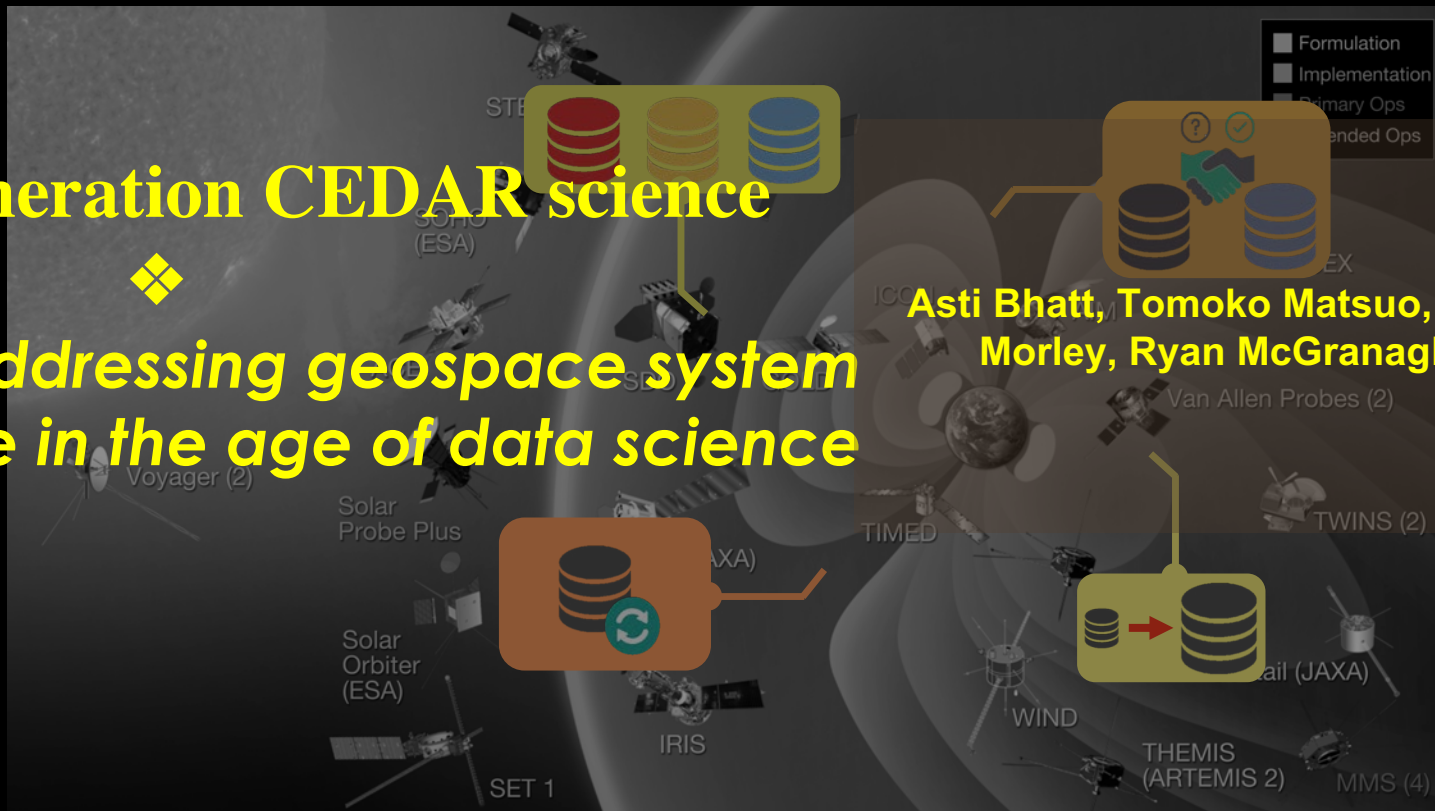


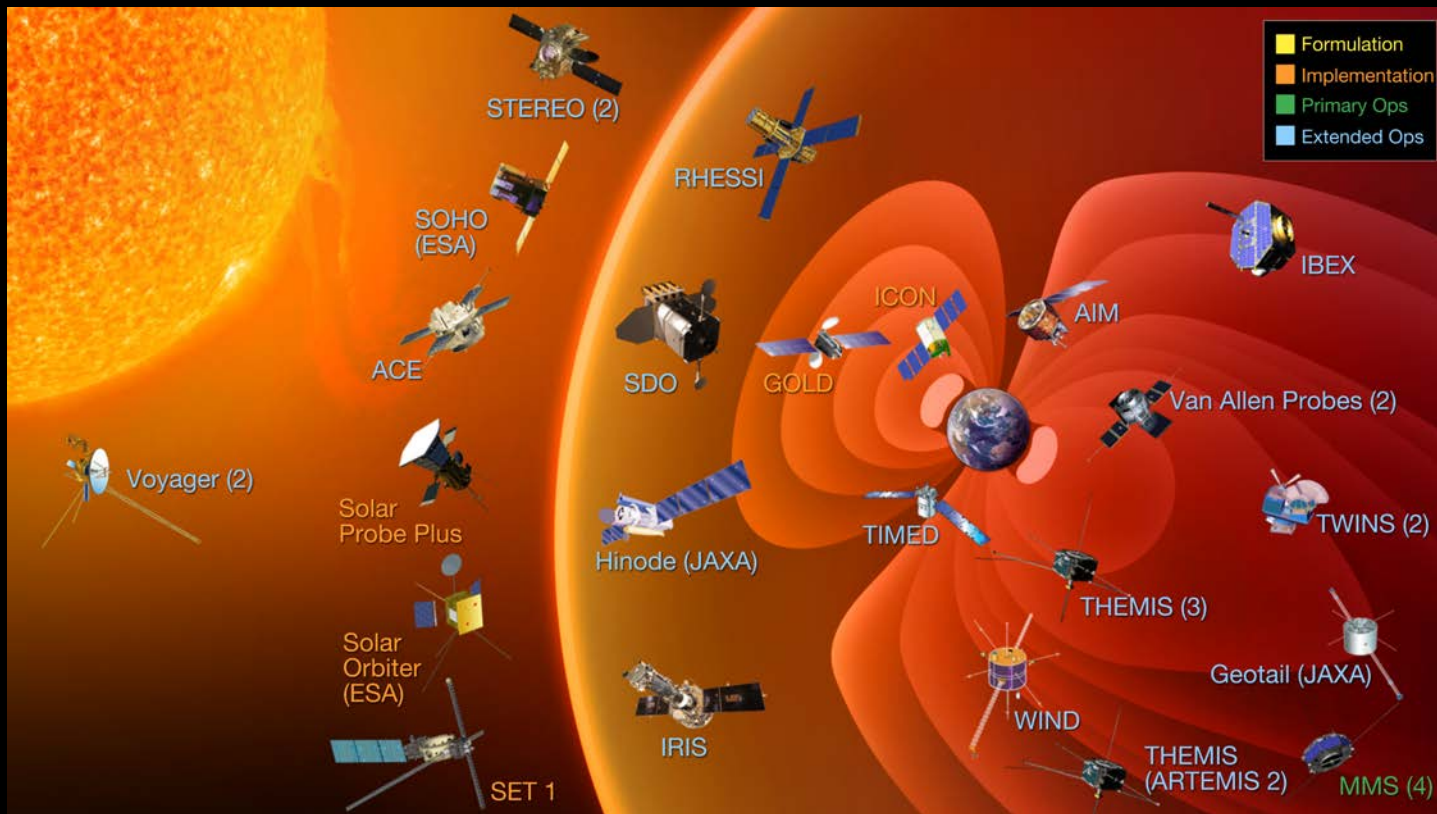
Next generation CEDAR science

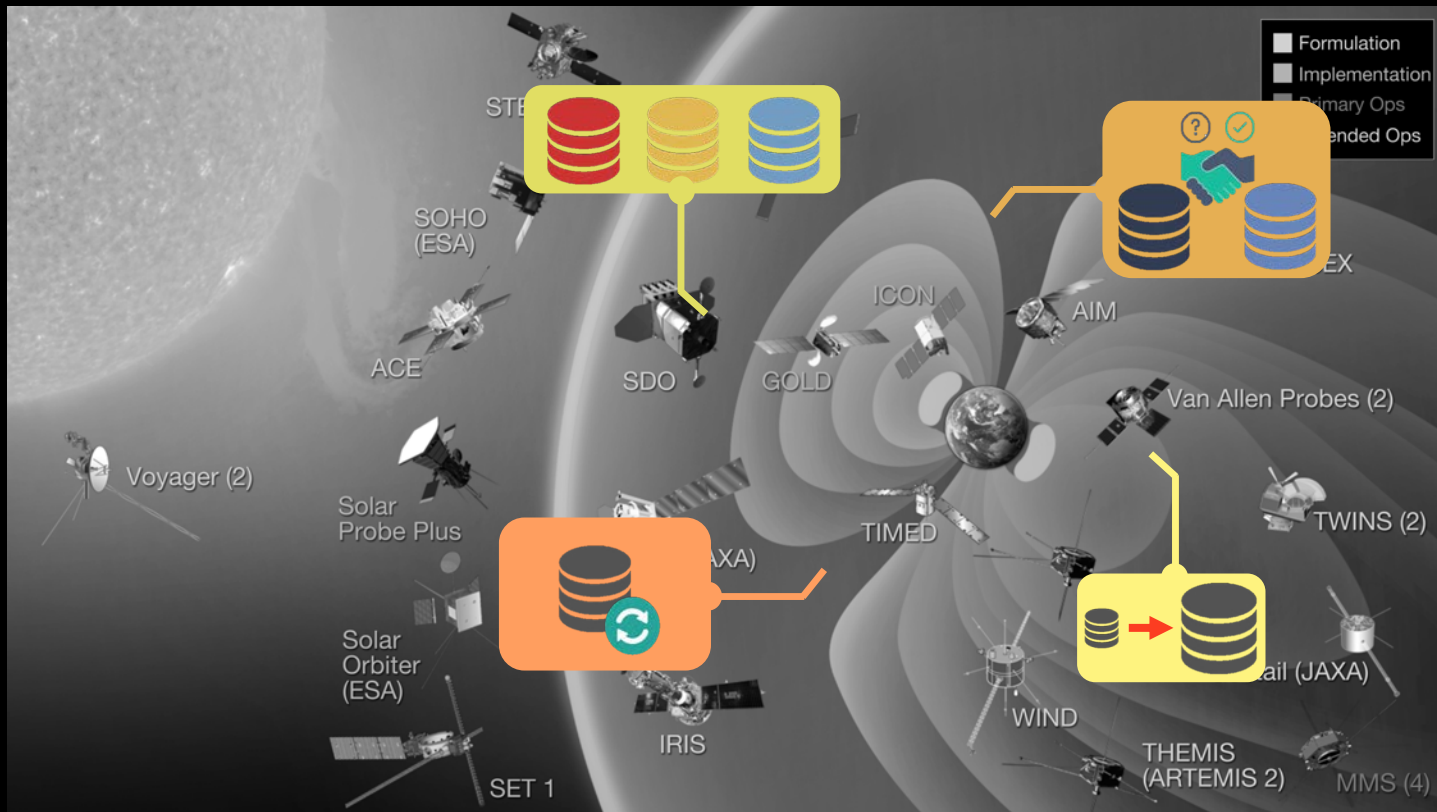


Addressing geospace system science in the age of data science

Asti Bhatt, Tomoko Matsuo, Steven Morley, Ryan McGranaghan







Opportunity:

- Evolve traditional approaches
- Embrace data-driven discovery
- Enable interdisciplinary work

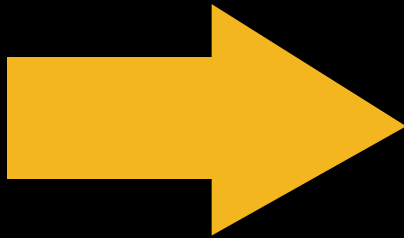




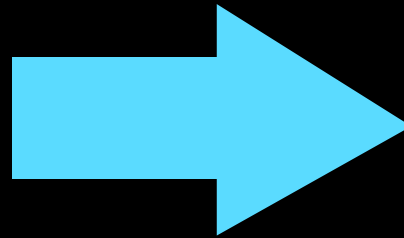
Questions:

- To what extent do traditional methods of geospace research overlap with data-driven techniques?
- What are the powerful use cases for applying data science in geospace?
- What is needed for more efficiency in geospace data processing and analysis?
- What are the immediate next steps to embrace data science approaches in geospace?

