BACKGROUND AND OVERVIEW

We can all likely agree that 2020 is a year of change, both in personal as well as professional lives. While it is easy to dwell on the negatives, the year has been ripe with creative and entrepreneurial opportunities to embrace new forms of sharing science. In particular, virtual conferences and summits have opened a space where researchers, students, and curious individuals across the globe can engage in open discourse. In doing so, the increasing availability and transparency of virtual tools have allowed scientific communities not only to connect in a forward-thinking way, but also to overcome barriers to participation that have long existed, such as travel and registration fees.

With these needs and opportunities in mind, a grassroots group of scientists convened for the first “Virtual Summit: Incorporating Data Science and Open Science in Aquatic Research” from 23 through 24 July 2020. For those who may be curious but less familiar, data science combines mathematics and statistics, computer science, and domain expertise to enable insights for problems that are otherwise too computationally demanding or data-intensive to be analyzed with traditional tools (Hayashi 1998). Open science is the practice of making tools that enable transparency into scientific design, analysis, and reporting, such that future researchers—and curious individuals in general—can access and reproduce others’ work (Bartling and Friesike 2014). Together, data science and open science techniques allow aquatic researchers to tackle complex aquatic problems while increasing scientific transparency and efficiency between researchers and stakeholders. By bringing together speakers who practice data science and open science techniques, this virtual summit was intended to showcase how limnologists and oceanographers work with big data, expand modeling frameworks, develop tools and software for the larger community, and inform natural resource management and monitoring.

The virtual summit featured 18 prerecorded presentations split across two, 3-h sessions on successive days. There were also live question and answer sessions immediately following the presentations. The talks were divided into four main themes: (1) Big Data, (2) Data-Intensive Models, (3) Tools and Software, and (4) Applications of Open Science. While many of the presentations incorporated multiple themes, we broadly grouped talks such that the first day of the virtual summit pertained mostly to “Data Science,” and the second day centered around “Open Science” topics (Fig. 1). Following presentations on July 24, breakout groups paralleling the summit’s four themes allowed for participants and presenters to discuss certain data science and open science topics in a more casual setting.

COMMUNITY RESPONSE

Admittedly, we were not sure how many participants the virtual summit would attract and were enthused by the community response. We advertised for the virtual summit through professional networks that participants had to sign up for the virtual summit during a 3-week registration period in order to receive an invitation to a password-protected video-conferencing link (Fig. 2). Most registered during the initial advertising campaign, and we quickly reached our initial virtual meeting capacity of 300 people, which was then expanded to a 500-person capacity. Over 425 people registered for the virtual summit, and there were between 125 and 160 attendees each day as some registrants did not attend. Although most registrants were from North America, all continents except Antarctica were represented.

Throughout the virtual summit and immediately following, we received verbal and written feedback from a subset of attendees about their virtual summit experience. In general, attendees reported that they really appreciated the virtual format with many participants noting they had a great experience and that they would be interested in future virtual summits centered around aquatic data science and open science.

What worked well

Virtual summit attendees reported that they appreciated the breadth of talks, which ranged from “big picture” surveys to “zoomed in” descriptions of specific analyses or tools. From a logistical perspective, many attendees reported that they appreciated that talks were prerecorded. During the virtual summit, we queued all prerecorded talks for each session in advance, allowing the presentations to stream...
sequentially over a shared video conference screen. For attendees having bandwidth issues, we also shared links to private YouTube playlists for both the data science and open science sessions, where talks could be accessed during and following the virtual summit. Having talks in hand prior to the summit also allowed us to adhere to our schedule, which attendees noted as helping the summit run efficiently. Having prerecorded talks also afforded us the chance for closed-captions presentations prior to the summit. Numerous attendees commented that the use of closed-captions was particularly helpful, especially for non-native English speakers.

Aside from presentations, attendees reported that they appreciated the interactive question and answer sessions that immediately followed each of the four sessions. Prior to streaming the prerecorded talks, we sent session-specific Google forms to attendees, which allowed them to submit questions to particular speakers. While talks were streaming, session presenters and moderators had access to session-specific response documents, which would automatically populate questions as they were submitted. Following each sessions’ presentations, moderators asked submitted questions to each speaker, who were given approximately 4 min to respond or expand on their presentation. Attendees reported that they found this format very interactive, time efficient, and effective in ensuring equitable opportunity in asking questions, especially in instances where individuals may not feel comfortable voicing a question in front of an audience.

With respect to breakout groups, participants reported that they enjoyed the opportunity to discuss data science and open science topics in detail with presenters and other participants. In particular, the “Software and Tools” breakout group highlighted a need for practical training in data science and open science techniques. While the immediate group discussed resources for various skill levels, this group spurred an idea of how future virtual summits could include training workshops, targeted at priming attendees with fundamental, hands-on experience in numerical, machine learning, and statistical modeling approaches.

What needed to be improved

Although the prerecorded talks made the sessions efficient and on time, attendees suggested a transition between speakers would likely have helped the audience reset mentally and finish note taking. While we were concerned about stopping and starting talks may increase risk for technical glitches, attendees noted that even a brief introduction to the successive speaker would be beneficial in orienting thoughts toward a new topic.

