Strengthening Research Capacity

Lessons learned from a large, international, cross-disciplinary research expedition

MOSAiC

Multidisciplinary drifting **O**bservatory
for the **S**tudy of **A**rct**i**c **C**limate

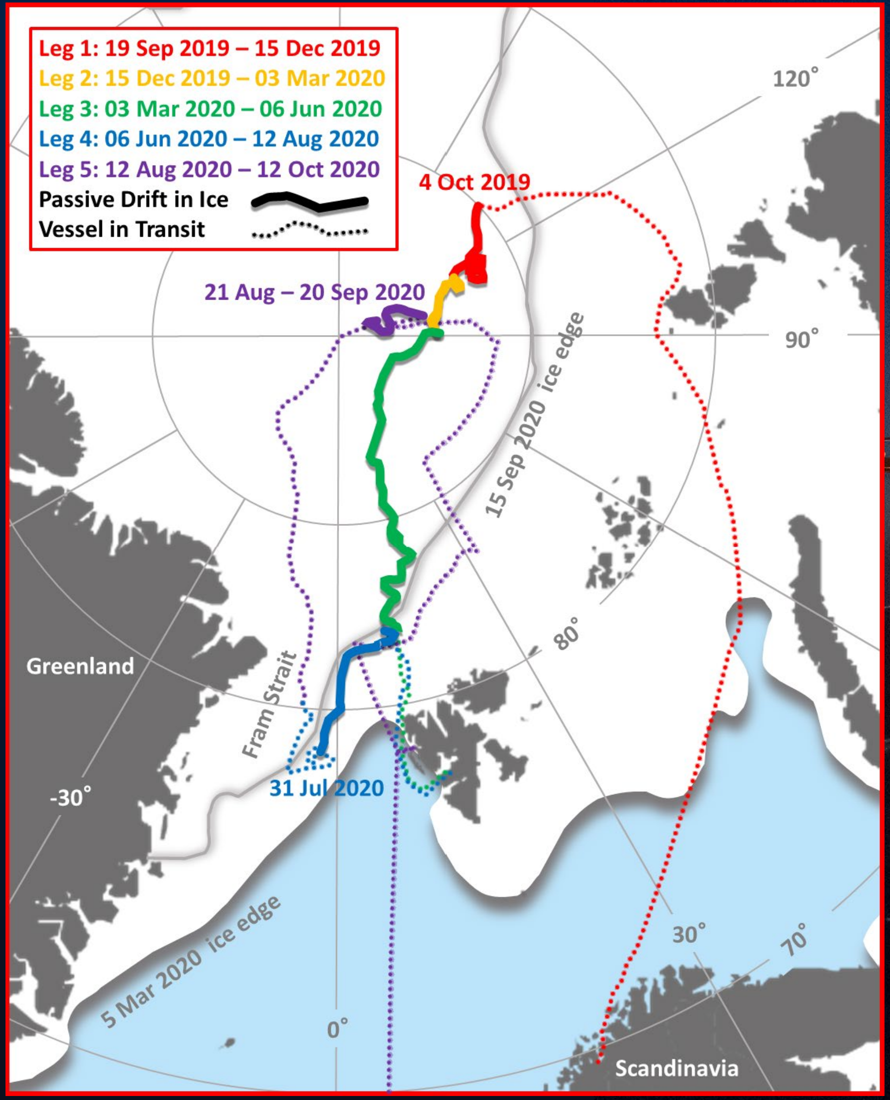
Photo: Markus Rex

Matthew Shupe

CIRES/NSIDC - University of Colorado

NAS IPY Workshop, 21 May 2025

What is MOSAiC?



The Expedition: Sept 2019 – Oct 2020

- *20 Nations, 7 Ships, 400 Field people, 80 Institutions, $170M*
- *"Causes and Consequences" of rapid sea ice loss*
- *Motivated by model development and "coupled processes"*
- *Polarstern & Central Observatory within 2 km*
- *Distributed network out to 200 km, plus satellite observations*
- *Periodic resupply, mostly by Russia*
- *Data publicly available: PANGAEA, DOE ARM, Arctic Data Center, etc.*

Building and sustaining capacity

Dave Stenssen

Leadership
Early Career
Infrastructure
Cross-disciplinary

Enabling Leadership

Big science does not happen without champions

They do the hard work, bring people together, look out for the big picture and the small details

Esther Horvath

Enabling Leadership

Leaders need support:

- Mechanisms for identifying and building engaged communities
- Leadership training and support
- Coordination and travel resources
- Explicit leadership funding
- Many cannot afford to be champions (esp. ECRs)

Sebastian Grote

Empowering Early Career Researchers

Engage Early Career Researchers in all Aspects – Leadership, Design, Fieldwork, Analysis, Synthesis

Fresh new perspectives!

Lianna Nixon

Empowering Early Career Researchers

- ECRs are adept at cross-disciplinary research, products, & synthesis
- ECRs bring new perspectives and tools for research (i.e., AI approaches)
- Need to support ECRs in leadership (often support-limited)
- Need to ensure the ‘safety’ of ECRs in the field (and everyone else)

Lianna Nixon

Large Research Infrastructure (more and more effective)

Often the foundation & building block of big science

Nucleator for drawing participation

Need commitments early

Example: An available icebreaker or station enables participation of many

Stephan Hendricks

Large Research Infrastructure (more and more effective)

Stephan Hendricks

US approach can be limited

- Icebreakers capacity is limited
- Few people dominate the large infrastructure leadership (both management and science)
- Need more distributed facilitation of large infrastructure