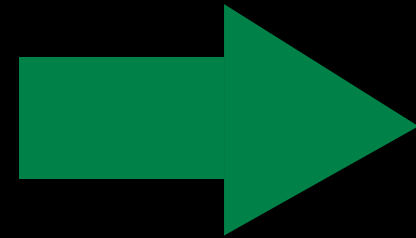
Blueprint for today



**Contributed and
solicited talks**



**Round-
table/breakout
discussion groups**



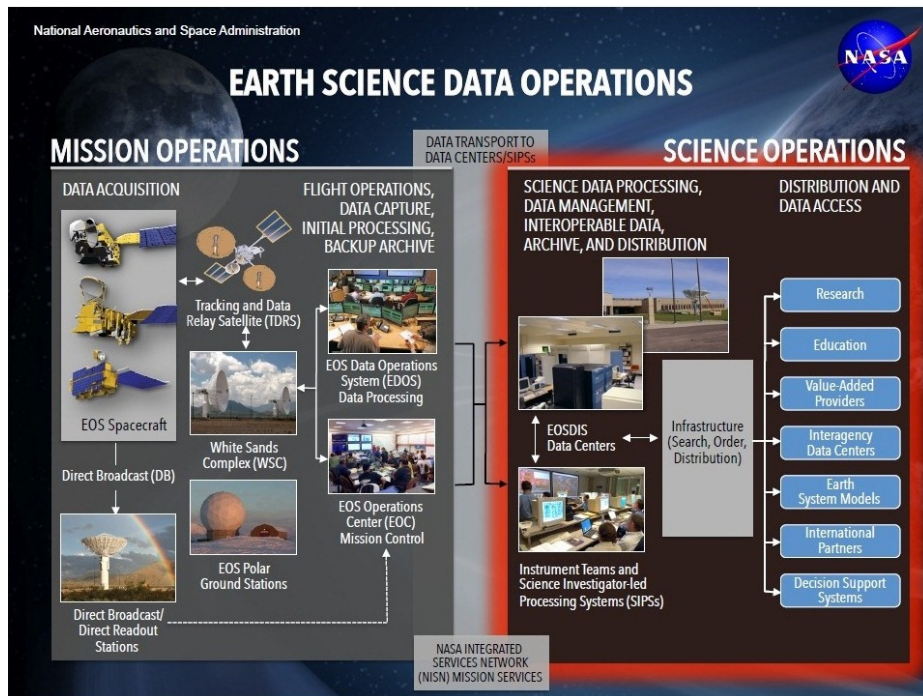
Convergence

Ryan McGranaghan

Discussion Slides

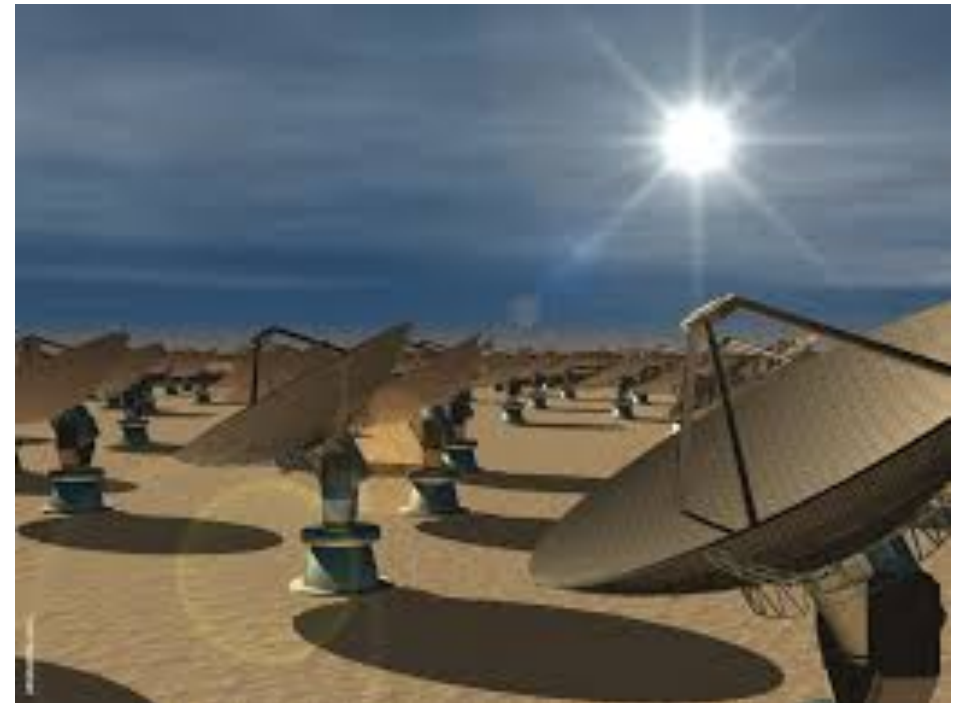
How do we work with and learn from 'sister disciplines' (e.g., Earth Science)?

