UNIVERSITY of WASHINGTON

“Intellectual Infrastructure” at the UW MSDSE

Bill Howe

UNIVERSITY of WASHINGTON

2005-2008

"All across our campus, the process of discovery will increasingly rely on researchers' ability to extract knowledge from vast amounts of data… In order to remain at the forefront, UW must be a leader in advancing these techniques and technologies, and in making [them] accessible to researchers in the broadest imaginable range of fields."

UNIVERSITY of WASHINGTON
eScience Institute

In other words:

- Data-intensive science will be ubiquitous
- It's about **intellectual infrastructure**

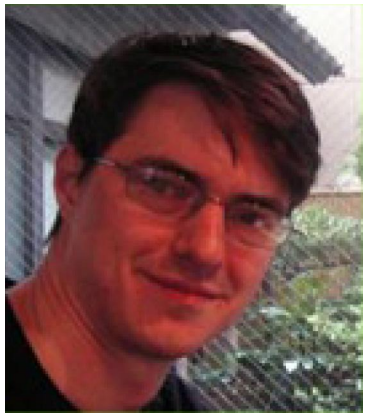
http://escience.washington.edu

eScience Research Consulting circa 2010…

- Technical staff

David Beck

Jeff Gardner



Bill Howe

Erik Lundberg

Chance Reschke

UNIVERSITY of WASHINGTON

eScience Research Consulting 2016

Data Scientists (fully supported)

Jake VanderPlas
Director of Research, Physical Sciences
Ph.D., Astronomy

Ariel Rokem
Data Scientist
Ph.D., Neuroscience

Valentina Staneva
Data Scientist
Ph.D., Applied Mathematics and Statistics

Bernease Herman
Data Scientist
BS, Stats
was SE at Amazon

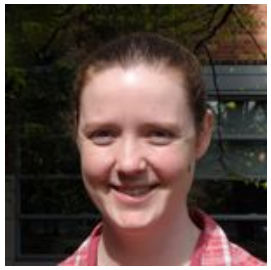
Research Faculty

Dave Beck
Director of Research, Life Sciences
Ph.D. Medicinal Chemistry, Biomolecular Structure & Design

Research IT

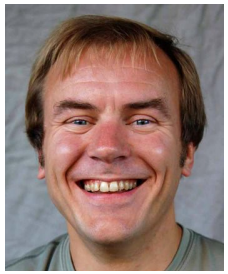
Rob Fatland
Director of Cloud and Data Solutions
Senior Data Science Fellow
PhD Geophysics

Research Scientists (partial support)



Bryna Hazelton
Research Scientist
Ph.D., Physics

Andrew Gartland
Research Scientist
Ph.D., Biostatistics



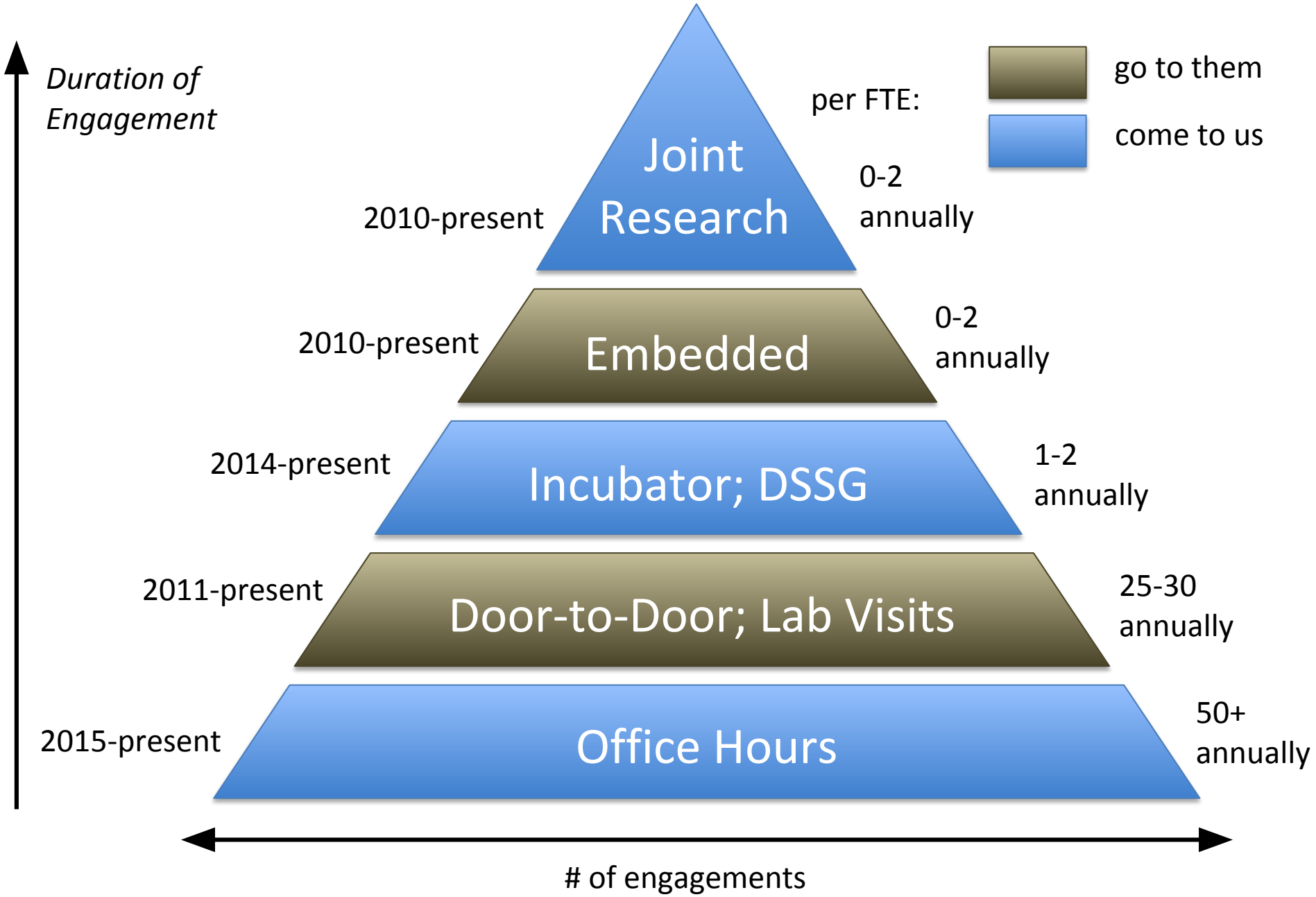
Vaughn Iverson
Research Scientist
Ph.D., Oceanography