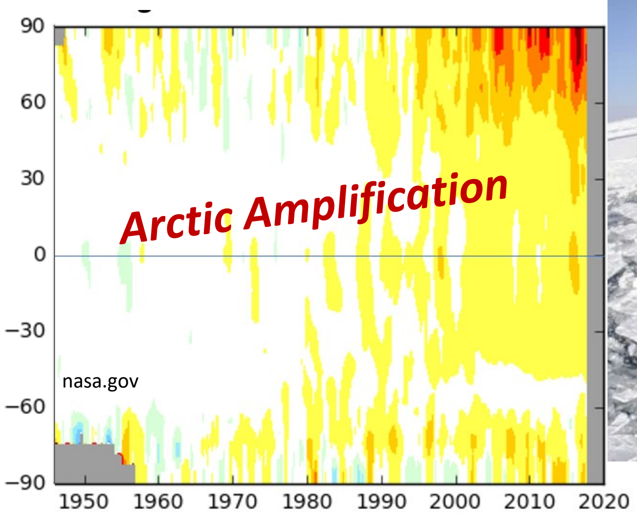
Observing System Development

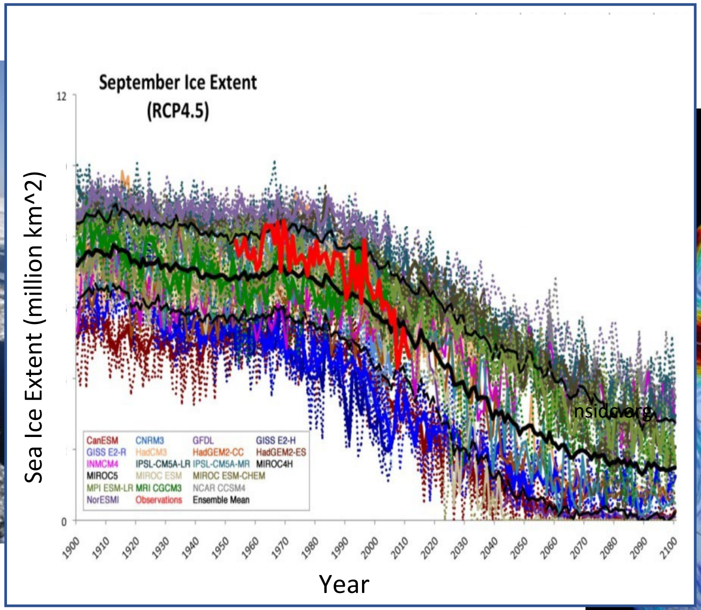
Need systems that…

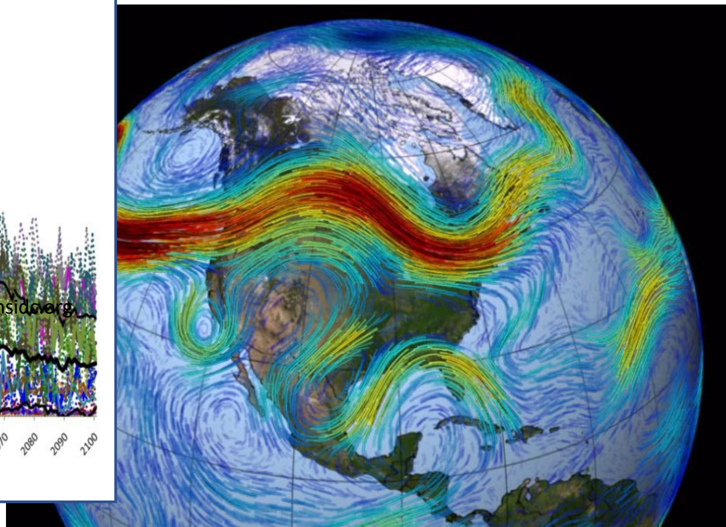
- Bridge scales and cross interfaces
- Enable autonomy, networking
- Fill critical gaps (e.g. energy budgets, etc.)
- Sustainable power systems

Why cross-disciplinary?

- Rapid change is a cross-disciplinary problem (sea-ice decline, amplification, ecosystems, land surface)
- Models often lack cross-disciplinary capabilities
- Emerging operational/management needs are cross-disciplinary (physical, chemical, social, economic, ecosystem, geopolitical,…)
- Dearth of cross-disciplinary observations







Cross-disciplinary Research

Need to develop cross-disciplinary……

- Thinking – Brings disparate people together and amplifies impact
- Planning – Shared vision of science goals and outcomes
- Organization – Leadership and governance structures
- Funding – Better link different mechanisms, programs, agencies & nations!

Cross-Disciplinary

Research activities,
observing systems, meeting structures, data products, synthesis activities….

Cross-disciplinary

- Modelers can be a good cross-disciplinary link
- Reach beyond science (economics, communities, resources, etc.)

Lianna Nixon

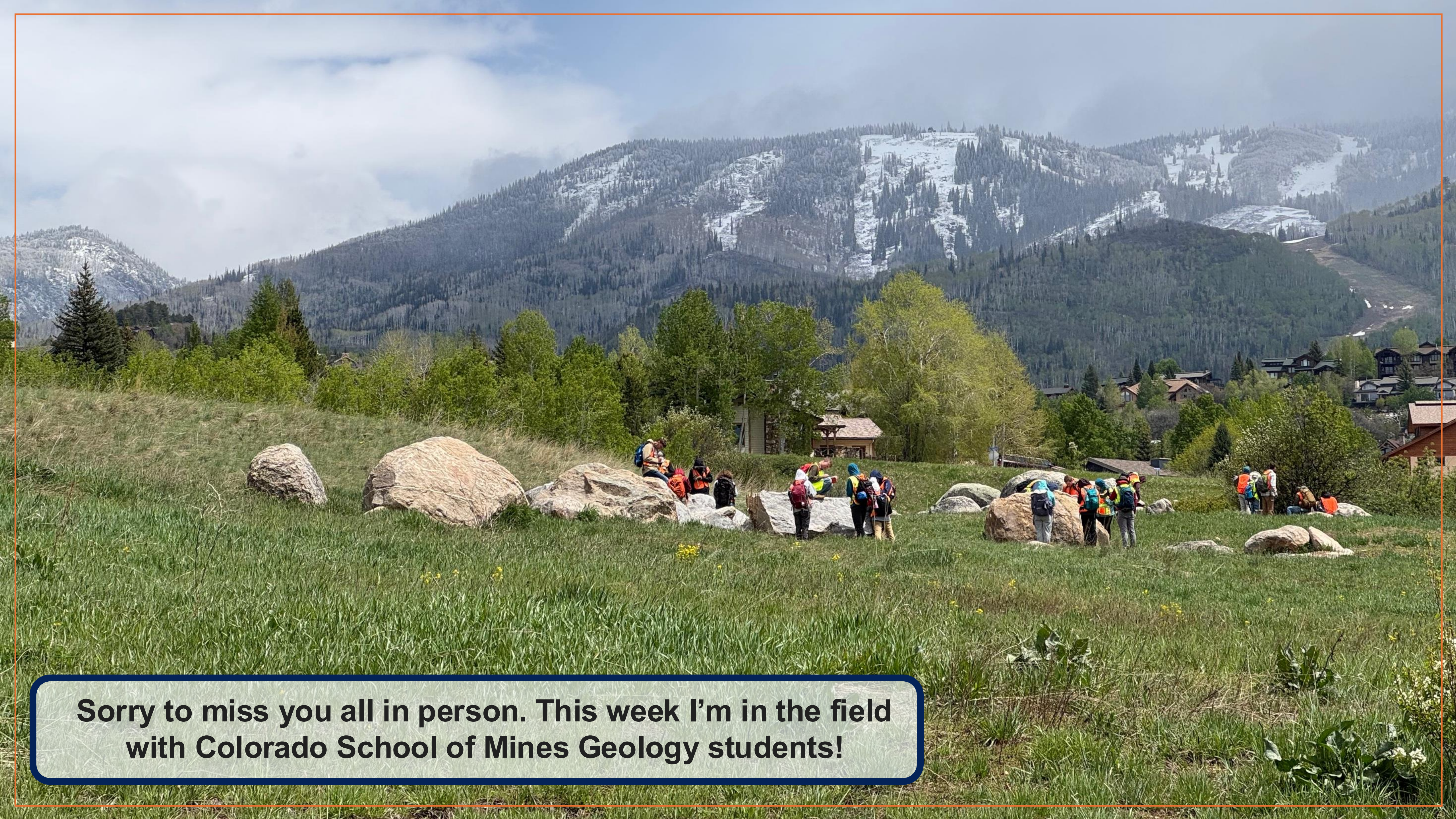
Summary (perspectives from MOSAiC)