Admittedly, our finalized speaker list was not as diverse as was originally intended. Our original speaker solicitation conducted at the beginning of the COVID-19 pandemic was relatively balanced with respect to diverse sex and gender representation; however, that diversity was not realized in the summit’s speaker list, which was finalized during the pandemic. Virtual summits, especially during the COVID-19 pandemic, pose additional or exacerbate existing barriers to participation that are considered sex-biased, such as childcare or eldercare (Malisch et al. 2020), and are heightened for intersectingitized identities (Louisas and Marrast 2020; Stanisicuaski et al. 2020). Advanced notice for summit participation, relaxing the requirement for prerecording talks, flexible presentation times (e.g., evenings, weekends), or options for speaking on a panel rather than organizing a talk could help overcome participation barriers posed by the COVID-19 pandemic and virtual summits in general. Future summit or conference organizers may improve diversity and inclusivity by carefully considering the implications if participant diversity metrics are not met (Stadnyk and Black 2020).

Virtual conferences provide opportunity for global participation while avoiding travel-associated fees. However, meeting times for when the virtual summit or conference is hosted can be a barrier to participation for globally distributed participants. For example, several participants from India and the Philippines were joining in the early morning hours local time, and many other registrants likely could not participate due to the time of day we hosted the summit. Making available recordings of sessions or presentations can help with disseminating talks to a globally distributed audience, but interactive participation (e.g., live Q&A sessions) will be biased toward certain time zones. One option for increasing interactive participation from a globally distributed audience is to follow the model by the Global Lake Ecological Observatory Network (GLEON) virtual conferences, where they rotate meeting times to accommodate time zones across the world in which participants are concentrated.

For breakout groups, participants suggested that having predetermined goals or discussion topics would have helped guide conversation, especially as many graduate students were newer to breakout groups or “unconference” formats. While we preemptively assigned facilitators for each of the breakout groups, including concrete yet flexible discussion topics in advance may help facilitate conversation, such that more junior colleagues can more easily orient to small breakout group formats.

NEXT STEPS

Based on the attendance, overwhelming positive feedback from attendees, and enthusiasm for these topics, we have identified opportunity for future engagement in the aquatic sciences community—a space where modelers from numerical, machine learning, and statistical backgrounds can converge, discuss ideas, and share experiences centered around data science and open science topics.

For the immediate future, there is a virtual workspace to encourage active engagement among this newly formed community throughout the year all interested participants are welcome. This virtual workspace can be a space for mentoring and networking, where new data science and open science practitioners can get in touch with more experienced scientists, as well as troubleshooting and brainstorming, where researchers can crowdsource errors and new ideas.

In addition to the virtual workspace, future virtual summits could continue engagement among members of this aquatic science community. One variant is to host punctuated mini-summits halfway between annual virtual summits. Future meetings can be improved by continuing to receive input from the aquatic community with respect to ideas for programming as well as help in coordinating future virtual summits.

ACKNOWLEDGMENTS

We would like to reiterate our gratitude to all the speakers and participants, as well as to everyone who shared excitement for the virtual summit with their colleagues and collaboration networks, regardless whether or not they were able to attend. We thank the speakers who put in the time to prerecord talks and make their science accessible to all participants. We also thank

---

Footnotes:

1https://www.youtube.com/playlist?list=PL52fZTCEWS4muaL7jYLxgESZ2KT-v12C.
2https://forms.gle/YE16p5PDLDyjnxWkxR7.
3https://www.youtube.com/playlist?list=PL52fZTCEWS54kkg9mWUaM7gM6zhr0HB.
all the participants who carved out time from their schedule to join in this event. We know many people tuned in from many different time zones, where it was very early in the morning or in the middle of the night. We appreciate your enthusiasm to engage with this exciting science. We would also like to thank Stephanie E. Hampton for helping us develop this Bulletin piece. Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

References

Bartling, S., and S. Friesike [eds.]. 2014. Opening science. Springer International Publishing.

Hayashi, C. 1998. What is data science? Fundamental concepts and a heuristic example. In Data Science, Classification, and Related Methods, pp. 40–51, Kobe, Japan. Springer.

Louisias, M., and L. Marrast. 2020. Intersectional identity and racial inequality during the COVID-19 pandemic: Perspectives of Black physician mothers. J. Womens Health 29: 1148–1149. https://doi.org/10.1089/jwh.2020.8677.

Malisch, J. L., and others. 2020. Opinion: In the wake of COVID-19, academia needs new solutions to ensure gender equity. Proc. Natl. Acad. Sci. USA 117: 15378–15381. https://doi.org/10.1073/pnas.2010636117.

Stadnyk, T., and K. Black. 2020. Lost ground: Female academics face an uphill battle in post-pandemic world. Hydrol. Process. 34: 3400–3402. https://doi.org/10.1002/hyp.13803.

Staniscuaski, F., and others. 2020. Gender, race and parenthood impact academic productivity during the COVID-19 pandemic: From survey to action. bioRxiv 2020.07.04.187583. doi: https://doi.org/10.1101/2020.07.04.187583

Michael F. Meyer†, School of the Environment, Washington State University, Pullman, WA, USA; michael.f.meyer@wsu.edu

Jacob A. Zwart†, Integrated Information Dissemination Division, U.S. Geological Survey, Middleton, WI, USA