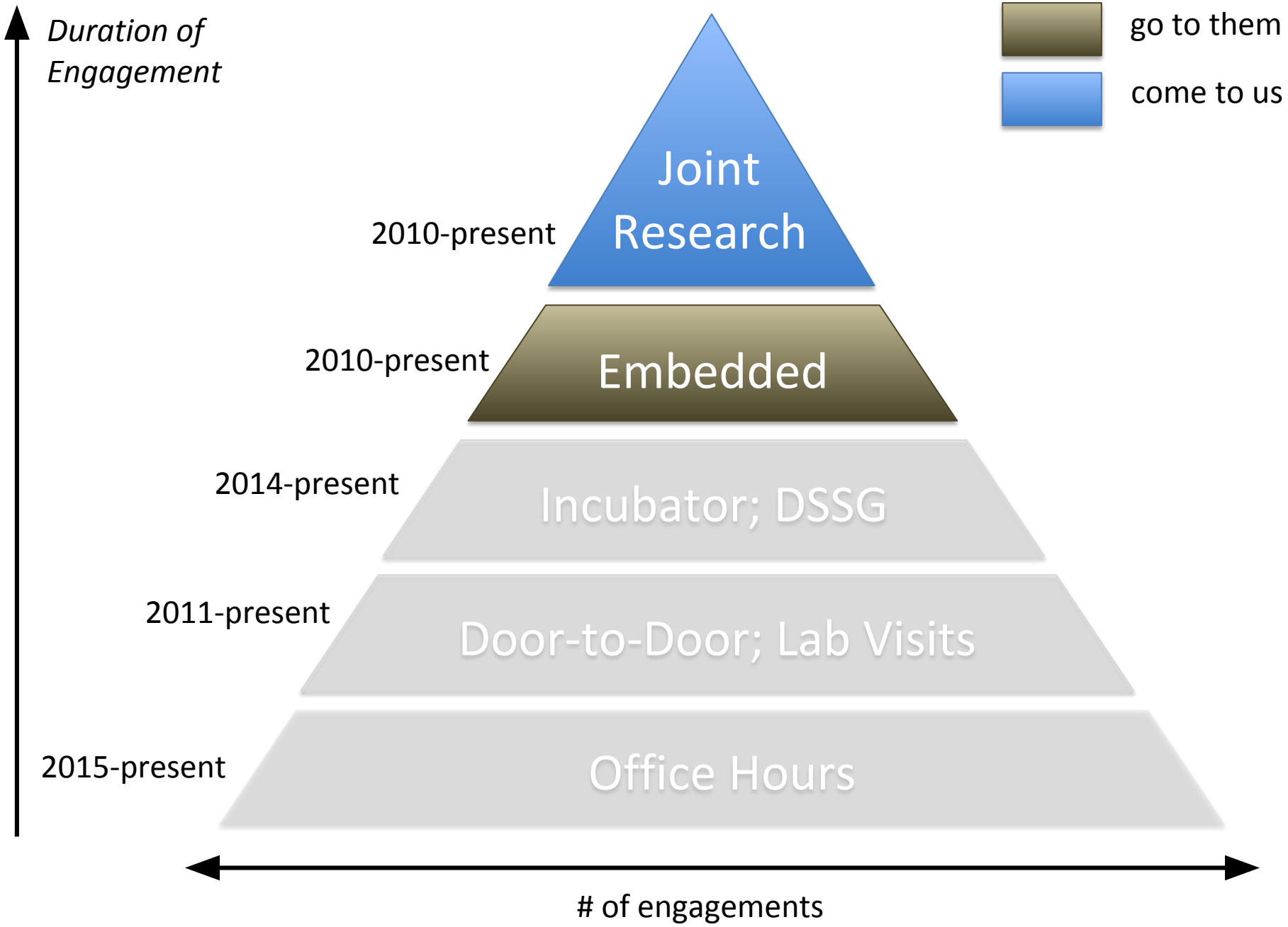
Joe Hellerstein
Senior Data Science Fellow
IBM Research, Microsoft Research, Google (ret.)