- Need to cultivate & support our leaders & champions
- Early career researchers are essential and impactful, but need meaningful roles, guaranteed safety, and dedicated support
- Large infrastructure should be identified and committed early to promote project development
- Need targeted development of observing technologies to fill critical gaps (i.e., autonomous surface energy budget measurements)
- The most pressing research challenges are increasingly cross-disciplinary and require cross-disciplinary thinking, approaches, and solutions
- As sea ice declines society needs more focus on sea-ice understanding and forecasting for many important applications
- Open data policies more effectively build communities and serve broader stakeholder needs
- Focus on what is uniquely IPY. What cannot be accomplished if not for widespread coordination & collaboration? This must be a strong metric for prioritization.

Strengthening Research Capacity: Building on Lessons Learned

Ryan A. Venturelli | May 21, 2025 | **NASEM Workshop on 5th IPY**

COLORADO SCHOOL OF MINES



Sorry to miss you all in person. This week I'm in the field with Colorado School of Mines Geology students!

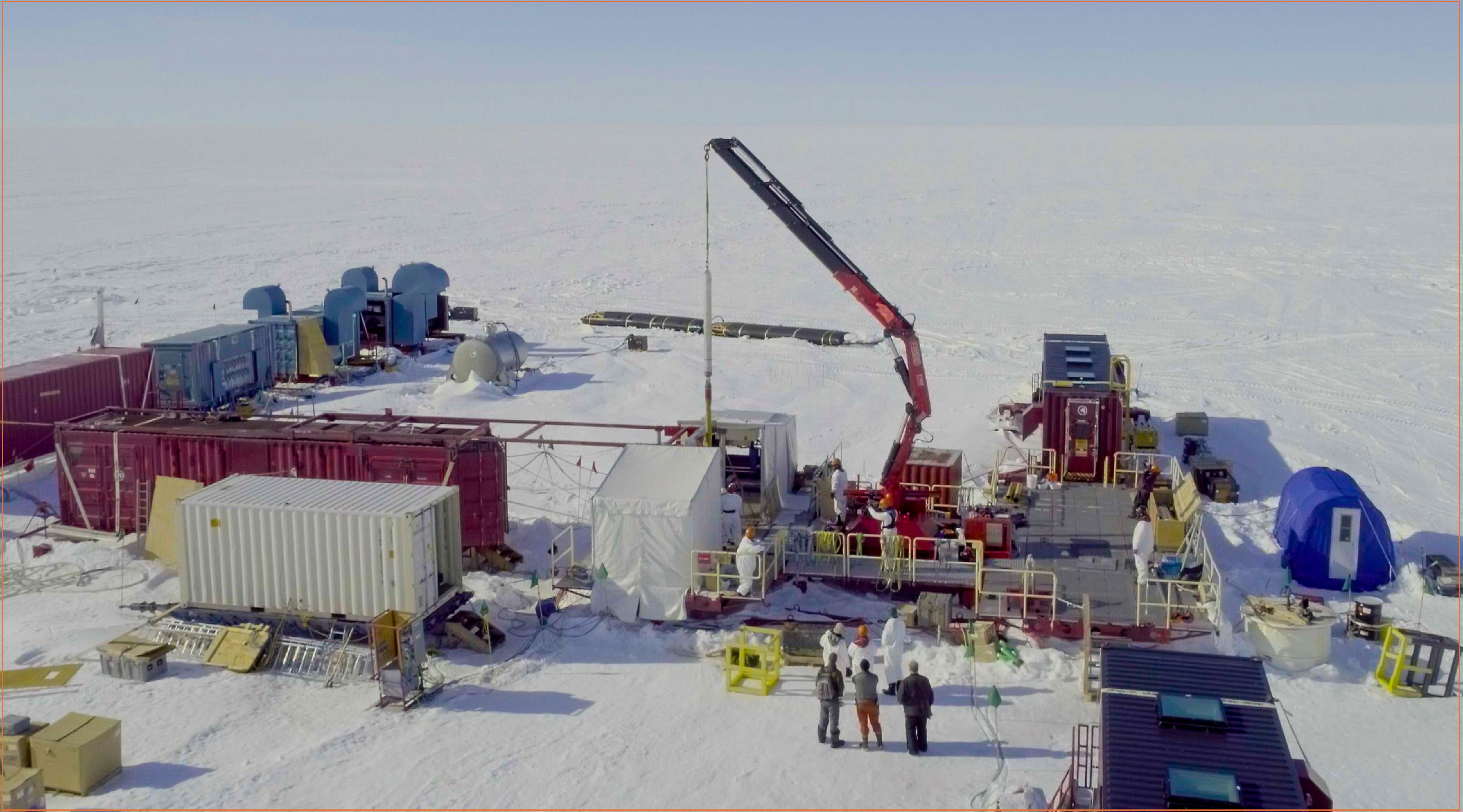
Among the most important questions at our poles:

1. When and under what conditions were the Antarctic and Greenland ice sheets smaller than today?
2. What drives variability at the grounding zone over timescales that range from tidal to millennial?
3. What controls biogeochemical variability in subglacial aquatic systems?

