. We aimed for a mentor-to-participant ratio of 1:5 in the breakout rooms, though this required a bit of flexibility as without requiring registration we did not know how many DOERs would be in attendance.

Lesson 3: Consistent attendance is a challenge. There are so many online trainings, webinars, and early-career groups, it is understandable that early-career researchers with limited time available are pulled in many directions. This training was no different, with a decline in attendance each quarter. Reflecting after the fact, it may have been beneficial to have delivered modules at higher frequency, for example, one module every two to three weeks as in the Openscapes Champions program (https://openscapes.org/champions-program), or to have provided a certificate of completion or other incentive, such as the digital badge offered by NASA's Open Science 101 curriculum (https://nasa.github.io/Transform-to-Open-Science/).

CONCLUDING REMARKS

The central guiding principle of DOOS is to seek collective solutions to global challenges and in this way move the deep ocean observing community forward through collaboration, sharing, and building together. We hope that this course and this paper help continue to grow the use of FAIR data across the entire ocean community. We have many great challenges to tackle—finding and using data should not be chief among them.

REFERENCES

Beaulieu, S.E., L. Raymond, A. Mickle, J. Futrelle, N. Symmonds, R. Mazzoli, R. Brey, D. Kinkade, and S. Rauch. 2020. "Data Science Training Camp at Woods Hole Oceanographic Institution: Syllabus and Slide Presentations in 2020." Woods Hole Open Access Server, https://doi.org/10.1575/1912/26103.

Beaulieu, S.E., K. Stocks, and L.M. Smith. 2024. "FAIR Data Training for Deep Ocean Early Career Researchers: Syllabus and Slide Presentations." Woods Hole Open Access Server, https://doi.org/10.1575/1912/67631.

Bell, K.L.C, M.C. Quinzin, D. Amon, S. Poulton, A. Hope, O. Sarti, T.E. Cañete, A.M. Smith, H.I. Baldwin, D.M. Lira, and others. 2023. Exposing inequities in deep-sea exploration and research: Results of the 2022 Global Deep-Sea Capacity Assessment. *Frontiers in Marine Science* 10:1217227, https://doi.org/10.3389/fmars.2023.1217227.

Deep Ocean Observing Strategy. 2024. "FAIR Data Training: Deep Ocean Early-career Researchers (DOERs)." https://www.deepoceanobserving.org/pages/doers-data-training.

Hampton, S.E., M.B. Jones, L.A. Wasser, M.P. Schildhauer, S.R. Supp, J. Brun, R.R. Hernandez, C. Boettiger, S.L. Collins, L.J. Gross, and others. 2017. Skills and knowledge for data-intensive environmental research. *BioScience* 67:546–557, https://doi.org/10.1093/biosci/bix025.

Lowndes, J.S.S., H.E. Froehlich, A. Horst, N. Jayasundara, M.L. Pinsky, A.C. Stier, N.O. Therkildsen, and C.L. Wood. 2019. Supercharge your research: A ten-week plan for open data science. *Nature*, https://doi.org/10.1038/d41586-019-03335-4.

Stocks, K., S. Beaulieu, C. Berys-Gonzalez, M. Biddle, A. Mills, S. O'Connor, L.M. Smith, M. Vardaro, and D.J. Wright. 2023. "Easy Resources for Managing Your Ocean Data." Earth Science Information Partners, https://doi.org/10.6084/m9.figshare.21766709.v2.

Wilkinson, M.D., M. Dumontier, I.J. Aalbersberg, G. Appleton, M. Axton, A. Baak, N. Blomberg, J.-W. Boiten, L. Bonino da Silva Santos, P.E. Bourne, and others. 2016. The FAIR guiding principles for scientific data management and stewardship. *Scientific Data* 3:160018, https://doi.org/10.1038/sdata.2016.18.

AUTHORS

Stace E. Beaulieu (sbeaulieu@whoi.edu), Woods Hole Oceanographic Institution, Woods Hole, MA, USA. **Leslie M. Smith**, Your Ocean Consulting LLC, Knoxville, TN, USA. **Karen Stocks**, Scripps Institution of Oceanography, University of California, San Diego, CA, USA.