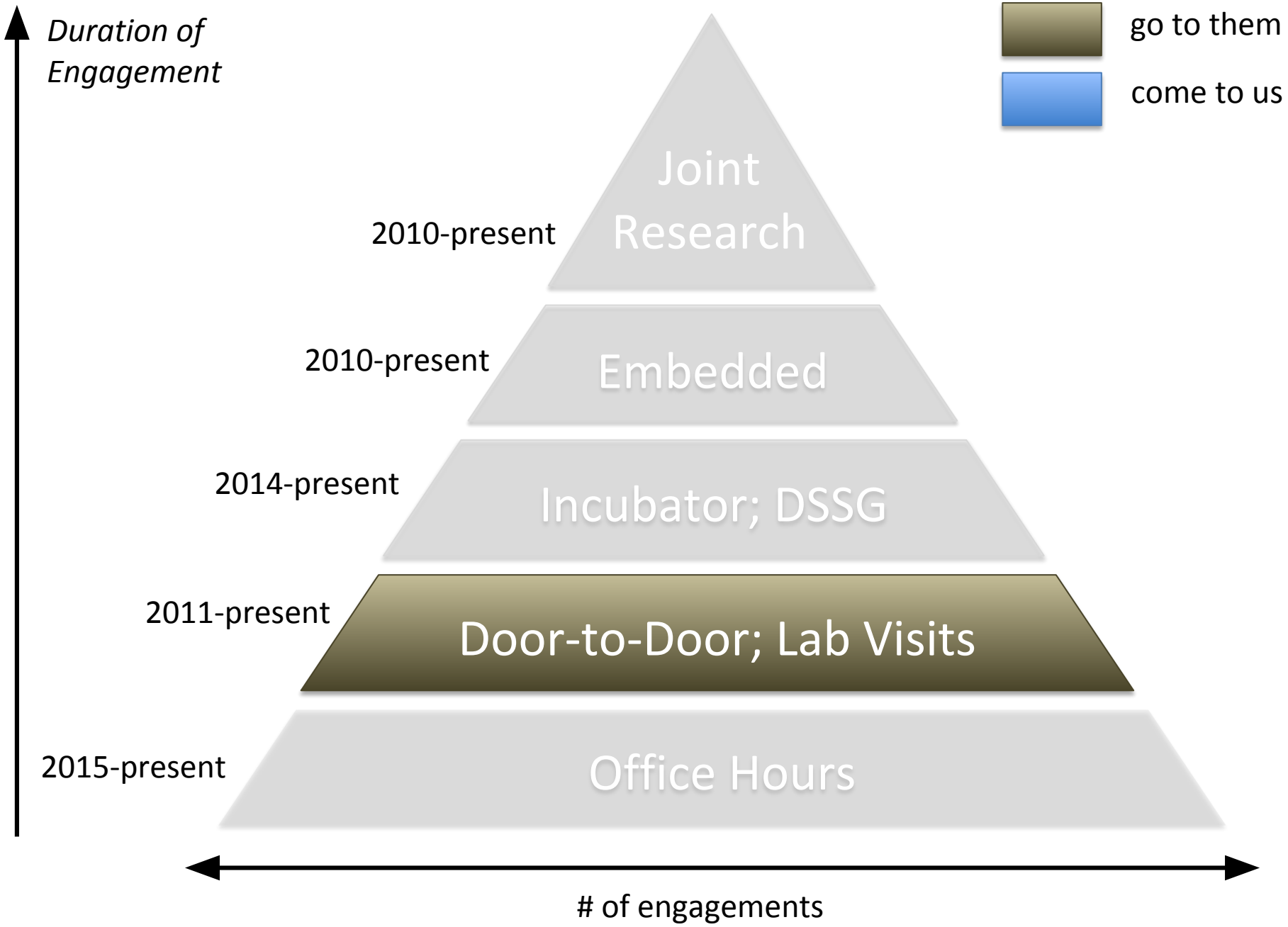
Dir. Ethnography

Brittany Fiore-Gartland
Ethnographer
Ph.D Communication

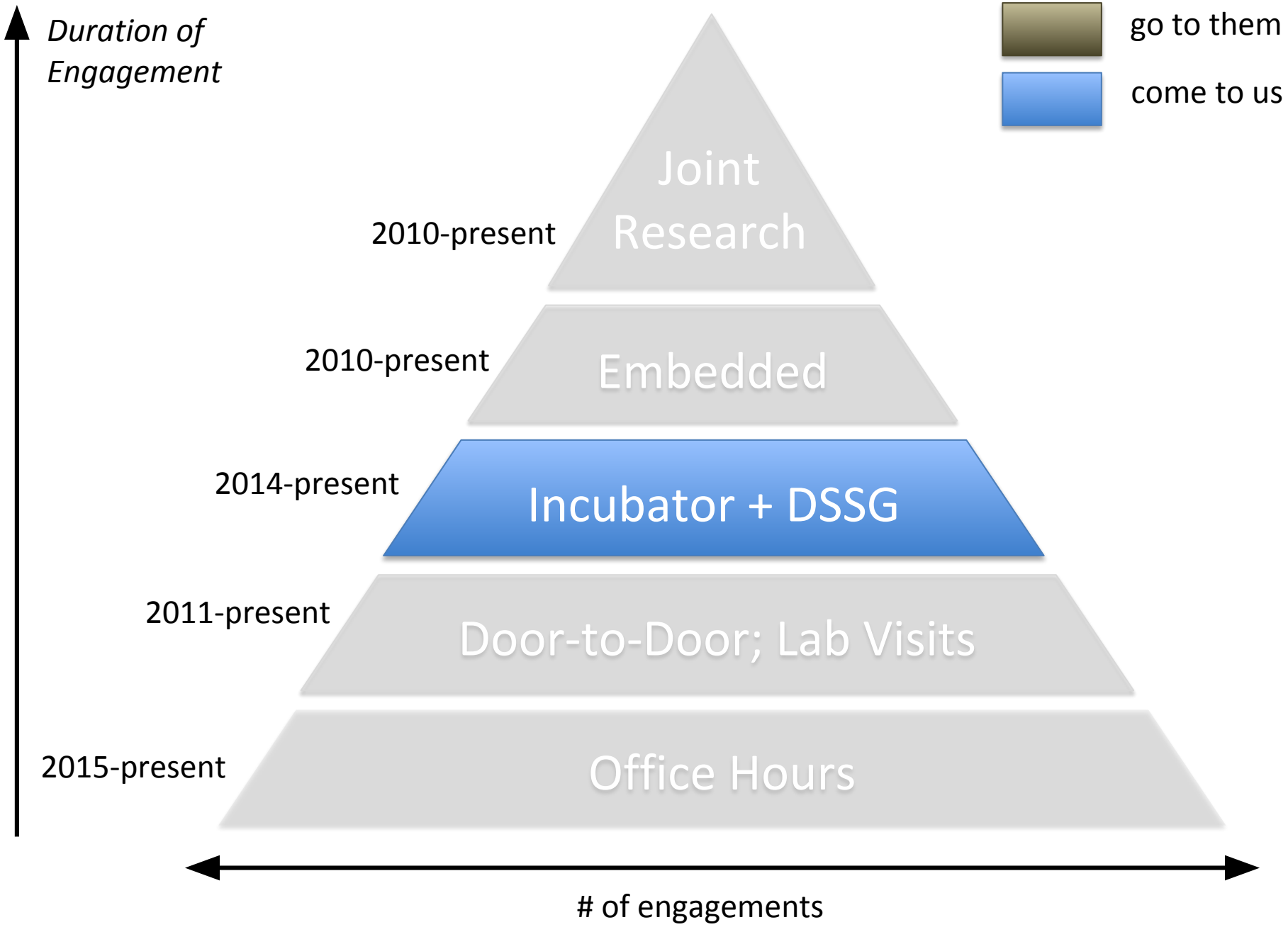
“Intellectual Infrastructure”




Duration of Engagement
go to them
come to us
Joint Research
2010-present
Embedded
2010-present
Incubator; DSSG
2014-present
Door-to-Door; Lab Visits
2011-present
Office Hours
2015-present
# of engagements


Duration of Engagement
go to them
come to us
Joint Research
2010-present
Embedded
2010-present
Incubator; DSSG
2014-present
Door-to-Door; Lab Visits
2011-present
Office Hours
2015-present
# of engagements


Data Science Kickoff Session:
137 posters from 30+ departments and units

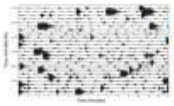

Duration of Engagement
go to them
come to us
Joint Research
2010-present
Embedded
2010-present
Incubator + DSSG
2014-present
Door-to-Door; Lab Visits
2011-present
Office Hours
2015-present
# of engagements

Incubation Program Overview

- Quarter-long, on-site projects, engagement two days per week
  - Simple two-page proposals
  - 4-6 concurrent teams: Network effects among cohort beyond 1:1 interactions
  - Each team is ~50% project lead + ~50% eScience FTE