Earth Science → Earth Analytics



Data Variety: space- and ground-based, airplanes, diverse sensors

Astrophysics → AstroInformatics



Data Volume: Doubling of data every ~1.5 years

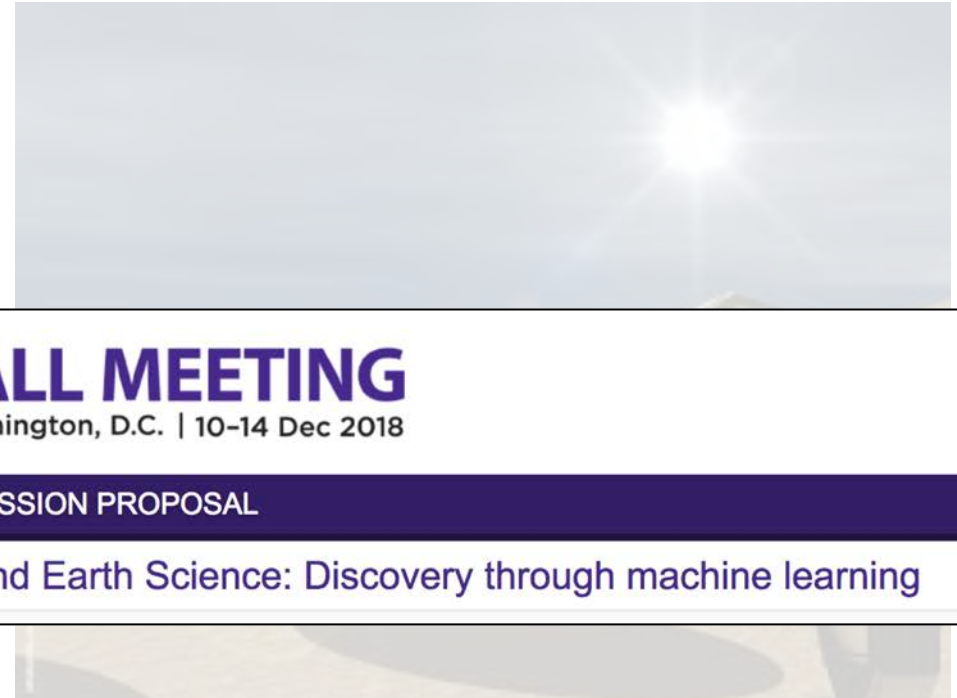
How do we work with and learn from 'sister disciplines' (e.g., Earth Science)?

Earth Science → Earth Analytics



Data Variety: space- and ground-based, airplanes, diverse sensors

Astrophysics → AstroInformatics



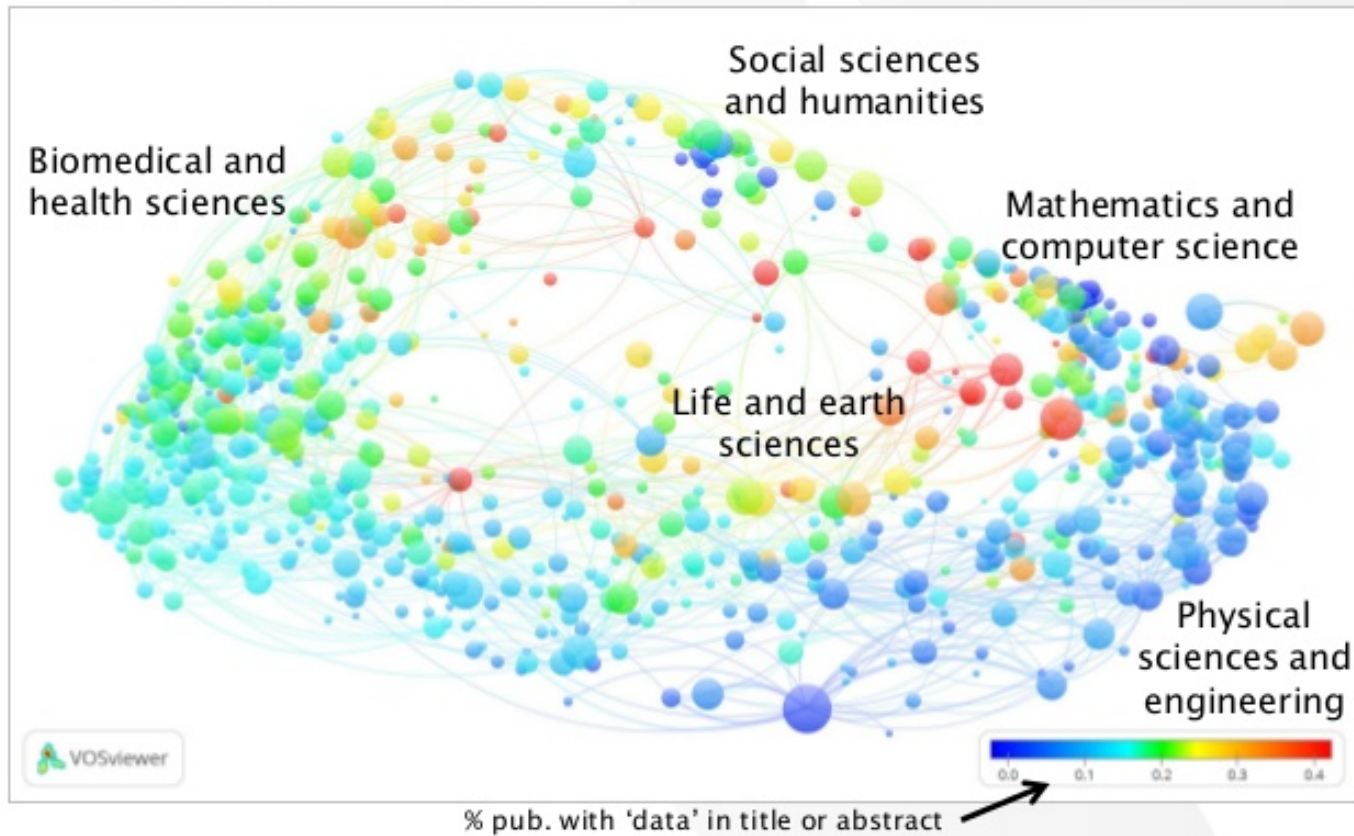
Data Volume: Doubling of data every ~1.5 years

