**Welcome Remarks** – *Jay Larsen*

* According to code.org, 70% jobs are from software space
* Passion is important

Idaho’s economy

* Process
* Materials research
* Energy research
* Computer science research
* Bio/ag sciences research

Natural Resource Economy – Nonfarm Jobs in Idaho

* Idaho before was resource-based jobs (farming)
* Tech expansion, 1973 and 1977 two branches of HP emerged, also Micron
* Idaho powered through recession

**Plenary Panel: This is Your Hub**

Eric Jankowski – Associate Professor

**Two Thoughts -** David E. Culler

* *Speaker talked about UC Berkley Data Science workshop*
* Data science for education
* Promoting research and enthusiasm for students as success factor for data science

**Community-driven Principles for Ethical Data Sharing (CPEDS)** - Natalie Evans Harris, BrightHive

* How to do data science responsibly and ethically?
* Responsible data use, gathering
* Transparency
* Foster diversity
* Guarantee security of data, subjects and algorthims
* Advance social initiatives on hackathons, etc

How do we make good?

* Strengthen community
* Building partnership
* Develop ethical framework

**California Safe Drinking Water Data Challenge** - Debbie Franco - Community and Rural Affairs Advisor – California Governor’s Office of Planning and Research

* Speaker not a data person or tech person, a water person
* Millions of water data points in California, limited effort for taking advantage of the data points
* Open data portal

**Expanding data science training opportunities across the western region** - Sarah Stone – UW eScience Director

* Carpentry – 2-day foundational training on data science and coding
* Data Carpentry Instructor Trainings (event) – train to be a data carpentry instructor
* Structured around volunteer instructor
* Carpentry & Data Awareness Event Pala Reservation (event) – topics: python, intro to data, git, visualization.
* Instructor: Ariel Rokem

*Eric Jankowski – question, what did you do with the Hub system?*

*Answer by David Culler: hub is a vehicle, the mission.*

*Ethics to human context. Not only “am I ethical?” it should be “is it good for the public interest”*

*“Safe drinking water as a human right”*

*Answer by Natallie Harris: the hub, can take an idea, and talk about how to apply it.*

*There are legal challenges*

*The human-side*

*How to apply to benefit for the day-to-day lives of the people*

*Answer by Debbie Franco: not familiar with the hub, but my colleagues, yes.*

*Session: information to action, was very informative. Meridith Lee helped.*

*While the focus was on data, Meridith was fine to do a broader conversation.*

*Set meta data standards*

*Importance of data access, instead of people researching, filtering for hours on data.*

*Answer by Sarah Stone: focus on regional community. Opportunity to collaborate with the other hubs and focus on the national level.*

*Question from audience: private sector on ethics?*

*Natalie: Private sector for ethics. Apple, Kaggle, Bloomberg.*

*David: research coming from the government and the industries. Workforce gap in data science is everywhere. Lacking of industrial partnership with schools. As university, get the breadth of students from different programs. Student-driven efforts.*

*Data science for social good*

*Question from audience: volunteer, software carpentry?*

*Hard to be a general data scientist*

**The State of the Hub Address**

*Meridith Lee*

**National Transportation Data Challenge – How Might We Reduce Traffic Fatalities?**

Info: <https://bigdatahubs.io>

Datathon

California Safe Drinking Water Data Challenge #CAWaterDataChallenge

Global Climate Action Summit

**Remarks from the National Science Foundation**

Big Data Hubs and Spokes

*Beth Plale, PhD*

Big Data Hubs

* West
* South
* Midwest
* Northeast

Midwest

* Water Quality Issues in the Upper Mississippi
* Smart Big Data Pipeline for Aging Bridges

Northeast

* Data-Driven Discovery and Rational Design Paradigm in Chemistry
* Collaborative for Computational Social Science
* Data Integration of Ecological Logic Trail

West Spokes Awards

* Big Data to Promote Community Learning and Impact
* Breaking down barriers for reproducible neuroimaging data analyses
* Building the Community for the Open Storage Network

Anticipated accomplishment areas:

* Programmatic activities that accelerate academic, industry, and community
* Socio- and technical- shared resources/services
* Data science education

**POWDER: Platform for Open Wireless Data-driven Experimental Research**

Kobus Van der Merwe, University of Utah

End-to-end software defined city-scale wireless living lab

Novice through repeatable wireless research platform

State-of-the-art programmable wireless, networking, cloud equipment

Extensibility with BYOD style research

Progammable/observable massive MIMO (RENEW)

<https://powderwireless.net>

POWDER and Data..

