

The [University of Washington eScience Institute](#), one of the nation's first data science institutes, grew out of the Moore-Sloan Data Science Environment effort which was focused on identifying and tackling impediments to the broad and sustainable adoption of data-intensive discovery. With a mission to empower researchers and students to address fundamental questions through the use of large, complex and noisy data, key eScience personnel now hold leadership roles in the construction, commissioning and operations of federally-funded research initiatives such as Pangeo, the Rubin Observatory, and Interactive Oceans as well as NSF-funded programs including CloudBank and the West Big Data Innovation Hub. With expertise in advanced statistical and computational techniques including artificial intelligence, machine learning, database management, visualization, and research software engineering, eScience has developed a suite of education, research, and community building programs. 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 Incubators

The annual [Data Science Incubator](#) program connects researchers from across campus with eScience data scientists for 10 weeks of close collaboration. Over the past 8 years, we have supported 50 projects with collaborators from 29 different departments. Previous incubator projects have led to long-term collaborations, publications, and grant awards. The annual **UW Data Science for Social Good (DSSG) program** was entirely remote again in Summer 2021, with 2 project teams and 8 student fellows (selected from over 120 applicants). This year's program supported [timely and relevant projects](#): 1) developing ensemble methods for evaluating redistricting plans and 2) examining the Seattle \$15 minimum wage ordinance as a solution to income inequality. To date, the DSSG program has provided training for over 80 student fellows and supported 19 research projects proposed by faculty, local government, non-profits, and industry. In collaboration with the [West Big Data Innovation Hub](#), we continue to leverage lessons learned from more than 10 university "data for good" programs to create a new guide for other universities interested in building similar programs that will be available later this year.

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. The eScience Institute has extensive experience (check out [this toolkit](#)) developing and facilitating hackweeks focused in particular domains, e.g. [Neurohackademy](#) and the Electrochemical Society HackWeek, as well as around particular datasets, e.g. the [ICESat-2](#) and [SnowEx](#) Hackweeks.

Extending Data Science Beyond STEM: the Data Science Minor

UW launched a new [Data Science Minor](#) in fall 2020 that is hosted out of the eScience Institute. Students in the Data Science Minor gain literacy and fluency in data science methods and their implications for society and the world. Graduates of the Minor are ready to work in emergent “translator” roles on heterogeneous teams composed of data scientists and domain experts. In the 9 months since launch, over 250 students have enrolled in the minor. Two new faculty have been hired in Geography and Digital Humanities, respectively, to support the development of curricula for the Minor, with more hires to come next year. In a related initiative, eScience created a curriculum 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.

Building Research Collaborations: Dynamic Systems, Astronomy, and the Cloud

The UW is leading a new artificial-intelligence research institute, which is part of an NSF investment that supports 11 new AI Institutes with \$20 million each over the next 5 years. The [AI Institute in Dynamic Systems](#) will focus on fundamental AI and machine learning theory, algorithms and applications for real-time learning and control of complex dynamic systems, which describe chaotic situations where conditions are constantly shifting and hard to predict. Partner institutions include the University of Hawaii at Mānoa, Montana State University, the University of Nevada Reno, Boise State University, the University of Alaska Anchorage and Portland State University, Harvard University and Columbia University.

The [Legacy Survey of Space and Time \(LSST\)](#), which will be carried out by the Vera C. Rubin Observatory, is the flagship ground-based astronomical survey of the 2020s. With an estimated start date of October 2023, LSST will generate the deepest-ever, multi-color, 10-year-long movie of the southern sky, detecting 30 billion stars and galaxies and amassing 100 PB of imaging and catalog data. The Schmidt Futures Foundation recently announced \$15M of support over 5 years for LINCC, a collaboration between UW, Carnegie Mellon, University of Arizona, Northwestern University, and LSST Corporation. LINCC will deliver key algorithms and code needed to extract knowledge from the data, frameworks to enable petabyte-scale analyses, mechanisms to search for one-in-a-million events in continuous streams of data, and community organization mechanisms and communication channels that enable researchers to develop and share their algorithms and software. The UW part of this award will be run through eScience and the [DiRAC Institute](#) and will build out a new team of data scientists and research software engineers.

The eScience Institute along with UC San Diego, UC Berkeley, and Strategic Blue continue production operations of the \$5M NSF [CloudBank](#) program, which has developed a suite of managed services to simplify the use of the public cloud for university researchers. Over the past 2 years, the CloudBank team has 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.