AGU 100 | **FALL MEETING**
ADVANCING EARTH AND SPACE SCIENCE | Washington, D.C. | 10-14 Dec 2018

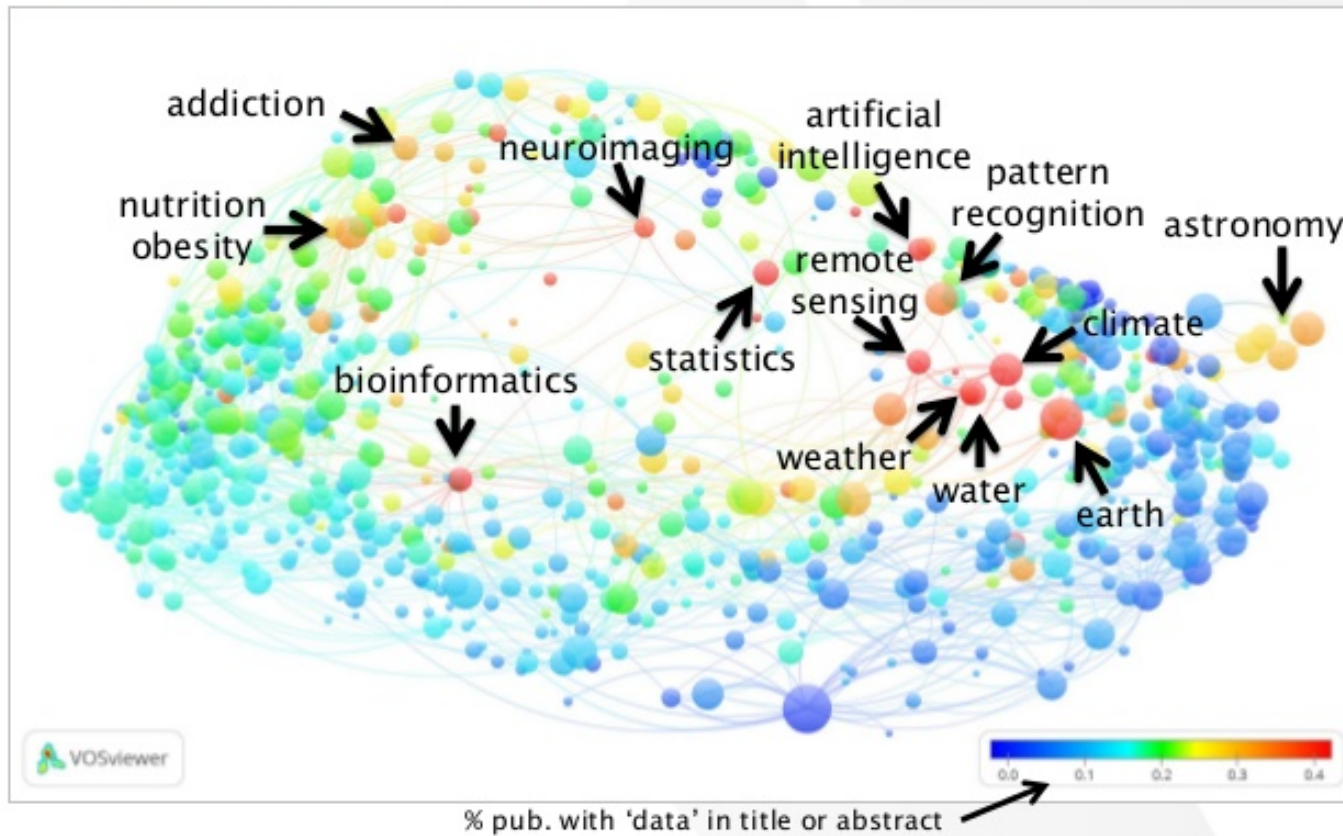
SESSION PROPOSAL

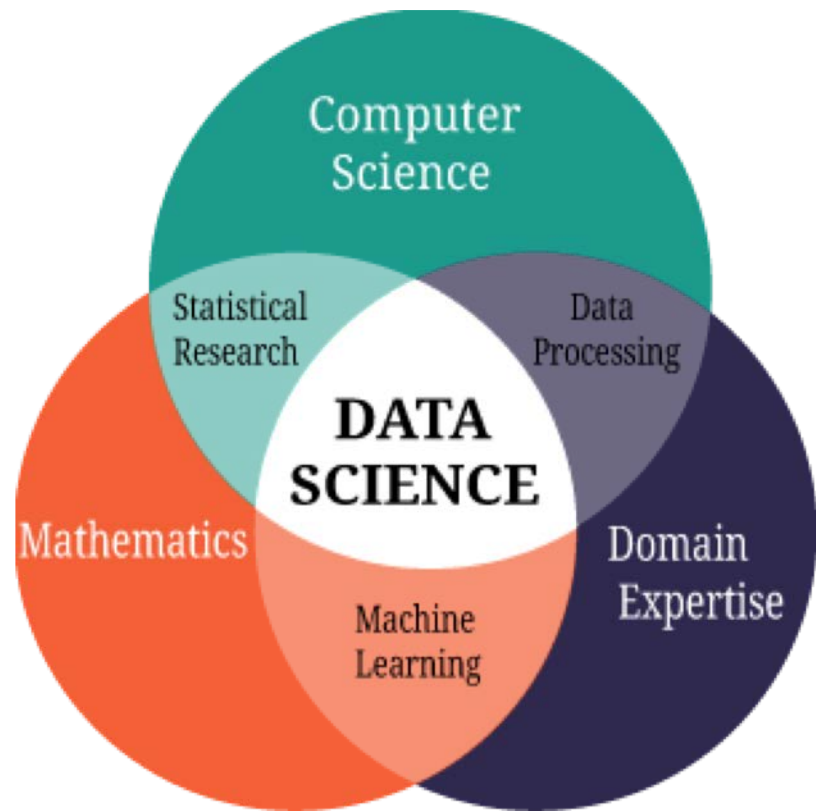
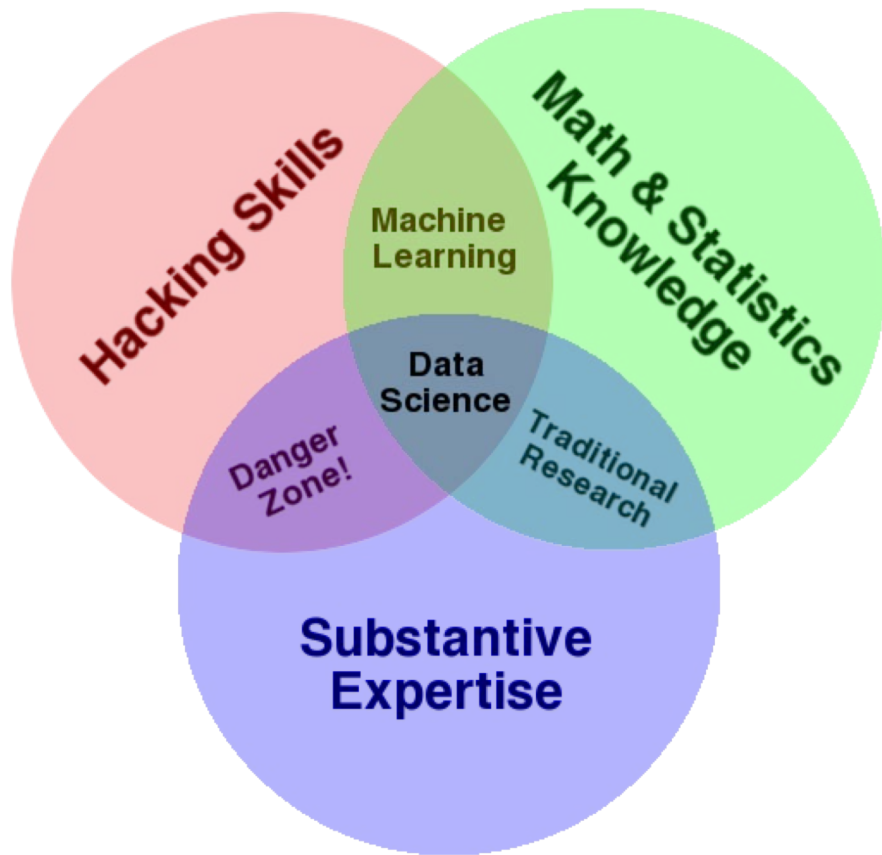
IN018: Convergence in Space Physics and Earth Science: Discovery through machine learning