Data from the platform:

* Spectrum monitoring
* Networking
* Sensor
* Data

Architectural

* Data processing platform research
* Highly distributed and diverse compute storage

**Open Data: Research Datasets in the Cloud**

*Vani Mandava @vanimt*

microsoftopendata.com

simplify access to data

Data Growth vs Internet Bandwidth Growth

Dual Word Embeddings Trained on Bing Queries

**CoMSES Net: Promoting accessibility, transparency**

*Allen Lee computer scientist*

CoMSES

* Accessibility
* Transparency
* Reusability
* Interoperability

CoMSES BD Spoke Activities

* Open access publishing and peer-review of model code
* International & interdisciplinary collaboration
* Knowledge exchange

Only 10% of code is available.

<https://catalog.comses.net>

**Translational Data Science: Upcoming Workshop Opportunity**

*Rene Baston*

Translational Data Science – 3.0 Advancing a Virtuous Cycle

October 25-26 at NYU.

TDS 1.0 focus

* Definition of TDS
* Success stories and failures

TDS 2.0

* Ethics
* Reproducibility
* Land grant systems
* Incentives
* Education

TDS 3.0

* Challenges of Translational Data Science
* Incentives
* Organizational
* Cultural
* Technical
* Responsible Data Science
* How to embed

What We are Driving At

* Concrete Recommendations – good enough to engage community
* Action – low-hanging fruit + commits to act on them

Pipeline Paradigm

Continuous Knowledge Exchange and Co-creation in Complex Ecosystems

**Women in Data Science (WiDS) Conference and Datathon**

*Maria Klawe*

**Pick one for Community-Led Deep Dive Sessions**

* Data Science for Social Good
* Data Science Pedagogy & Practice
* Water Task Force
* Health & Medicine: Data-Sharing & Collaboration
* Smart & Connected Communities in the West

**Water Task Force**

Agenda:

1. Introduce the Western Water Collaborative that addresses water scarcity by integrating watershed hydro-environmental processes and human socio-technological water systems
2. Brainstorm about a potential hypothesis driven framework to support the links between data, science, and communities in the collaborative
3. Explore big data approaches to sharing, analyzing, and communicating water-related data

Interest in wildfires and water

Not successful in spoke proposal funded

Seeded open water information architecture

Review design of test-bed development

Water-balance

Water-budget

Water-management

Water

* Not only measuring, but who uses it?

Internet of Water concept

* Can generalized beyond California
* Open scientific best practice
* Fish count

Seeding (seed funding) goes a long way

Asking people how much water is in your property is like asking how much money is in your bank

California law – disclose usage of water

Not really about technology, it is social. Political.

Companies don’t want to give out cash cow.

You are fighting privacy and corporate interest on water owners.

The victory: open water architecture – community driven and shared information.

Intermountain water initiative

Scarcity is not only hydro-logical issue, it is also human management issue.

Water scarcity in the west is changing.

Amount of precipitation changes, and timing.

Composition of forest, species.

How can we share data that is different?

Learn from California’s open water architecture

Water budget in the west, 90% of water goes back.

Run the water to plants? To make it beneficial.

Use big data techniques

Social science, gradients of social systems, what variables to capture. Adaptive capacity?

* Individual land owners
* Big entities
* Agencies

How do we draw the boundaries in the conservation on water?

You couldn’t put it in a disciplinary box on water

Look at each state has its own water politics

If you have high-fidelity models…

People demonized agriculture for using water. But it could actually be a resilient buffer. To recharging land.

Duke it out, or bribe a politician.

Let’s try to duke it out with information.

What should we be measuring?

Factors interacting with other factors instead of taking an isolation perspective.

How can we put more money on the bank?

Slow the water down

Reverse the trend on the ground water

Montana – beaver mimicry

Dealing with Big Data is very challenging

Is there a way to be hypothesis driven?

Unified system of water accounting.

Millennium drought in Australia.

Urban water efficiency.

Problem in water rights, they are only in theory and not in practice. Regulations are a bit easier.

Lawsuits are very specific and not helping for the systemic

Library information science problem.

Industry-level water footprint

Hazards – input and output connectivity.

Science gateway community institute

Water managers to respond to climate/water/precipitation change

Forum to put water managers in the West. If it doesn’t exist, we built on that.

1950s people moved to the West due to cheaper housing, sun.

Could we see a reverse migration? Due to water?

Scannable landscape, talk to water managers on what keeps them up at night.

Internet of Water – is kind of vague and a marketing piece.

USGS – matured as a data portal, water census for water use.

