# Today

- Introduction to the eScience Institute

- Overview of the Incubator Program

- Example Projects

- Your questions….

Our Mission

The eScience Institute **empowers** researchers and students in all fields to answer fundamental questions through the use of large, complex, and/or noisy data.

As the **hub** of data-intensive discovery on campus, we lead a **community** of innovators in the techniques, technologies, and best practices of data science and the fields that depend on them.
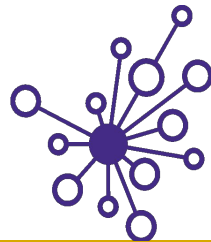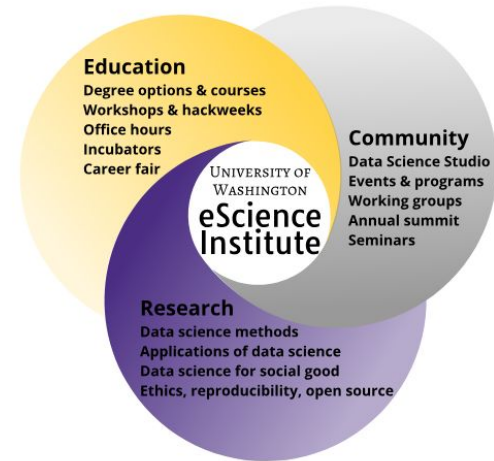
UNIVERSITY *of* WASHINGTON
eScience Institute
ADVANCING DATA-INTENSIVE DISCOVERY IN ALL FIELDS

eScience Institute Role at the U of Washington and Beyond

- **Education**
  - Disseminate Data Science Expertise and Best Practices
  - Lead Data Science Education at UW
- **Research**
  - Advance the State of the Art in Data Science
  - Use Data Science for Social Good
- **Community**
  - Hub of data science community
  - Partnerships

Education
Degree options & courses
Workshops & hackweeks
Office hours
Incubators
Career fair

UNIVERSITY OF
WASHINGTON
eScience
Institute

Community
Data Science Studio
Events & programs
Working groups
Annual summit
Seminars

Research
Data science methods
Applications of data science
Data science for social good
Ethics, reproducibility, open source

eScience Research Team

Director of Research

David Beck
Ph.D. Medicinal
  Chemistry,
Biomolecular Struct.
  & Design

Data Scientists
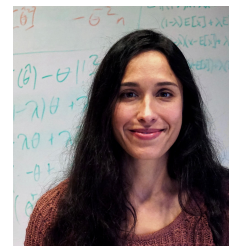
Noah Benson
Ph.D. Biomedical
& Health Informatics

Bernease Herman
B.S. Statistics
Formerly SE at Amazon

Jose Hernandez
Ph.D. Measurement
& Statistics

Valentina Staneva
Ph.D. Applied
Mathematics and
Statistics

Anissa Tanweer
Ph.D. Communication

Research Scientists

Anthony Arendt
Ph.D. Geophysics
APL

Nicoleta Cristea
Ph.D. Environmental
Engineering

Bryna Hazelton
Ph.D. Astrophysics
Physics

Joe Hellerstein
Ph.D. Computer Science
IBM Research, Microsoft
  Research, Google (ret.)

Scott Henderson
Ph.D. Geological
Sciences

Vaughn Iverson
Ph.D. Oceanography

Spencer Wood
Ph.D. Zoology

Data Science Incubator Program - *Scalable Research Impact*

Move from "accidental" encounters to engineered quarter-long+ partnerships

Identify emerging opportunities around campus

Provide a shared environment where researchers can learn from our team *and each other*

49 projects over the past 7 years, 28 departments represented

5-7 projects per year

UNIVERSITY *of* WASHINGTON
eScience Institute
DATA SCIENCE INCUBATOR

Each team will consist of…

- 1 Project Lead
  - domain expertise, overall project responsibility
  - faculty, postdoc, graduate student, research staff
  - from any discipline or department/unit
- 1 Data Science Lead
  - technical expertise, technology support
- (in some cases) Stakeholder(s)
  - faculty advisor, industry sponsor, government subject matter expert, etc.

Program Logistics

Project proposals due online: by 11:59 p.m. PT on Nov. 10th

Notification: Dec. 10th

Kickoff meeting: Tuesday Jan. 4th

Project Leads and Data Science Leads commit to ~16 hours/week

– **Flexible scheduling in coordination with data scientist lead**

– **Weekly full group meeting**

– **Other weekly events TBD: tutorials, social activities**

– **Final Presentation**

Program will be primarily remote via Zoom & Slack, but there may be some opportunities for in-person work

Areas of Interest

- No Limits!

