HACKING LIMNOLOGY WORKSHOPS AND DSOS23

Growing a Workforce for the Nexus of Data Science, Open Science, and the Aquatic Sciences

Michael F. Meyer, Merritt E. Harlan, Robert T. Hensley, Qing Zhan, Carolina C. Barbosa, Nahit S. Börekçi, Jonathan J. Borrelli, Tuba Bucak, Alli N. Cramer, Johannes Feldbauer, Robert Ladwig, Jorrit P. Mesman, Isabella A. Oleksy, Rachel M. Pilla, Jacob A. Zwart, Elisa Calamita, Nicholas J. Gubbins, Mary E. Lofton, Daniel A. Maciel, Nicholas S. Marzolf, Freya Olsson, Audrey N. Thellman, R. Quinn Thomas, and Michael J. Vlah

The 3rd Aquatic Ecosystem MOdeling Network—Junior (AEMON-J) Hacking Limnology Workshop and 4th Virtual Summit: Incorporating Data Science and Open Science in the Aquatic Sciences (DSOS) took place 24–28 July 2023. This joint event followed a similar structure to previous years, comprising three days of workshops followed by two days of the virtual summit (Meyer and Zwart 2020; Meyer et al. 2021*a*, 2022). During the week, over 100 aquatic science practitioners and enthusiasts gathered to exchange knowledge and to share experiences working in the nexus of data science, open science, remote sensing, and the aquatic sciences.

As in previous years, accessibility for a global audience was at the forefront of both events. To prevent financial restrictions, there were no registration fees. All recordings and workshop materials were made freely available on the Open Science Framework archive (Meyer et al. 2021*b*; https://osf.io/682v5/) for asynchronous viewing and access following the event. Having videos subtitled helped overcome language barriers, allowing attendees to translate subtitles from English into their preferred language using available software. Relative to the 50 countries represented in the 2021 and 2022 workshop and summit (Meyer et al. 2021*a*, 2022), the 2023 workshop and summit experienced an expansion in global representation, with registrants from more than 60 countries. The most represented countries included the United States (42% of registrants), Canada (8.2%), Nigeria (8.2%), Germany (5%), and Brazil (3.8%). The majority of registrants from newly represented countries were largely located in the Caribbean. With respect to career-stage, most participants identified as being in early career positions (i.e., graduate students and postdoctoral researchers; ∼ 57%).

Leading up to the 2023 summit, the leadership team began a period of restructuring, a process aimed to usher in fresh perspectives and formalize the group as a community of practice (Fig. 1). This restructuring included an expansion of the leadership teams, including a steering team, a summit team, a workshop team, and an external advisory team. This expansion was realized in part to further establish the growing AEMON-J and DSOS communities, as well as enable clearer succession paths, new leadership opportunities, and organizational sustainability. The steering team, as highlighted in Fig. 1, will be instrumental in advising long-term development and goals, ensuring the community can move past short-term planning for individual summits to envision and achieve long-term goals. As the combined AEMON-J + DSOS organizing teams strive to be composed of early career researchers, this long-term planning is crucial for the broader community. To involve community members in the strategic development, a listening session was organized as a part of the 2023 summit. The goal of the listening session was to receive feedback from community members on what they appreciated about the workshops and summit. By engaging community feedback, the DSOS and AEMON-J leadership teams sought to create a more indelible mark in the nexus of data science, open science, as well as the larger aquatic sciences community by keeping the diversity, equity, inclusion, and accessibility of the larger community of practice at the forefront.

THE "HACKING LIMNOLOGY" WORKSHOP

The AEMON-J "Hacking Limnology" workshop series ran during the first three days of the week. The goals of the workshops were to introduce early career scientists to various open datasets and to provide coding skills necessary to access and work with the data. The structure of the workshops was similar to previous years (Meyer et al., 2022; Meyer et al., 2021*a*,*b*). For each day, a pre-recorded introductory video was made available to watch the week before the workshop, so participants could familiarize themselves with the fundamental concepts needed. A pre-recorded, 15-minute keynote presentation was live streamed, followed by a live question-and-answer session with both the introductory and keynote speakers. After a short break, the bulk of each workshop consisted of a two-hour, live coding session, where attendees were encouraged to live code along on their own machines. Each day concluded with a virtual social hour, during which attendees could ask further questions, network, and discuss future collaborations.