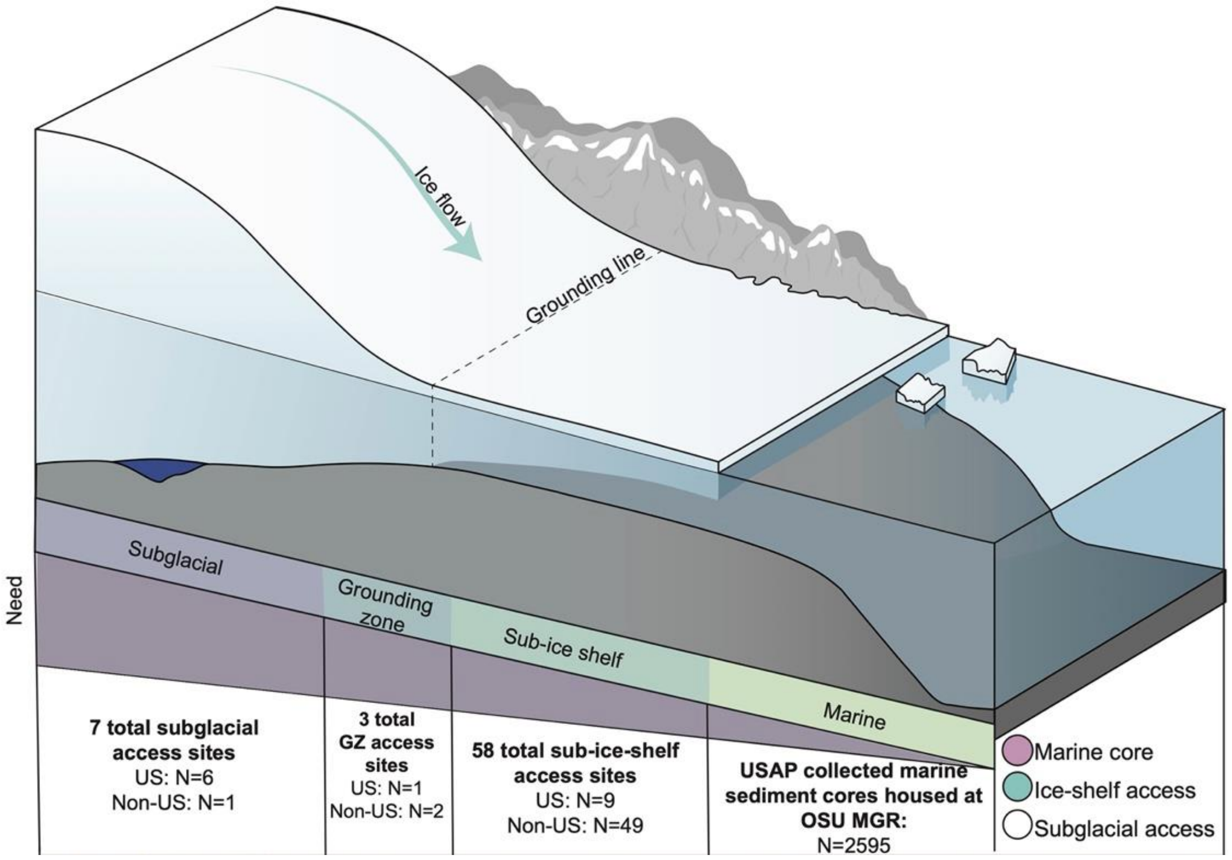
Answering these most important questions will require transdisciplinary teams with strong leadership, vertical and horizontal mentorship frameworks, knowledge co-production, and inclusion of all parties as team members, and international partnership.

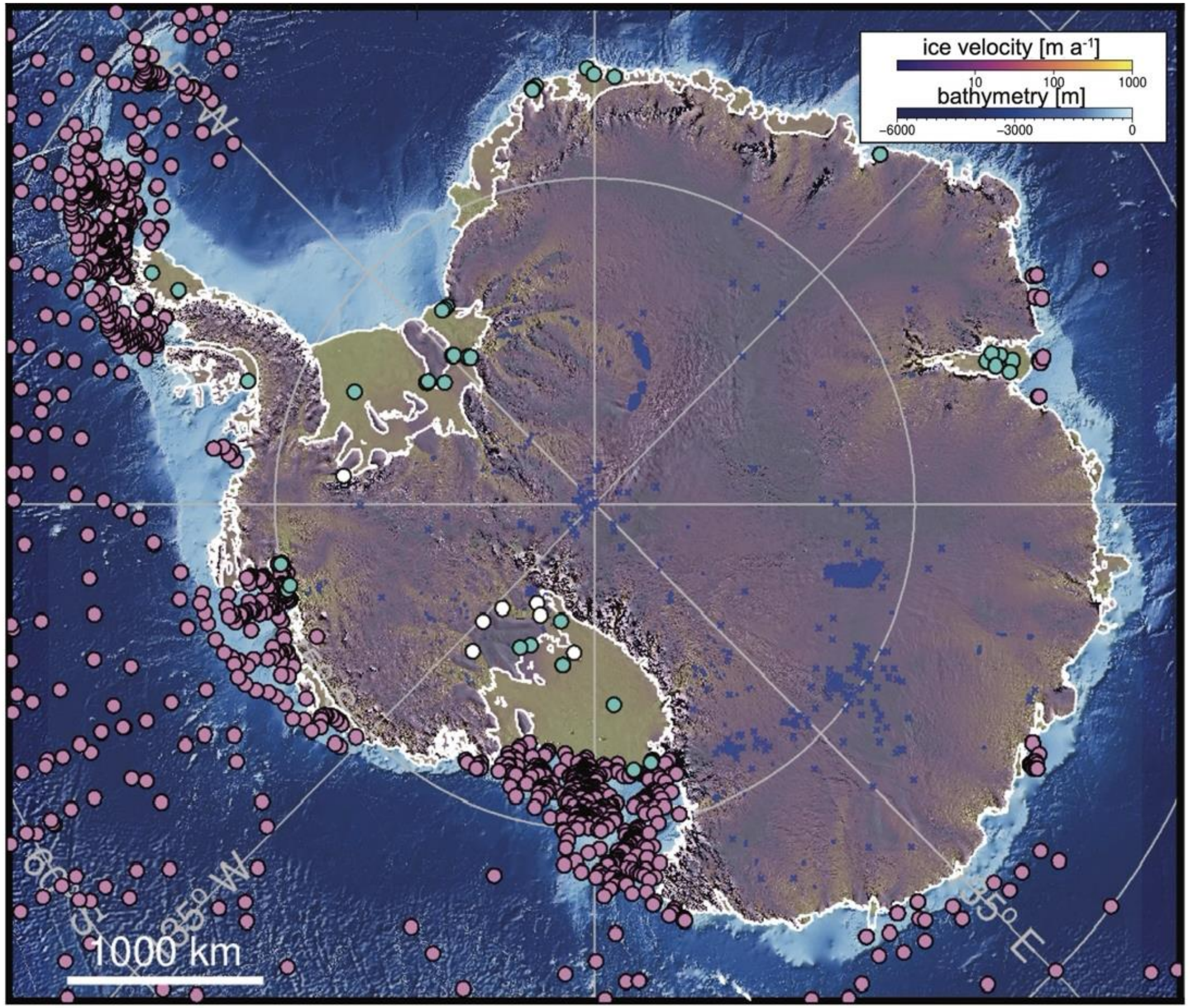
These most important questions require sub-ice access:

1. Enabled by scalable and modular drilling technology with clean-access capability.
2. Achieved through international coordination and collaboration.
3. That explores a range of environments (sub-ice shelf, grounding zone, subglacial) and a range of timescales (tidal, mid-Holocene, Last Interglacial).
4. That leave behind observatories to enable sustained monitoring.



Some context in numbers:





To summarize:

- IPY5 presents the opportunity to build sub-ice drilling campaigns to address the most exciting science targets.
- This work can’t be achieved without our international partners.
- If we want a seat at this table, we need to invest in scalable and modular drilling technology now to enable this type of work.
- These big, transdisciplinary collaborations offer an opportunity to build capacity for future leadership in polar research if we approach them thoughtfully.