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

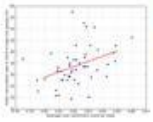
Spring 2014 Incubation Projects



Automated Detection and Analysis of Repeating Earthquakes

Alicia Hotovec-Ellis, Kate Allstadt, Jon Connolly, and John Vidale — Earth and Space Sciences

eScience Contact: Jake Vanderplas



Using social media data to identify geographic clustering of anti-vaccination sentiments

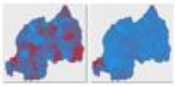
Benjamin Brooks, Abraham Flaxman — Institute for Health Metrics and Evaluation

eScience Contact: Andrew Whitaker

Analysis of Kenya's Routine Health Information System data

Gregoire Lurton, Abraham Flaxman, Emmanuela Gakidou — Institute for Health Metrics and Evaluation

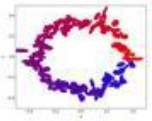
eScience Contact: Dan Halperin



Efficient Computation on Large Spatiotemporal Network Data

Ian Kelley, Josh Blumenstock — Information School

eScience Contact: Andrew Whitaker



Scalable Manifold Learning for Large Astronomical Survey Data

Marina Meilă — Statistics

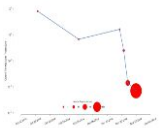
eScience Contact: Jake Vanderplas

ASPASIA: Adult Service Providers and Some Incidental Addenda

Sam Henly — Economics

eScience Contact: Andrew Whitaker

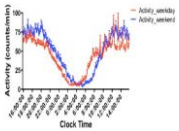
Fall 2014 Incubation Projects



Kernel-Based Moving Object Detection

Andrew Becker (Astronomy)

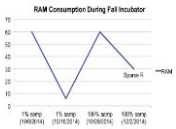
eScience Contact: Daniel Halperin



Students' sleep and academic performance

Ângela Katsuyama (Biology)

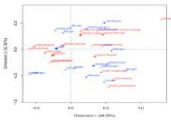
eScience Contacts: Bill Howe and Daniel Halperin



Simulating Competition in the U.S. Airline Industry

Carlos A. Manzanares (Economics)

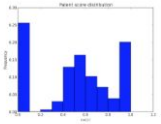
eScience Contact: Andrew Whitaker



Analysis of .gov web archive data

Emily Gade (Political science), Andrew Whitaker (eScience)

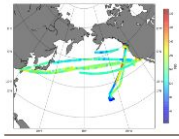
eScience Contact: Andrew Whitaker



Innovation: Evidence from Patents

Matthew Denes (Finance and Business Economics)

eScience Contact: Andrew Whitaker



Analysis of large-scale patterns in phytoplankton diversity

Sophie Clayton (Oceanography)

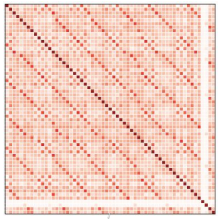
eScience Contact: Daniel Halperin

Winter 2016 Incubation Projects

Developing a Workflow for Managing Large Hydrologic Spatial Datasets to Assist Water Resources Management and Research

Lead: Nicoleta Cristea, **Civil and Environmental Engineering**

eScience Liaisons: Anthony Arendt, Rob Fatland



Methods for Characterizing Human Centromeres

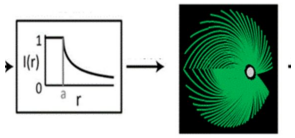
Project Lead: Siva Kasinathan, **UW School of Medicine**

eScience Liaison: Andrew Fiore-Gartland, Bryna Hazelton

Target Detection for Advanced Environmental Monitoring of Marine Renewable Energy

Project Lead: Emma Cotter, **Mechanical Engineering**

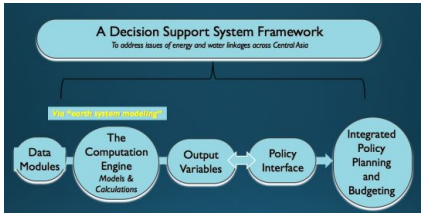
eScience Liaison: Bernease Herman



Improved Stimulation Protocols for Sight Restoration Technologies

Project Leads: Ione Fine, Geoffrey M. Boynton, UW **Psychology**

eScience Liaison: Ariel Rokem



AralDIF: A Cloud-based Dynamic Information Framework for the Aral Sea Basin

Project Lead: Amanda Tan, Department of **Oceanography**

eScience Liaisons: Rob Fatland, Anthony Arendt

Damage Speaks: Acoustical Monitoring Framework for Structures Subjected to Earthquakes

Lead: Travis Thonstad, **Civil & Environmental Engineering**

eScience Liaison: Valentina Staneva

Home About Us Research Education Get Involved

Data Science for Social Good

UNIVERSITY of WASHINGTON eScience Institute
DATA SCIENCE FOR SOCIAL GOOD

Inaugural 2015 program:
16 spots
140 applicants
…from 20+ departments

Are you interested in using data-driven discovery for societal benefit?