As in past years, each day had a specific theme. The theme for the first day was "Remote Sensing" and focused on how satellite-based observations, coupled with machine learning, can be used to monitor water quality over large spatiotemporal scales (Lehmann et al., 2023). The theme of the second day was "Ecological Forecasting" and focused on the Ecological Forecasting Initiative's National Ecological Observatory Network Ecological Forecast Challenge (Thomas et al., 2023), aimed at generating real-time forecasts of water temperature and dissolved oxygen in freshwater lakes across the continental United States. The theme of the third day was "Catchment Modeling" and focused on the MacroSheds dataset (Vlah et al., 2023), which contains discharge and chemistry data from hundreds of watersheds and can be used to estimate solute fluxes.

All workshop materials (introductory and plenary videos, recording of live coding sessions, links to Github repositories with R code) are archived and freely available on the Open Science Framework repository (Meyer et al., 2021*b*; https://osf.io/682v5/), where they can serve as useful

Michael F. Meyer, Merritt E. Harlan, Robert T. Hensley, and Qing Zhan contributed equally and are listed as co-first authors.

[Correction added on Nov 22, 2023, after first online publication: Carolina C. Barbosa and Jonathan J. Borrelli has been added as co-authors along with their affiliations.]

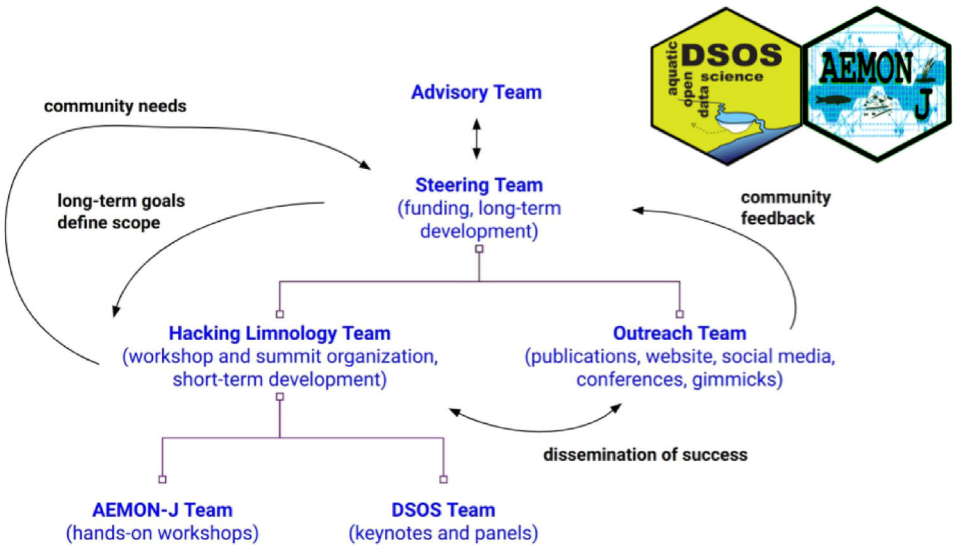

FIG. 1. Organigram for the combined AEMON-J + DSOS research coordination network.

training material for independent study as well as organized graduate training (Table 1).

THE VIRTUAL SUMMIT FOR INCORPORATING DATA SCIENCE AND OPEN SCIENCE IN THE AQUATIC SCIENCES

The 4th Virtual Summit for Incorporating Data Science and Open Science in the Aquatic Sciences (DSOS23) ran 27–28 July 2023 and featured 17 speakers who presented on the various ways they engaged in the nexus of data science, open science, and the aquatic sciences. The talks were divided into four sessions: "Big Data," "Data Intensive Models," "Applications of Open Science," and, new this year, "Remote Sensing of Aquatic Environments". Similar to previous years, 4–5 prerecorded 10-minute talks were streamed sequentially, followed by a 20-minute live Q&A session moderated by organizers. The Q&A session was entirely in English, although live captions could be generated during the session with dictation software. As in the case of the Hacking Limnology workshops, talks include subtitles and are archived in the Open Science Framework repository (Meyer et al., 2021*c*). Topics within the sessions ranged from metagenomic sequencing to modeling of lake evaporation, runoff-driven soil erosion, optimal stream gage placement, phytoplankton succession, lake metabolism, as well as satellite-derived water quality estimates and community science earth observations. Each presentation highlighted the importance of access and maintenance of open data as well as reproducible science.

The summit included a career panel highlighting careers in data science and open science. Panelists included Joshua Fisher (Hydrosat and Chapman University), Stephen Kennedy (UrbanFootprint), Roxane Maranger (University of Montreal), Anika Pyle (Colorado State University), and Emily Read (U.S. Geological Survey). This group spanned a wide range of careers, perspectives, and approaches to the data-intensive sciences, and many focused on the importance of data visualization to increase accessibility and develop common ground with interdisciplinary work. Discussion included advice that panelists would tell their former selves, perspectives on leading interdisciplinary teams, suggestions for being creative in data visualization and interpretation, and strategies for keeping data visualization fun and innovative.