Using the Madrigal Database for Polar Science

Katherine Cariglia
MIT Haystack Observatory
cariglia@mit.edu

MIT HAYSTACK OBSERVATORY

IPY5 Workshop
Washington, DC

Moving Towards IPY5 - What Makes Data Last?

- Standard data format
- Complete documentation
  - Consistent naming conventions
  - Parameters
  - Units
  - Error bars/quality flags
- Science ready data
- Storage capacity and backups
- Data is citable
- User perspectives

Research Campaigns - Technical Perspective

- How easy is it to find the data? To use and/or combine the data? To organize the data?
- What is the essential information needed from a user perspective?
- How can lessons learned from Madrigal be useful for IPY5?

Introduction: the CEDAR Madrigal Database

Coupling, Energetics, and Dynamics of Atmospheric Regions (CEDAR)

- CEDAR is a research community investigating the near-Earth space environment
- CEDAR Madrigal database is an open source community resource for science data from various ground and space-based instruments from around the world
- Madrigal is not just limited to space and atmospheric data sets
- Madrigal is a robust web-based system capable of managing and serving archival and real-time data, in a variety of formats
- MIT Haystack Observatory Atmospheric and Geospace Science Group maintains the CEDAR Madrigal database
- Madrigal is part of Millstone Hill Geospace Facility - supported by NSF

MIT
HAYSTACK
OBSERVATORY

How to Get Your Instrument’s Data into Madrigal

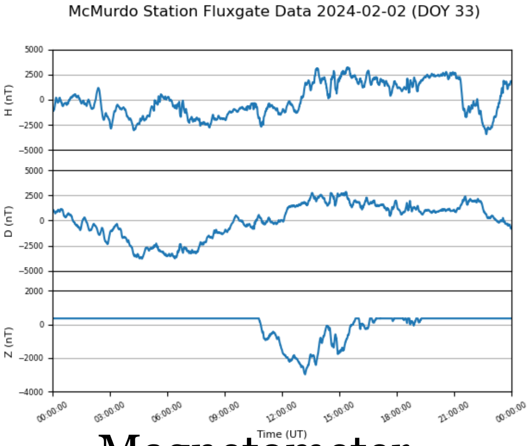
Method 1: Send data to CEDAR Madrigal

- Send data, documentation, and summary plots to MIT Haystack
- Loading program written by MIT Haystack, verified by you
- Add new data in batches or via automated upload

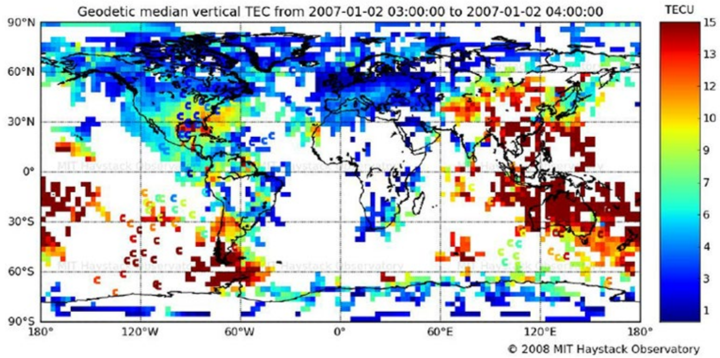
Method 2: Set up your own Madrigal site

- MIT Haystack will help with installation and loading programs
- You control when the data is uploaded
- Automated backup to CEDAR Madrigal site

What Kind of Data is Stored in Madrigal?



Magnetometer



GPS/GNSS TEC

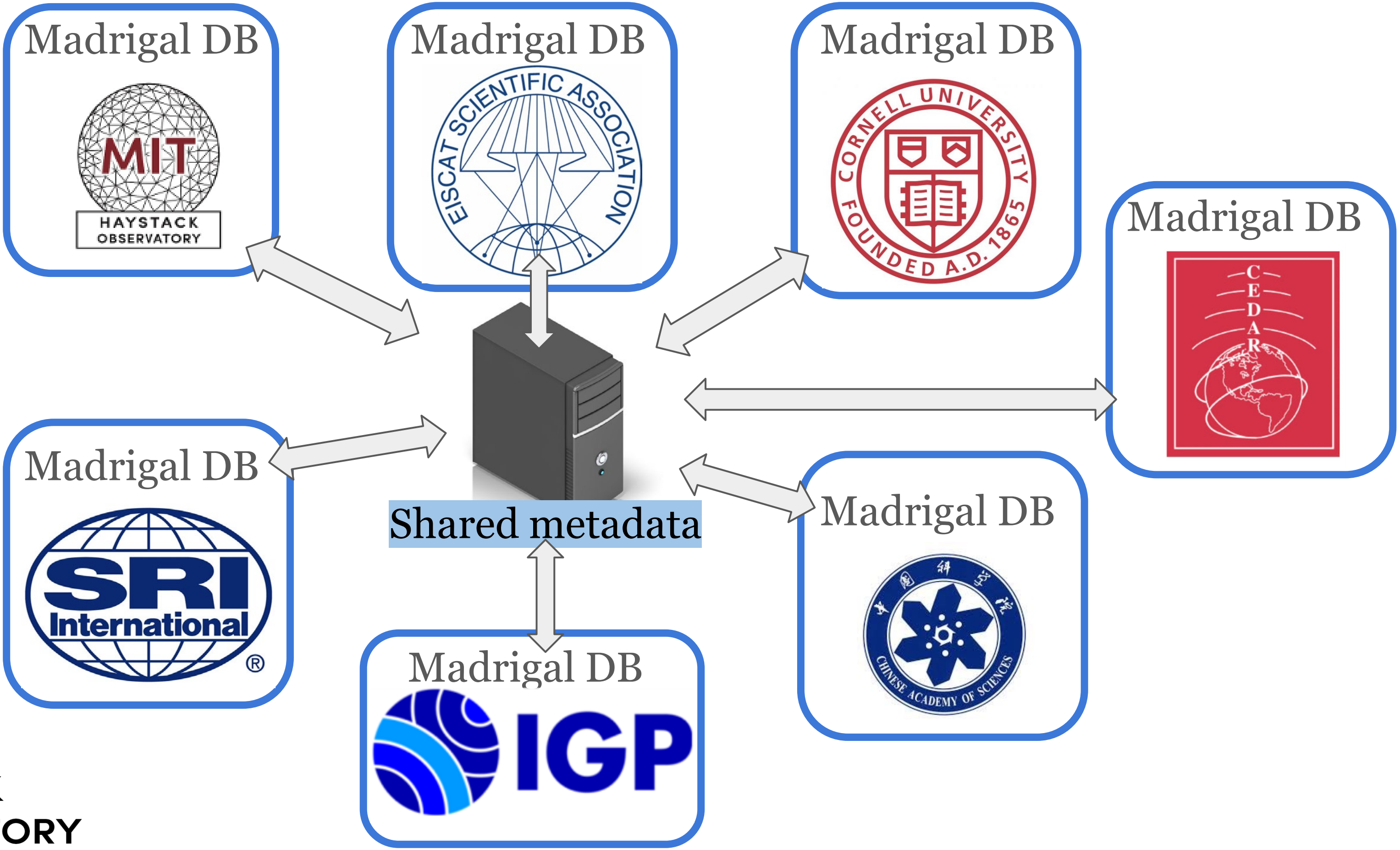
Over 200 Diverse Instruments in Madrigal, ~300 TB of data