Apply for the 2016 eScience Institute Data Science for Social Good summer program!

The University of Washington eScience Institute, in collaboration with **Urban@UW** and Microsoft, is excited to announce the 2016 Data Science for Social Good (DSSG) summer program. The program brings together data and domain scientists to work on focused, collaborative projects that are designed to impact pub

Modeled after similar programs at the U
sixteen DSSG Student Fellows will be sel
intensive research projects. These are p

Assessing Community Well-Being
Third-Place Technologies
Optimization of King County Metro Paratransit
Computer Science & Engineering
Predictors of Permanent Housing for Homeless Families
Bill and Melinda Gates Foundation
Open Sidewalk Graph for Accessible Trip Planning
Computer Science & Engineering

The Seattle Times

Education | Education Lab | Local News | Transportation

UW student project taps ORCA cards, unlocks data trove

Originally published August 19, 2016 at 10:21 pm | *Updated August 21, 2016 at 6:37 pm*

GeekWire NEWS JOBS EVENTS RESOURCES DEALS ABOUT Search

STRATOS MEDICAL DEVICE DEVELOPMENT IS RISKY. LET STRATOS' ENGINEERS CLEAR THE PATH FOR PRODUCT DEVELOPMENT. learn more >>

Trending: Hewlett Packard Enterprise cuts staff as it struggles to keep pace in the cloud

Newsletter signup Space & Science

Could Amazon reviews keep you from getting sick? Researchers analyze text to predict food recalls

BY CLARE MCGRANE on August 28, 2016 at 11:16 am

Xconomy Xperience Tech + Life EXOME Biotech + Health Our Regions Tech Channels Meet the Xconomists Our Events

ADVERTISEMENT

Seattle Home Seattle Events Local Jobs Archives Xconomists VC / M&A Deals

Budding UW Data Scientists Use Their Powers for Social Good

Benjamin Romano
August 24th, 2015

@bromano @xconomy Like Us

Earn a degree in the field of data science these days and your ticket is punched: Google, Amazon, Facebook, leading-edge academic research, a well-funded startup

GeekWire NEWS JOBS EVENTS RESOURCES DEALS ABOUT

Trending: Microsoft reveals the 'Xbox Onesie' and the internet goes nuts

Could data help solve Seattle's transportation challenges?

BY CLARE MCGRANE on August 20, 2016 at 3:30 pm

Student projects leapfrog governments and industry in 'Data Science for Social Good' program

Posted Aug 26, 2016 by Devin Coldewey, Contributor

Next Story

ADVERTISEMENT

Enter your forecasts for a chance to win prizes totaling $1.2 million.

megaman: Manifold Learning with Millions of points

James McQueen JMCQ@UW.EDU
Department of Statistics
University of Washington
Seattle, WA 98195-4322, USA

Marina Meilă MMP@STAT.WASHINGTON.EDU
Department of Statistics
University of Washington
Seattle, WA 98195-4322, USA

Jacob VanderPlas JAKEVDP@UW.EDU
e-Science Institute
University of Washington
Seattle, WA 98195-4322, USA

Zhongyue Zhang ZHANGZ6@CS.WASHINGTON.EDU
Department of Computer Science and Engineering
University of Washington
Seattle, WA 98195-4322, USA

Editor:

Abstract

Manifold Learning (ML) is a class of algorithms seeking a low-dimensional non-linear representation of high-dimensional data. Thus ML algorithms are, at least in theory, most applicable to high-dimensional data and sample sizes to enable accurate estimation of the manifold. Despite this, most existing manifold learning implementations are not particularly scalable. Here we present a Python package that implements a variety of manifold learning algorithms in a modular and scalable fashion, using fast approximate neighbors

Home

Edit New Page Search this wiki...

Alicia Hotovec-Ellis edited this page on Jan 11 · 20 revisions

Clone this wiki locally

https://github.com/ahotovec

Clone in Desktop

Motivation

Monitoring earthquakes is a challenge because signals from hundreds of seismograms record up to several hundred samples a second. Automatic event detection is required, yet important signals are too buried in noise and variable in character to identify with simple schemes. We aim to systematically mine the numerous events that are near-repetitions.