Its not the ‘mean’, it’s the ‘variant’. Its all about the ‘tails’.

Plenty of water, but distribution is a problem.

Harvesting or creating new water. There would be more need for water in the future.

Geo-engineering, like cloud seeding, use of solar flare.

Global climate change is a temporal change.

**Plenary Panel: Spotlight on Education & Training**

**UW eScience Institute**

*Ed Lazowska – founding director and senior data science fellow – UW eScience Institute*

eScience Institute

* Community
* Education
* Research

Creating Institutional Change in Data Science – January 2018

* What worked and didn’t work.

<https://msdse.org/creating_institutional_change>

Data Science Leadership Summit – March 2018

* Decisions (table) you have to make to ramp-up data science in your school
* Compendium of choices

Committee on Envisioning the Data Science Discipline: The Undergraduate Perspective – May 2017

Education

* Formal
* Informal
* Tutorials
* Hackweeks

**Carpentries**

*Kari L. Jordan, PhD @drkariljordan* – Ohio State University

Transferred over to data carpentry

* First question to the speaker in the program, R or Python?

Carpentries (Software Carpentry and Data Carpentry)

* Founded in 1998
* 2-day workshops on git, python, cloud computing
* focus on non-traditional approach
* very popular and oversubscribed. Labs in UW were flipped several times to accommodate.

*Clark’s notes: email Kari – be bumped up for the instructor application process.*

**Big Data to Promote Community Learning and Impact**

*Pamela Scott-Johnson, PhD* – Dean, College of Natural and Social Sciences California State University, LA

Education for the public good

Partnership between city of LA, GeoHub (data portal of City of LA)

Partnered with Community Partners – non-profit orgs

SEED – social equity education

Students flock to Data Science for ‘Social Good’

**Online M.S. in Information Systems**

*Vandana Janeja – UMBC*

Data Science must have a focus on service in communities

Develop a draft “playbook” for data science projects and volunteer policies.

Create an advisory board

Identify potential pilot projects

**The Public Cloud in Research Computing**

*Bhardwaj, G.*

Group tends to do invading.

Reproducibility, digital curation, collaboration

Build-Test-Share

Research computing is moving to the cloud

Resistant factor for the cloud is entrechment

**Collaboration Opportunities**

**Moore – Sloan Data Science Environments**

*Micaela Parker –* program coordinator – Moore – Sloan Data Science Environments

Started by inviting the university in the design phase

3 universities did a 5-year grand experiment. Tackle data deluge. For domain researchers.

**Earth Science Information Partners**

*Annie Burgees, PhD – Director ESIP Lab*

Observation to Algorithms

Modify algorithm for satellite observation

Operational workflow

Validation

Analysis

ESIP Lab supports Earth science developers build novel, applied technologies through rapid funding and community input

Challenges: Idea, technical and scientific competitions open to a broad community of solvers

Incubation: Explore new tech, standards, workflow; succeed and fail fast

Workshops: In-person synthesis-center style activities to work on a common problem

Evaluation: tackling the ‘valley of death’ between development and implementation

Adoption: lab projects are open for reuse by the broader science community

**Utilizing Collaboration Effectively**

*Beth Plale –* Program Officer, National Science Foundation

Google went to Michigan and digitized the library

**Open Storage Network: Distributed National-Scale Storage for Research**

Repository for Active Publicly-Funded Research Data

Six Prototype Development Sites

* John Hopkins
* Northeastern Storage Exchange
* Northwestern University
* UC San Diego
* UNC
* Univ of Illonois

Openstoragenetwork.org

RDA Plenary 13

[www.rd-alliance.org](http://www.rd-alliance.org)

**Water management in the West**

SCUBA – Strategic California UrBan water Analytics

$20 million saved through improved water demand forecasting

**Table Talk**

Steve Kutchin

* Boise University
* Get the faculty be interested and involved
* Computing Research program as a way to have a “data science” program

Sam Fernald (Professor) – New Mexico Water Resources Institute

Water Data

* Water test data issues
* Still hard to use. So much data. Need data science.
* Scarcity in water

Vani Mandava - Director

* Launched Open Data from Microsoft

Google public dataset from schema.org – will be indexed by the search engine

Jeffrey Baehr – BI Developer – JR Simplot company - Boise

* Agriculture Data
* Has a lot of data, lacks the people and community to do data science
* Uses Power BI, Web Focus
* Growing pains with Tableu

*Clark’s notes: I suggested doing hackathons, doing tie-ups with coding bootcamps, software carpentry to create a community.*