Incoherent scatter radars (ISR): 28
Lower/middle atmosphere radars: 21
Photometers: 7
Fabry Perot Interferometers: 38
Michelson Interferometers: 6
Lidars: 10
Meteor radars: 18
GNSS Total Electron Content (TEC)

MIT
HAYSTACK
OBSERVATORY

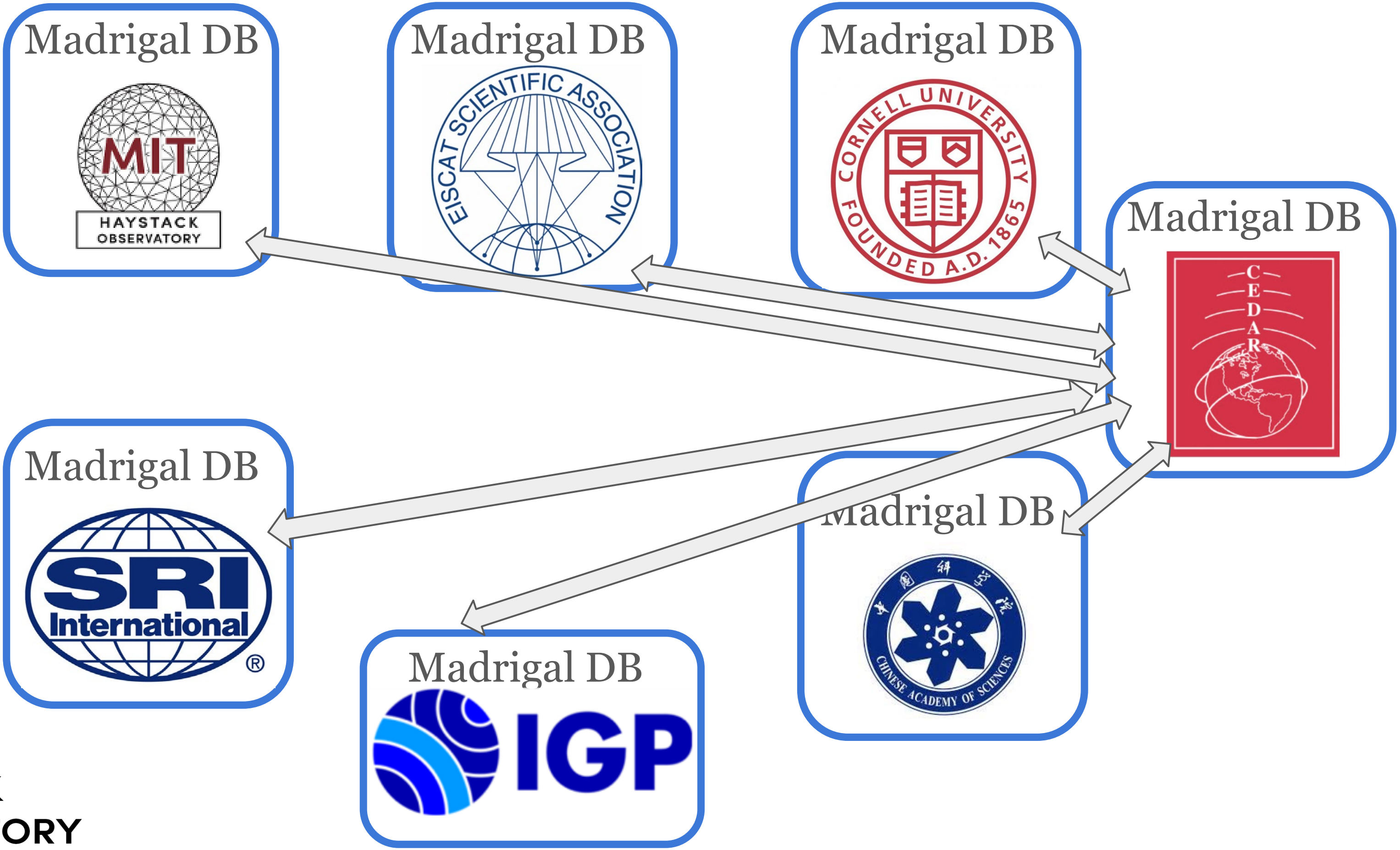
Easy to add new instruments too!