Data-driven nature of different scientific fields



Data-driven nature of different scientific fields





Is an 'analytics center' a way to focus? (i.e., harmonize data, tools, and computational resources to free researchers to focus on investigation)

How do we make progress toward this?



Image credit: NASA, *technology drives exploration*

Example: Earth Science Division Advanced Information Systems Technology (AIST) Program

Objectives

The objectives of the AIST program are to identify, develop and (where appropriate) demonstrate advanced information system technologies which:

- Reduce the risk, cost, size, and development time of Earth Science Division (ESD) space-based and ground-based information systems,
- Increase the accessibility and utility of science data, and
- Enable new observation measurements and information products.

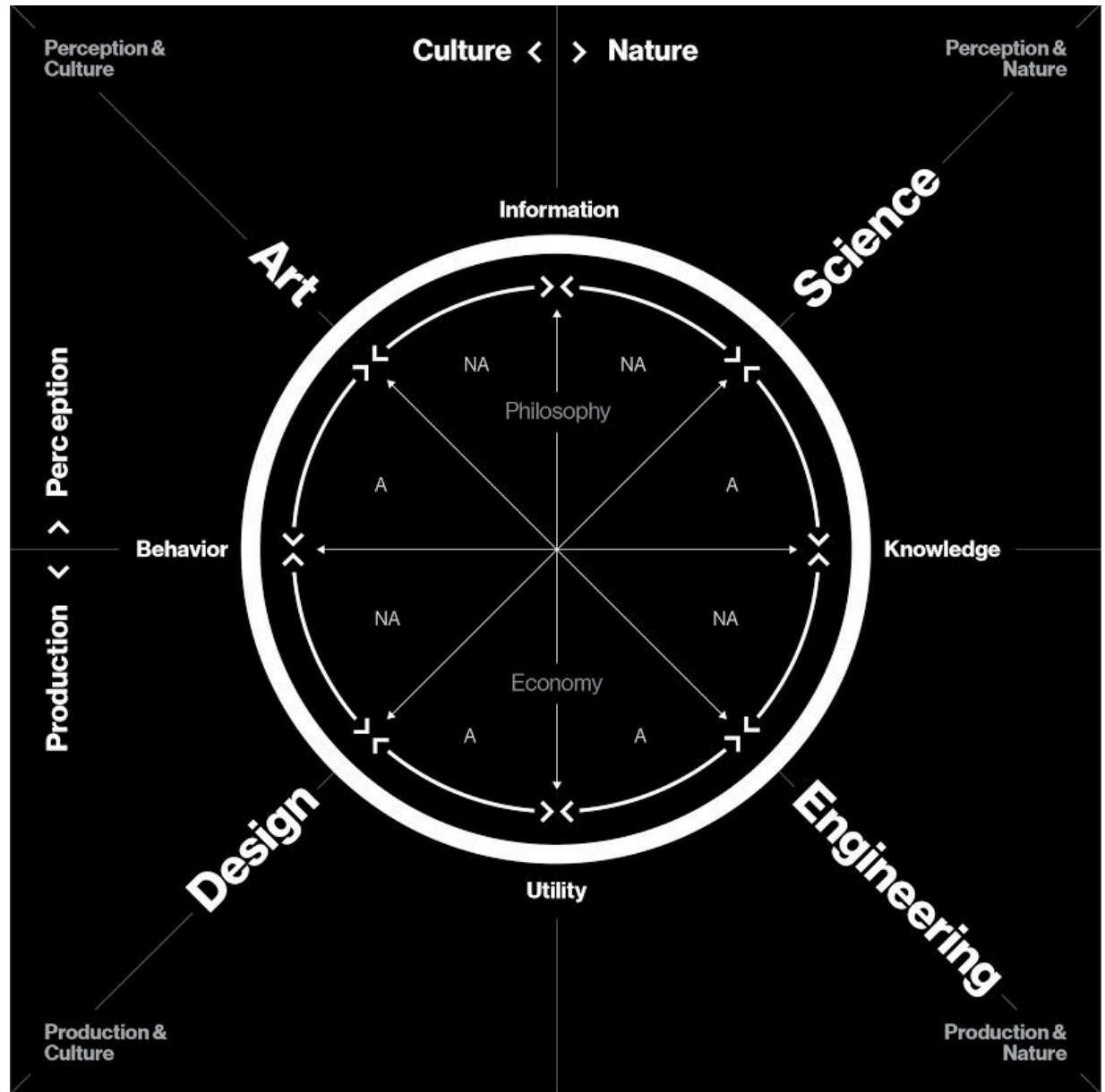
Founded on technology drivers: The AIST Program is founded on technology needs for Earth science measurement, analysis, and application objectives.

