

The University of Washington eScience Institute was established in 2008 with the mission to empower researchers and students in all fields to answer fundamental questions through the use of large, complex, and noisy data. Today, thanks in significant part to the support of the Gordon and Betty Moore Foundation, the Alfred P. Sloan Foundation, and the Washington Research Foundation, the eScience Institute has grown into a mature organization with sustained impact on the UW and beyond through our education, research, and community building programs. This impact has been recognized by the Provost's decision to provide additional permanent funding to the eScience Institute, ensuring longevity for the suite of programs and initiatives that helped to establish UW's reputation as a leader in data-intensive discovery.



In March of 2020 Astronomy Professor Andrew Connolly was appointed as the new Director of the eScience Institute and the UW Associate Vice Provost for Data Science. Connolly has been involved since the early days of eScience as a member of the original Steering Committee and as a member of the Executive Committee since its inception in 2013. Below we highlight recent accomplishments as well as how we have addressed the challenge of transitioning, and sometimes pivoting, our work in light of the ongoing pandemic.

Data Science in the Community: Data Science for Social Good

Leveraging lessons learned from more than 10 university “data for good” programs and 6 years of running our own [UW Data Science for Social Good program](#), we are creating a "D4G Roadmap" —a guide for organizations interested in developing similar programs that pair student fellows with data scientist mentors and project leads from academia, government, and nonprofits. These types of programs serve the dual purposes of advancing projects with positive societal impact and providing real-world data science training for students from diverse disciplinary backgrounds. Past D4G projects cover a breadth of topics such as transportation, public health, urban planning, voting rights, and disaster response. Release of the D4G Roadmap will be facilitated through our role as a PI Institution for the NSF [West Big Data Innovation Hub](#).

Moving Data Science Education Online: Virtual Hackweeks

The [hackweek model](#) has emerged within the data science community as a powerful tool for fostering the exchange of ideas in research and computation. In contrast to conventional conferences or workshops, hackweeks are intensive and interactive, facilitated by 3 core components: tutorials on state-of-the-art methodology, peer-learning, and on-site project work in a collaborative environment. This setup is particularly powerful for sciences that require not only domain-specific knowledge, but also effective computational tools and methodologies. In response to the pandemic, eScience convened organizers of 10 hackweek-style events (including our own Neurohackademy, Oceanhackweek, Waterhackweek, and Electrochemical Society HackWeek) to tackle the challenge of making these events virtual. Through support from the Sloan Foundation, we have developed best practices for the engagement of students, the creation of collaborative environments, and the facilitation of interactive programs. This expertise will be incorporated into a dedicated section of the [hack-week toolkit](#).

Extending Data Science Beyond STEM: the Data Science Minor

In winter 2019, the provost at the University of Washington convened a Data Science Task Force focused on undergraduate data science education for non-STEM majors, co-chaired by eScience leadership. The task force developed a new Data Science Minor, approved earlier this year and officially rolling out this fall supported by a new cohort of faculty. eScience serves as the organizational home for this minor, in recognition of our ongoing leadership in this space. As part of the initiative, eScience created a framework for department-specific data science options. Currently, 16 departments have adopted data science options at the graduate level and 8 departments at the undergraduate level (increasing at a rate of 3-4 per year).

Career Paths: an Emerging Need for Research Software Engineers

In the last decade, we have learned that data science researchers often have related yet distinct needs for software development. To understand these needs and how eScience might support them, we reached out to PIs across the three UW campuses this summer with a survey based on one distributed at NYU in 2019 (courtesy of Kyle Cranmer). We received 339 responses, with 41% coming from the School of Medicine alone. Half of respondents noted the difficulty of finding skilled software engineers and/or fully supporting such positions through grants. Approximately 60% of respondents expressed an openness to contracting for these services. We found that >20 software engineering FTEs would need to be hired to meet the current funded demand and a further 20 FTEs would be needed to meet all needs across campus. We anticipate using these data to explore building out a software engineering team within eScience in tight collaboration with partner units on campus.

Building Research Collaborations: Data Science, Algorithms, and the Cloud

As recently [announced](#), our partners at the Algorithmic Data Science Institute received a \$12.5M grant from NSF through the [Transdisciplinary Research in Principles of Data Science](#) (TRIPODS) program to create a new **Institute for Foundations of Data Science** (IFDS), a collaboration between the UW, University of Wisconsin-Madison, UC Santa Cruz and University of Chicago. This interdisciplinary research institute will focus on developing data science foundations across math, computer sciences, statistics and engineering. The effort will be led by Senior Data Science Fellow [Maryam Fazel](#) and includes Data Science Fellow [Zaid Harchaoui](#).

The eScience Institute along with UC San Diego, UC Berkeley, and Strategic Blue entered production operations of the \$5M NSF [CloudBank](#) program, which aims to develop a suite of managed services to simplify the use of public clouds across computer science research and education. Over the past year, the CloudBank team worked with cloud providers, early users, project advisors, and the computer science community to develop processes, tools, and materials to address the plethora of pain points in making effective use of public clouds.

The eScience Institute is home to a \$2M NSF Harnessing the Data Revolution Engineering Data Science Institute (EDSI) led by former eScience Director [Magda Balazinska](#). EDSI is extending the reach of the eScience Institute by developing new methods to accelerate the design and manufacturing life cycle with data science while also creating an Engineering Data Science Environment to support researchers through a synergistic set of collaboration and education activities closely integrated with current eScience offerings.