Madrigal: A Distributed Database

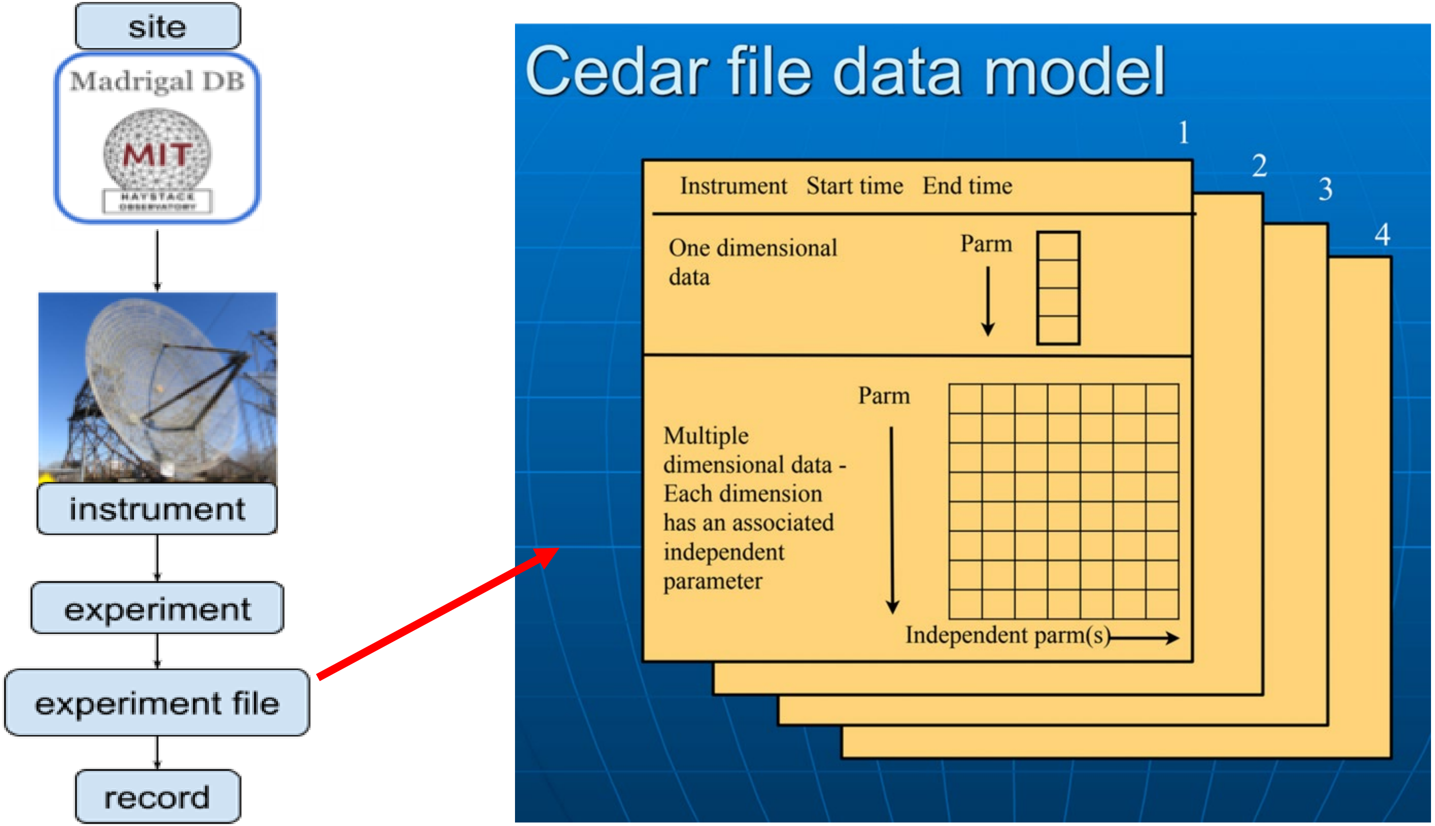


MIT
HAYSTACK
OBSERVATORY

CEDAR Madrigal Archive Imports All Metadata Weekly



The Madrigal Data Format and Metadata Model

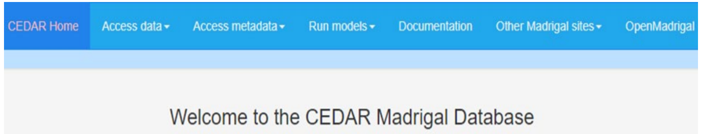


Data format designed to be **self describing**

Using Madrigal - Web or API Access

User

Web Interface



Typical use case: data discovery

Web Services API

- Python
- Matlab
- IDL

Create scripts using the MadrigalWeb API for all your data needs!

- Faster downloads compared to web interface
- Data flexibility - filter data by desired parameters/derived parameters

Thank you for listening!

Many thanks to the organizers of this workshop for supporting early career folks (such as myself!)

Any questions? Please contact cariglia@mit.edu

Summary

- Research campaigns need well-documented, standardized, science-ready data
- Data needs to be findable, accessible, citable, self-describing, easy to use
- More data means more storage and backups are needed
- Streamlined procedures for data provider upload / ingestion are key to ease data provision to the wider open science community
- The Madrigal Database is an example of robust technical infrastructure to store and serve real time and archival data for virtually any kind of data

Extra Slides

Madrigal Metadata Model

Madrigal site - a facility with scientists and a Madrigal installation

Instruments - ground based (set location) or satellites

Experiments - limited duration, single PI contact

Experiment files - data from one analysis of the experiment

Records - measurement over a single period of time

Madrigal DB

MIT HAYSTACK OBSERVATORY

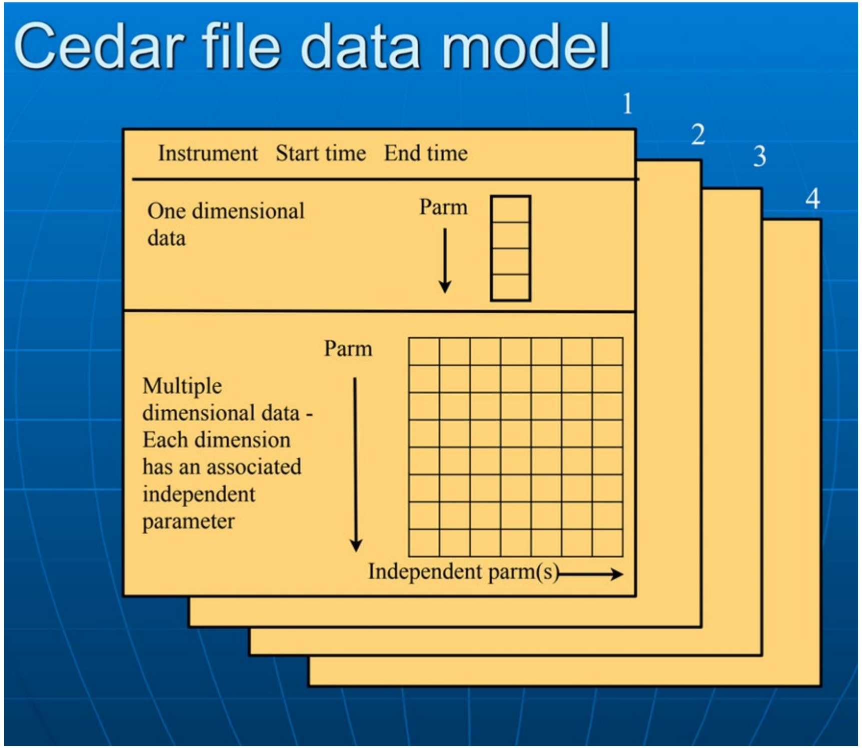
Data shared among all Madrigal sites

Data unique to one Madrigal site

MIT HAYSTACK OBSERVATORY

The CEDAR Madrigal Database Format

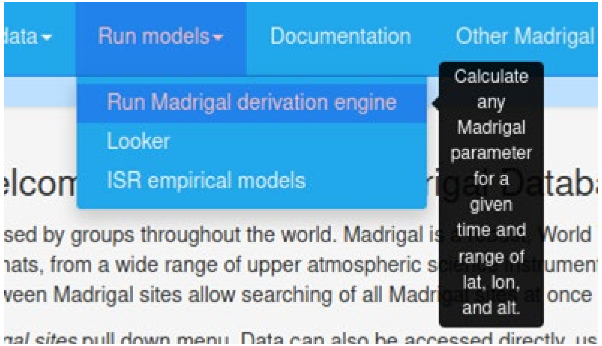
- Underlying data format is HDF5
- Well-defined parameters with standardized descriptions, units
- All parameters have corresponding uncertainty parameters
- Missing and assumed values included
- Well-defined parameter standards allow for the existence of a derivation engine, given a geographic location and datetime