The summit also included a community listening session, where members of the DSOS community could share their experiences with the present and past summits. The three main goals of the listening session were to identify (1) why participants come to Hacking Limnology and DSOS, (2) what participants wish existed or considered unnecessary, and (3) who might be interested in leadership roles for the community. Approximately, 30 community members outside of the organizing committee contributed feedback. They highlighted the unique format, prerecorded and captioned videos, the welcoming community, accessible workshops, and ambitious young researchers as reasons why they decided to join the summit. A challenge for participants was accessible time zones, especially for those located in Australia and Asia.

TABLE 1. The Open Science Framework archive houses all recordings, scripts, and subtitle files from the previous Hacking Limnology Workshops and DSOS Virtual Summits. The repository contains over 30 hours of training material and over 10 hours of conference talks, all of which distill several key advances in the nexus of data science, open science, remote sensing, and the aquatic sciences. As materials are intended for an intermediate skill level, the joint AEMON-J + DSOS website contains over 20 hours of introductory level material for participants to learn fundamental tools in preparation for the workshops (https://aquaticdatasciopensci.github.io/material/)

Introductory material	Introduction to Git/Github and the Tidyverse Introduction to markdown Introduction to statistical modeling Introduction to workflow management software Making webpages in Github
Remote sensing	Hacking Limnology Workshops 2021, 2022, and 2023 DSOS 2023 session "Remote Sensing of Aquatic Environments"
Big Data	Hacking Limnology Workshops 2021 DSOS 2020, 2021, 2021, 2023 session "Big Data"
Machine learning + modeling + data science	Hacking Limnology Workshops 2021, 2022, and 2023 DSOS 2020, 2021, 2022, and 2023
Tool and software development	DSOS 2020, 2021, and 2022
Open science	DSOS 2020, 2021, 2022, and 2023

MOVING FORWARD

Just like previous workshops and summits, the AEMON-J Hacking Limnology Workshop and DSOS23 attracted wide interest and earned overwhelmingly positive feedback from the community. Yet, much like the 2022 workshops and summit, synchronous engagement was lower than events in 2020 and 2021.

While over 550 people registered for the combined workshops and summit events, approximately 100 registrants (∼ 18%) engaged synchronously each day. This decrease from ∼ 22% in 2020 and 2021 to ∼ 18% in 2022 and 2023 possibly stems from the decline of restrictions associated with the COVID-19 pandemic, namely the increase in face-to-face conferences and workshops. We also observed declining attendance through the week, potentially reflecting a need to balance regular work requirements with a week-long virtual meeting.

Admittedly, these same challenges will likely continue to be common for most virtual events, and so the organizational team is committed to striking a balance between in-person and synchronous, virtual engagement. For example, a potential opportunity includes coordinating events with in-person meetings, while still streaming the event virtually to enable worldwide participation. In-person meetings can build on the momentum generated by virtual DSOS and AEMON-J events as well as the in-person social hours organized at the Joint Aquatic Sciences Meeting in 2022 and the ASLO 2023 Meeting. Also, by distributing events throughout the year and empowering community members to self-organize their own events, the organizational team is keen to create opportunities for engagement beyond just the Hacking Limnology workshops and virtual summit. The organizational team encourages those interested in joining the network or organizing potential workshops to contact an organizational team member directly (https://aquaticdatasciopensci.github.io/info/).

With the growth and potential for the combined AEMON-J and DSOS community, the organizational team is encouraged by the breadth and number of community members interested in realizing this growth. Numerous community members expressed interest in helping with social media presence, organizing workshops, and finding speakers. Likewise, the newly formed Advisory Team is intended to increase the footprint of AEMON-J and DSOS throughout the environmental sciences, thereby enabling new connections and collaborations, as well as longer-term visions. Furthermore, the increasing number of first-year organizational team members means that new ideas, topics, and initiatives can push forward, while also allowing founding members to focus on the sustainability of the larger community and provide institutional memory where necessary.