- Strong local expertise in Astronomy, Neuroscience, Applied Math, Physics, Earth Science, Finance, Education, Ecology

- Strong expertise in image processing and analytics, machine learning, Python, cloud computing, and big data systems

- Extracting knowledge from large, heterogeneous, and/or noisy datasets

NiPy

Cloud Computing

In collaboration with UW research computing, this year we are excited to be able to offer cloud computing support for the incubator projects.

Using cloud computing is not a requirement for incubator projects but it can help accelerate some projects.

Project desirables

- strong research, strong methods
- new directions, new questions
- availability, engagement, commitment
- "do only what we can do together"
- clarity and shovel-readiness
- capacity for measurable outcomes and sustained engagement

We expect you to…

- Summarize findings in a brief report on our website
- Publicly present findings in a final presentation

- Host all code on github

- Include eScience data scientist(s) as authors on all papers resulting from the project. Acknowledge eScience Incubator program on all presentations, publications, etc. related to this work.

- Stay in touch!!

Proposals

https://escience.washington.edu/get-involved/incubator-programs/overview/

- Lightweight! Short answer questions and 1-2 page project summary

- Questions focus on data details, what expertise you think would be helpful and the desired outcomes.

- Primarily remote, but we ask about interest in in-person work. This will NOT affect project selection.

- **** Please use our office hours to help decide if the program is a good fit and to help scope projects! ****

Data and Research Scientist Office hours

https://escience.washington.edu/office-hours/#eScienceDataScientists

Incubator reflection

"**The program is state-of-the-art with respect to mentoring, programming instruction, and machine learning didactic training**...The combination of these regular one-on-one meetings and a weekly group meeting with other incubator mentees and mentors provided a rich environment where I could ask a programming or data science question and obtain an answer immediately...**The UW eScience incubator has my full endorsement** and I would highly recommend it to anyone who is starting out in programming or anyone who has intermediate experience in programming and data science."

- Charles Zhou, Anesthesiology & Pain Medicine, Winter Incubator 2020

Past Projects

http://escience.washington.edu/get-involved/incubator-programs/

Example projects from Winter 2021 Incubator

- Using Gliders to Observe Submesoscale Flows **– Oceanography**

- The Universe of International Treaties **– Political Science**

- Dry Thunderstorm Forecast Using Machine Learning Techniques **– Atmospheric Sciences**

- Learning to See the Forest and the Trees: Using Computer Vision to Make Forest Stewardship More Accessible **– Environmental and Forest Sciences**

- Climate Adaptation for Future Maize – Novel Plant Traits and New Management **–Biology**

- Detecting Wildflowers in Spectral Imagery **– Biology**

- Using Social Media to Model Backcountry Use in Rainier National Park **– Environmental and Forest Sciences**

Example projects from Winter 2020 Incubator

- Deer Fear: Accelerometers, Video Collars, and GPS to Explore Deer Reactions to Wolves **– Environmental and Forest Sciences**

- British Justifications for Internment without Trial: NLP Approaches to Analyzing Government Archives **- Political Science**

- Automated monitoring and analysis of slow earthquake activity **– Earth & Space Sciences**

- Data Analytics for Decoding and Demixing Patterns of Population Neural Activity Underlying Addiction Behavior **– Anesthesiology & Pain Medicine**

- Systems level analysis of metabolic pathways across a marine oxygen deficient zone **– Oceanography**

- Predicting a drought with a flood of data: Evaluating the utility of data-driven approaches to seasonal hydrologic forecasts **– Civil & Environmental Engineering**

This program brings together students and researchers with data science and domain expertise to work on focused, collaborative projects for societal benefit.

Schedule

- 11/10: Project proposals due

- 12/10: Notification

- 1/4: Kickoff meeting

- Questions?
  - sstone3@uw.edu
  - brynah@uw.edu
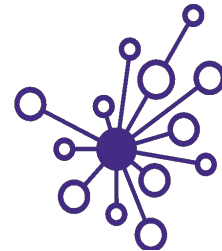
UNIVERSITY of WASHINGTON

eScience Institute

ADVANCING DATA-INTENSIVE DISCOVERY IN ALL FIELDS

GORDON AND BETTY MOORE FOUNDATION

ALFRED P. SLOAN FOUNDATION

Washington Research FOUNDATION

W UNIVERSITY of WASHINGTON