BAYLOR
UNIVERSITY

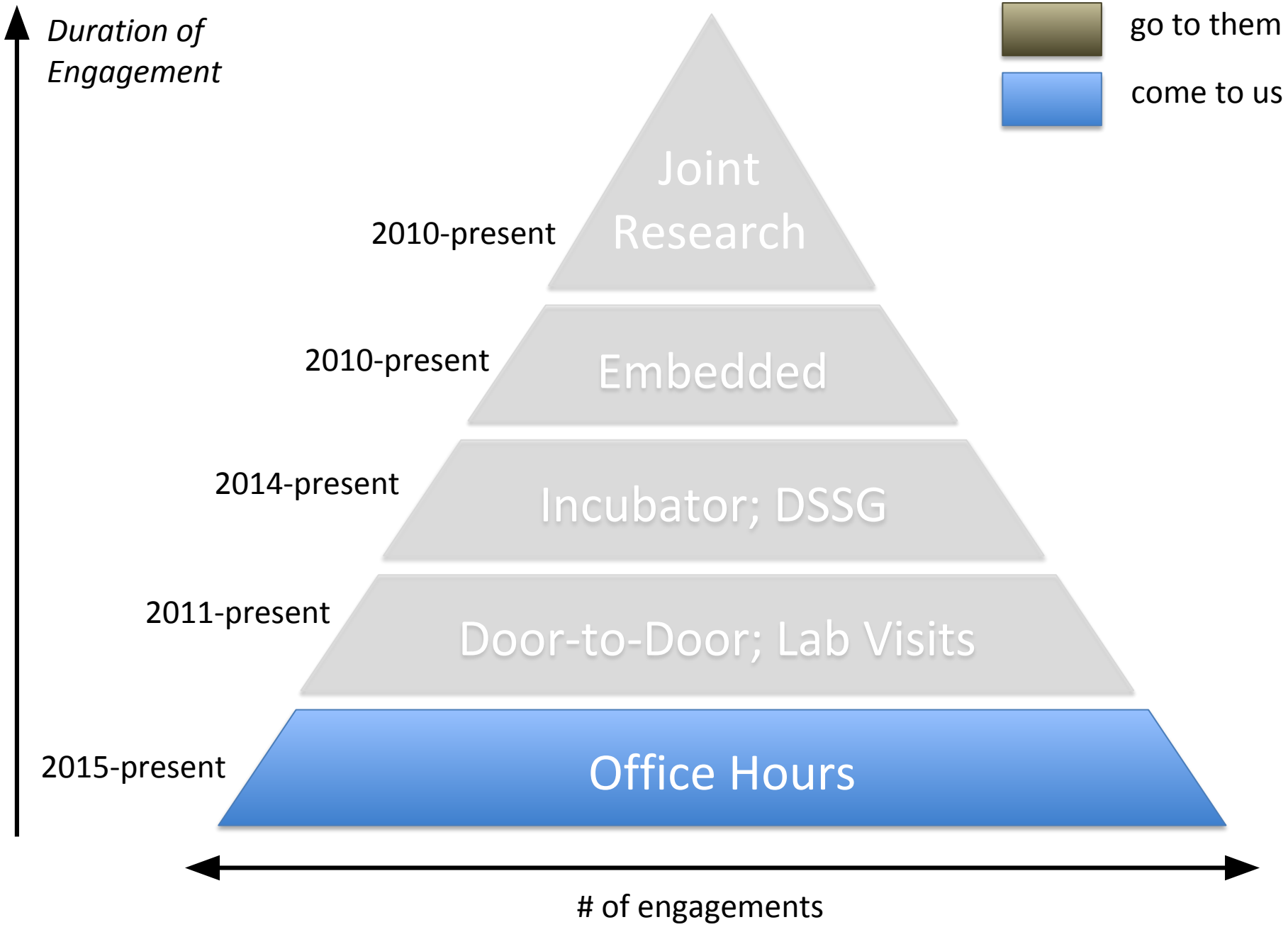
amazon


THE UNITED STATE OF WOMEN
THE WHITE HOUSE
WASHINGTON
THE UNITED STATE OF WOMEN
THE WHITE HOUSE
WASHINGTON
THE UNITED STATE OF WOMEN


STATE OF THE MAP US
CLASS OF 2020

BILL & MELINDA GATES foundation

MacArthur Foundation


Duration of Engagement
go to them
come to us
Joint Research
2010-present
Embedded
2010-present
Incubator; DSSG
2014-present
Door-to-Door; Lab Visits
2011-present
Office Hours
2015-present
# of engagements