What is the sociological of these changes? (i.e., how do we change the culture?)

Focus on the *home runs*? What are the *home runs* from these approaches?

What are the fantastic use cases in the field of geospace?

What does the intersection of science and design look like?



Why do we not practice open policies?

Barriers to Data and Code Sharing in Computational Science

Survey of Machine Learning Community (Stodden, 2010):

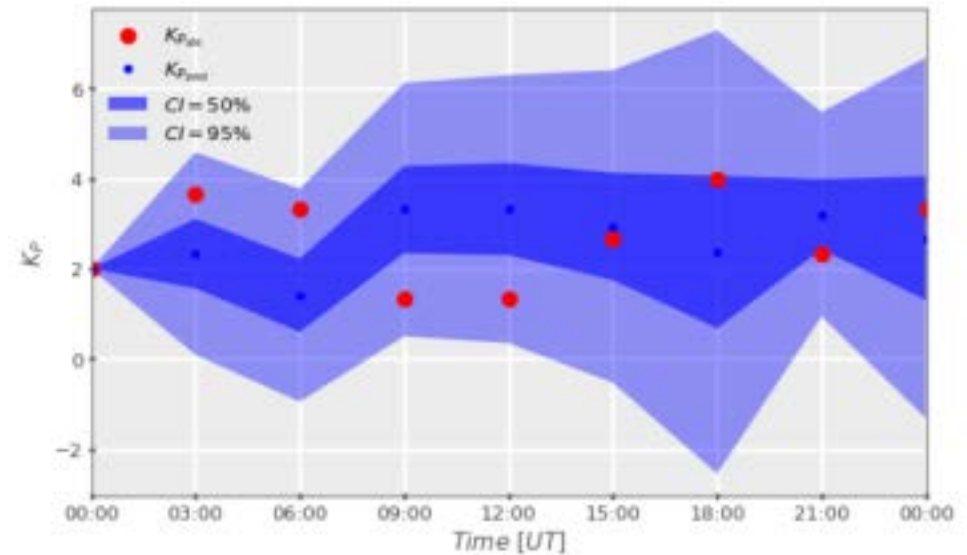
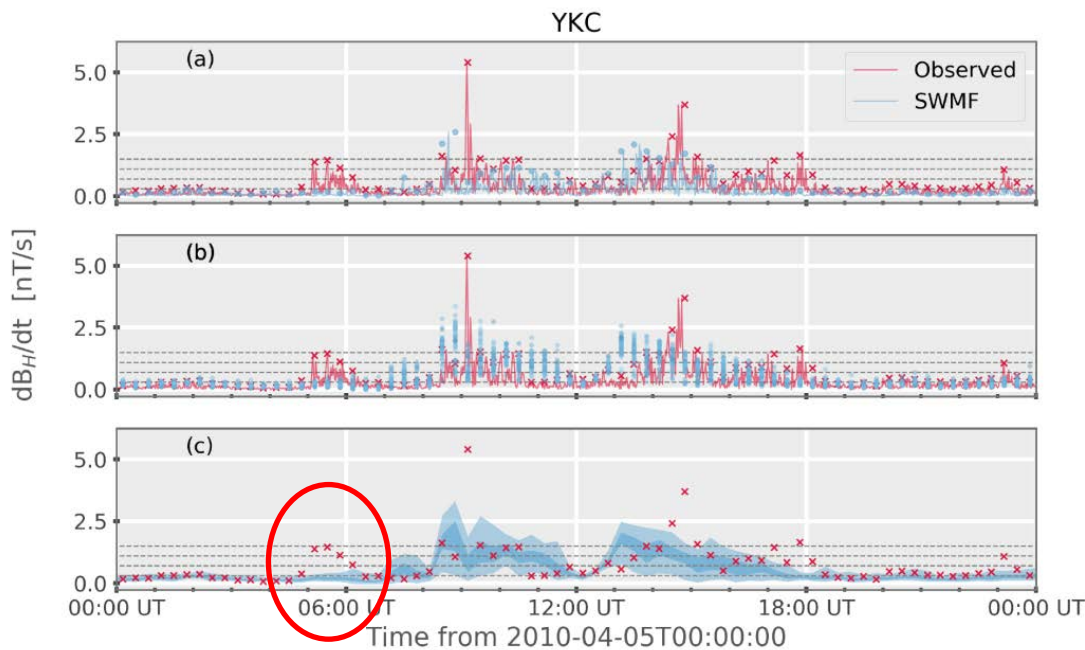
Code		Data
77%	Time to document and clean up	54%
52%	Dealing with questions from users	34%
44%	Not receiving attribution	42%
40%	Possibility of patents	-
34%	Legal Barriers (ie. copyright)	41%
-	Time to verify release with admin	38%
30%	Potential loss of future publications	35%
30%	Competitors may get an advantage	33%
20%	Web/disk space limitations	29%

Steve Morley

Discussion Slides

Physics Scavenger Hunt

How can we use data to learn the physics missing from our models?



