# 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.
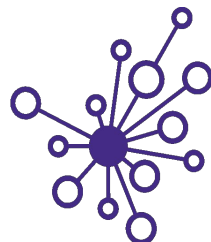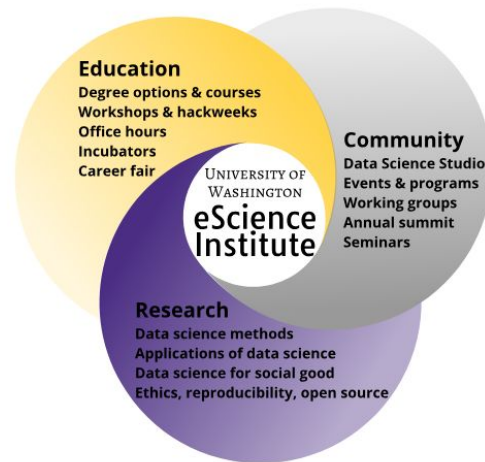
UNIVERSITY *of* WASHINGTON
eScience Institute
ADVANCING DATA-INTENSIVE DISCOVERY IN ALL FIELDS

# 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



Amanda Tan
Ph.D. Civil & Env.
Engineering



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 an in-house team, external mentors, *and each other*

**43 projects over the past 7 years, 28 departments represented**

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
- 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. 15$^{th}$

Notification: Dec. 9th

Kickoff meeting: Tuesday Jan. 5th

**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 in the UW Data Science Seminar**

Program will be entirely remote via Zoom & Slack

# 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 UW Data Science Seminar

- 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

- 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.

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

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

## 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

4-6 concurrent projects selected for each Incubator

43 projects over the past 7 years, 25+ departments represented

Participation from faculty, postdocs, graduate students, staff

*What we're looking for:*  **Projects where fruitful collaboration is possible, with potential for significant impact.  "Do only what we can only do together"**

# 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**

# Winter 2019 Incubator

- Atmospheric particulate matter source identification using excitation emission fluorescence spectroscopy – Chemical Engineering
- Beneficial competition under rationing: evidence from food delivery service – Economics
- A network analysis of tree competition: Which tree species make the best neighbors? – Biology
- Predicting human-mediated vectors for invasive species from mobile technology – Aquatic & Fisheries Sciences
- Affective state analysis of ultrasonic vocalizations in animal models of mTBI/PTSD and neuropathic pain – Psychology & Behavioral Sciences
- Interactions of tropical precipitation with atmospheric circulation and energy transport - Oceanography

# Winter 2018 Incubator

- Experimental diffusion analysis to extract tissue structure function in the diseased brain – Chemical Engineering

- Deciphering climate clues via carbon flux simulation – Atmospheric Sciences

- Incubating a DREAM – Civil and Environmental Engineering

- Political Twitter images project summary and goals – Political Science

- Hitting the mark: targeting strategy development for SDSS V with a robotic fiber positioning system – Physics & Astronomy

# Winter 2017 Incubator

- 3D Visualization of Prostate Cancer Using Light-Sheet Microscopy – Pathology
- Detecting Small Particles in Low-Contrast Images to Aid in Particle Tracking – Mechanical Engineering
- National Weather Watch: Monitoring Freshwater Vulnerability to Climate Change and Human Activity – Forestry & Environmental Science
- Cloud-Enabled Tools for the Analysis of Subsea HD Camera Data – Applied Physics Laboratory
- Discovering Marine Trophic Interactions Using Sonar Time Series from Ocean Observatories – Applied Physics Laboratory
- Applying Machine Learning to the Analysis of the Large-Scale Structure of Turbulence – Aeronautics & Astronautics

# Winter 2016 Incubator

- Developing a Workflow to manage Large Hydrologic Spatial Datasets *(Nicoleta Cristea, Environmental Engineering)*

- Methods for Characterizing Human Centromeres *(Seva Kasinathan, School of Medicine)*

- Target Detection for Advanced Environmental Monitoring of Marine Renewable *Energy (Emma Cotter, Mechanical Engineering)*

- Improved Stimulation for Sight Restoration Technologies *(Ione Fine, Psychology)*

- AralDIF: A Cloud-based Dynamic Information Framework for the Aral Sea Basin *(Amanda Tan, Oceanography)*

- Damage Speaks: an improved acoustical monitoring framework for structures subjected to earthquakes *(Travis Thonstad, Civil Engineering)*

UNIVERSITY *of* WASHINGTON
# eScience Institute
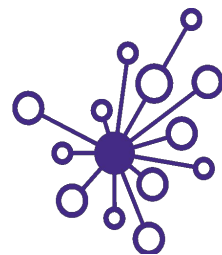## DATA SCIENCE FOR SOCIAL GOOD

2017 DSSG Student Fellows

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

# Schedule

- 11/15: Project proposals due

- 12/9: Notification

- 1/5: Kickoff meeting


- Questions?
  - sstone3@uw.edu
  - brynah@uw.edu