Office Hours

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

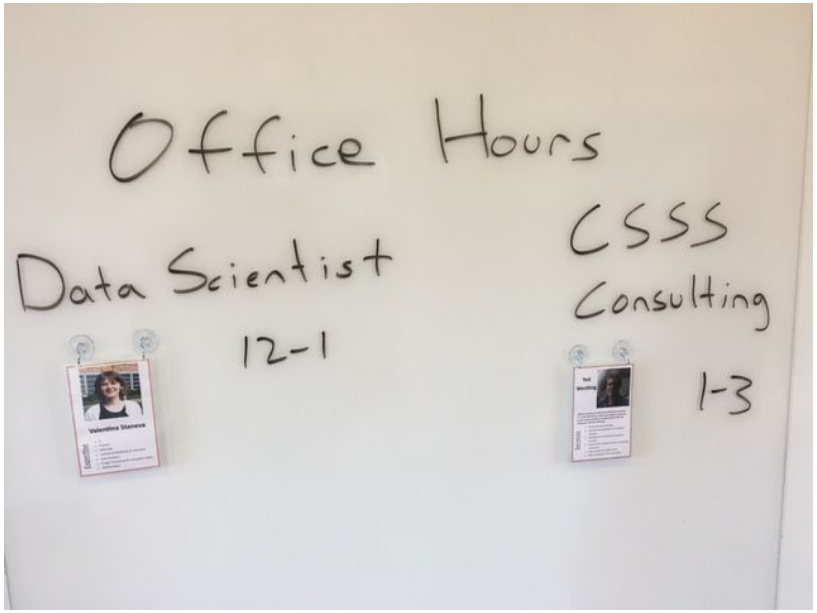
UW-IT

amazon web services

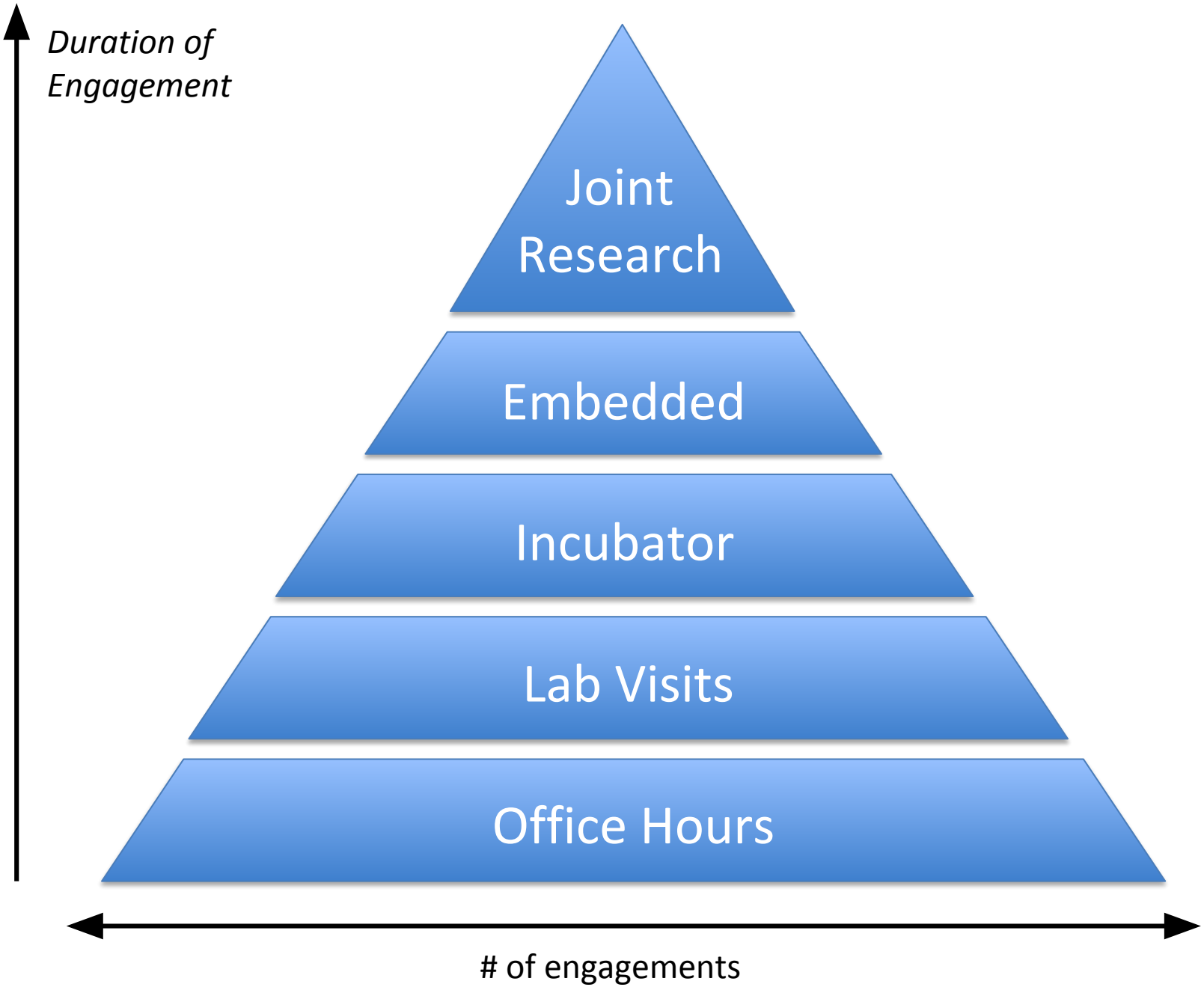
CENTER for STATISTICS and the SOCIAL SCIENCES

MathWorks

UNIVERSITY LIBRARIES



The MS DSEs have fundamentally transformed our ability to deliver the "Intellectual Infrastructure" to support data-intensive science


Duration of Engagement
Joint Research
Embedded
Incubator
Lab Visits
Office Hours
# of engagements