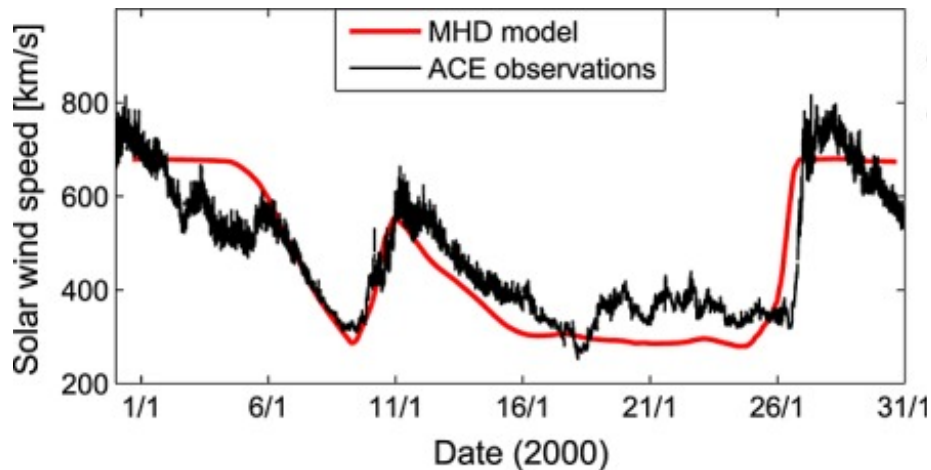
How can we use machine learning to give understanding of physical processes?
What can machine learning bring beyond a more accurate prediction?

UNCLASSIFIED

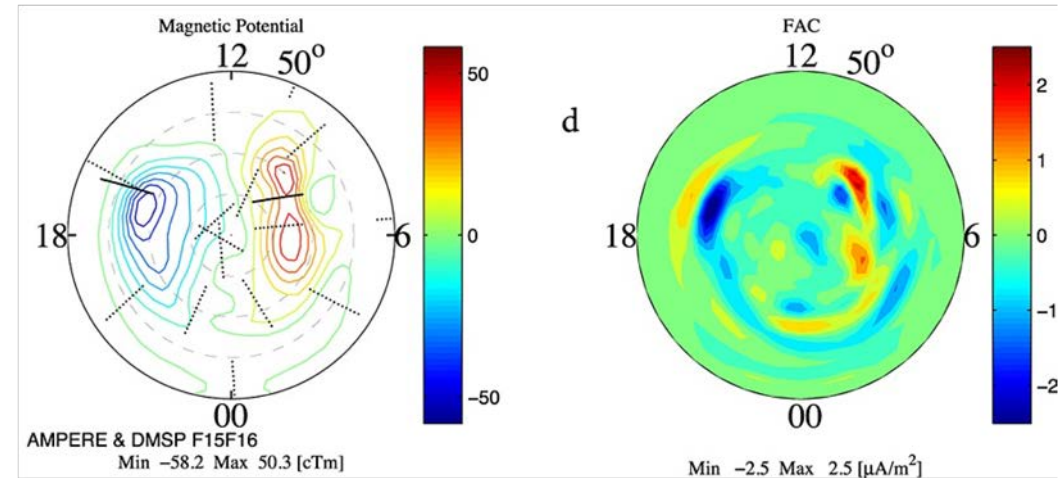
Science across scales

Models often miss small-scale features

Distributed data sets can give a low-resolution global picture



Owens et al., 2014



Knipp et al., 2013

How can we effectively include fine-scale structure? (AKA “downscaling”)

What about disparate data types or quantities?

UNCLASSIFIED

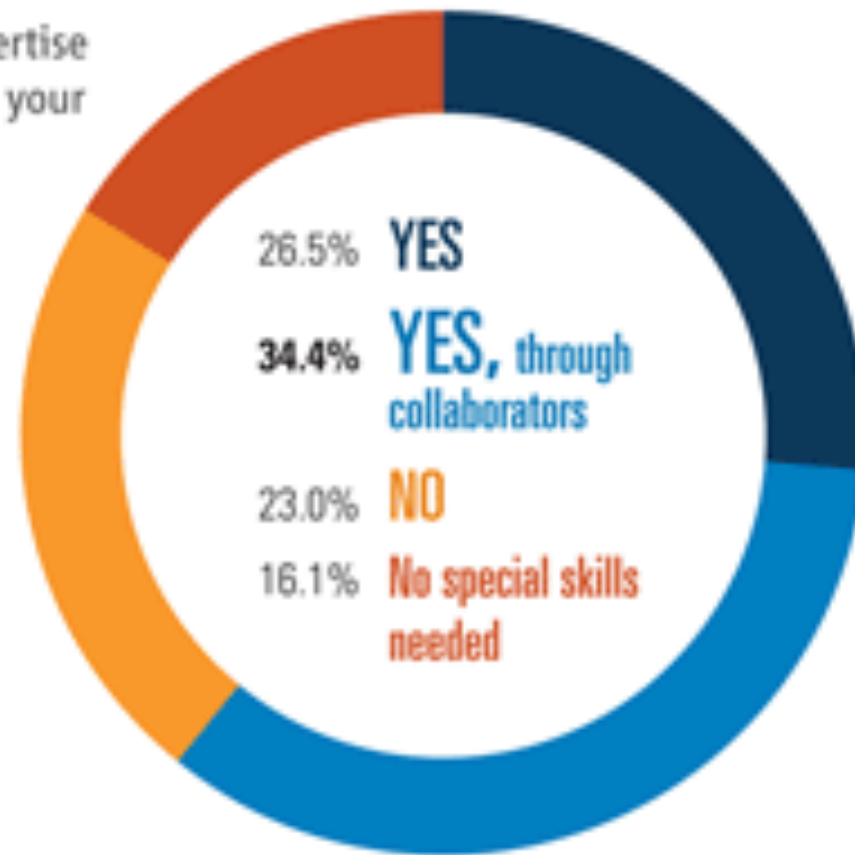
Tomoko Matsuo

Discussion Slides

How to leverage existing resources in other fields (NWP, computer science, applied mathematics) to advance our geospacer data science

Do you have the necessary expertise in your lab or group to analyze your data in the way you want?

“The next few years [particularly in medicine] the volume of data we need to analyze will expand exponentially.”



[Science, 2010]

Asti Bhatt

Discussion Slides

What are the barriers for accessing large geospace datasets?

What can we do to remove these?

What changes does the geospace community need to adopt to efficiently use data-driven techniques?

Suggested breakout groups

Open data and code

- Discussion lead: Asti Bhatt

Convergence with other fields

- Discussion lead: Tomoko Matsuo

Coding and software practices

- Discussion lead: Steve Morley

Machine learning

- Discussion lead: Ryan McGranaghan