ACKNOWLEDGMENTS

We would like to thank Hilary A. Dugan, Jill S. Baron, Stephanie E. Hampton, Mark D. Scheuerell, Cayelan C. Carey, Dexter W. Howard, Kaelin Cawley, Matthew R.V. Ross, Rafael Marcé, and Jordan S. Read for diverse creative and technical support throughout the development of DSOS and AEMON-J, as well as Wayana Dolan for providing a review of this manuscript prior to submission. MFM and MEH were supported by Mendenhall Fellowships from the US Geological Survey's Water Mission Area. RMP was supported by the US Department of Energy (DOE), Office of Energy Efficiency and Renewable Energy, Water Power Technologies Office, and Environmental Sciences Division at Oak Ridge National Laboratory (ORNL). ORNL is managed by UT-Battelle, LLC, for the US DOE under contract DE-AC05-00OR22725. NSB was supported by the YÖK 100/2000 PhD Scholarship from the Council of Higher Education of Türkiye. QZ was supported by the European Union's H2020 DRYvER (grant number: 869226). Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the US Government.

References

Lehmann, M. K., and others. 2023. GLORIA—A globally representative hyperspectral in situ dataset for optical sensing of water quality. Sci. Data **10**: 100. doi:10.1038/s41597-023-01973-y.

Meyer, M. F., and Zwart, J. A. 2020. Virtual summit: Incorporating data science and open science in aquatic research. Limnol. Oceanogr. Bull. **29**: 144–146. doi: 10.1002/lob.10411

Meyer, M. F., and others. 2021*a*. Virtual growing pains: initial lessons learned from organizing virtual workshops, summits, conferences, and networking events during a global pandemic. Limnol. Oceanogr. Bull. **30**: 1–11. doi: 10.1002/lob.10431.

Meyer, M.F., and others, 2021*b*. The AEMON-J "Hacking Limnology" workshop series & virtual summit: Incorporating data science and open science in aquatic research. Limnol. Oceanogr. Bull. **30**, 140–143. doi:10.1002/lob.10475

Meyer, M., and others. 2021*c*. AEMON-J/DSOS archive: "Hacking Limnology" workshop + virtual summit in data science & open science in aquatic research. doi: 10.17605/OSF.IO/682V5.

Meyer, M. F., and others. 2022. Hacking Limnology Workshop and DSOS22: Creating a community of practice for the nexus of data science, open science, and the aquatic sciences. Limnol. Oceanogr. Bull. **31**: 123–126. doi:10.1002/lob.10525.

Thomas, R. Q., and others. 2023. The NEON ecological forecasting challenge. Front. Ecol. Environ. **21**: 112–113. doi:10.1002/fee.2616.

Vlah, M. J., S. Rhea, E. S. Bernhardt, W. Slaughter, N. Gubbins, A. G. DelVecchia, A. Thellman, and M. R. V. Ross. 2023. MacroSheds: A synthesis of long-term biogeochemical, hydroclimatic, and geospatial data from small watershed ecosystem studies. Limnol. Oceanogr. Lett. **8**: 419–452. doi:10.1002/lol2.10325.

Michael F. Meyer, U.S. Geological Survey, Madison, WI, USA; mfmeyer@usgs.gov

Merritt E. Harlan, U.S. Geological Survey, Denver, CO, USA

Robert T. Hensley, National Ecological Observatory Network, Battelle, Boulder, CO, USA

Qing Zhan, The Netherlands Institute of Ecology, Wageningen, The Netherlands

Carolina C. Barbosa, University of Wyoming, Laramie, WY, USA

Nahit S. Börekçi, Mersin University, Mersin, Türkiye

Jonathan J. Borrelli, Rensselaer Polytechnic Institute, Troy, NY, USA

Tuba Bucak, Aarhus University, Aarhus, Denmark

Alli N. Cramer, University of Washington, Friday Harbor, WA, USA

Johannes Feldbauer, Technische Universität Dresden, Dresden, Germany

Robert Ladwig, University of Wisconsin—Madison, Madison, WI, USA

Jorrit P. Mesman, Uppsala University, Uppsala, Sweden

Isabella A. Oleksy, University of Colorado—Boulder, Boulder, CO, USA

Rachel M. Pilla, Oak Ridge National Laboratory, Oak Ridge, TN, USA

Jacob A. Zwart, U.S. Geological Survey, San Francisco, CA, USA

Elisa Calamita, Eawag, Dübendorf, Switzerland

Nicholas J. Gubbins, Colorado State University, Fort Collins, CO, USA

Mary E. Lofton, Virginia Tech, Blacksburg, VA, USA

Daniel A. Maciel, National Institute for Space Research, São José dos Campos, São Paulo, Brazil

Nicholas S. Marzolf, Duke University, Durham, NC, USA

Freya Olsson, Virginia Tech, Blacksburg, VA, USA

Audrey N. Thellman, Duke University, Durham, NC, USA

R. Quinn Thomas, Virginia Tech, Blacksburg, VA, USA

Michael J. Vlah, Duke University, Durham, NC, USA