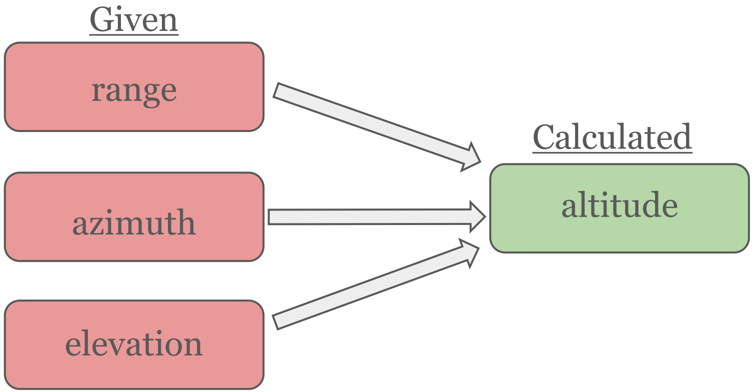
MIT
HAYSTACK
OBSERVATORY

Madrigal Derivation Engine

- Standardization of file parameters allows the existence of the derivation engine
- All parameters have corresponding uncertainty parameters
- Independent parameters built into standard
- Derived parameters appear in file
- Derivation engine determines which parameters can be derived
- New derived parameters (relatively) easy to implement in Python or C

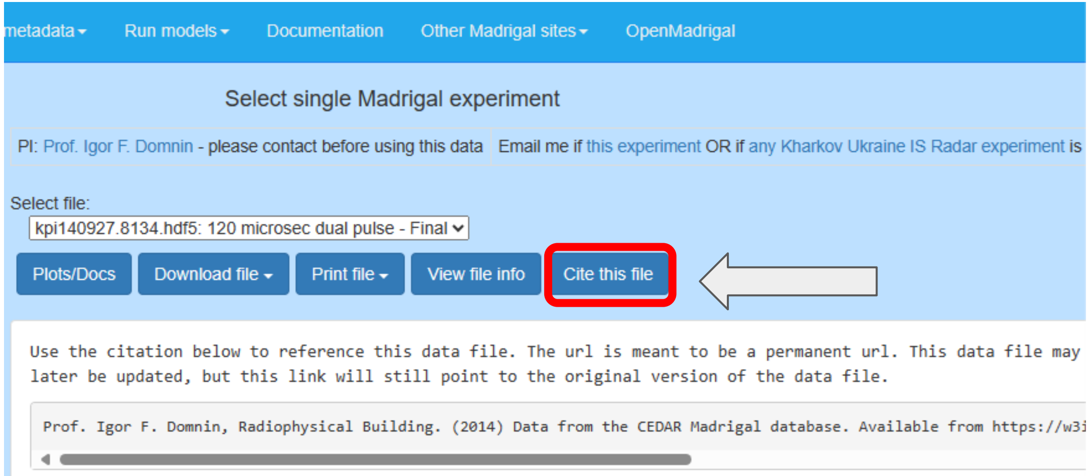


Derivation example for radar data:



Data Reproducibility

- Madrigal files are never deleted - when new versions of experiment files are created, old versions marked as “history”
- Every file has a citable URL
- Python API allows creation of a single citation for multiple files

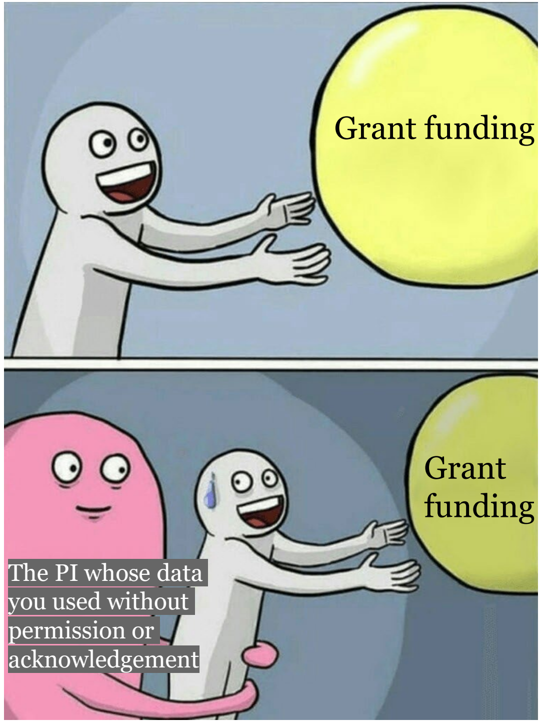


Citation generated:

Prof. Igor F. Domnin, Radiophysical Building. (2014) Data from the CEDAR Madrigal database. Available from https://w3id.org/cedar?experiment_list=experiments/2014/kpi/27sep14&file_list=kpi140927.8134.hdf5

Using Madrigal - Rules of the Road

- If you wish to use Madrigal data in a paper or talk, **please contact the PI**
- PI for every experiment listed on data download page



Use of data without informing the PI may lead to bad luck with grant writing-- don't let this happen to you!

MIT HAYSTACK OBSERVATORY

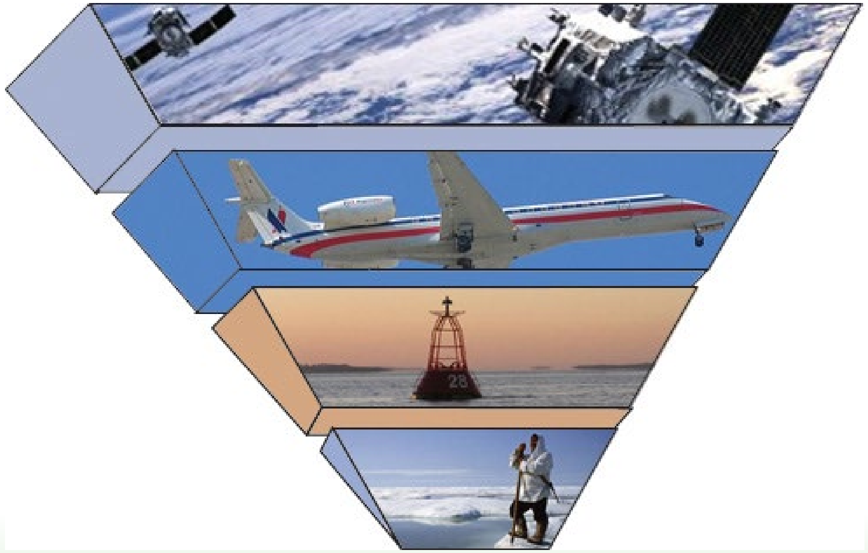
Community-observing Systems for the International Polar Year

Bering Sea Sub-Network (BSSN) / Community Observing Network for Adaptation & Security (CONAS)

- community observing for improving the understanding of environmental change in the Bering Sea
- 2007-2009; 2009-2014; 2014-2016

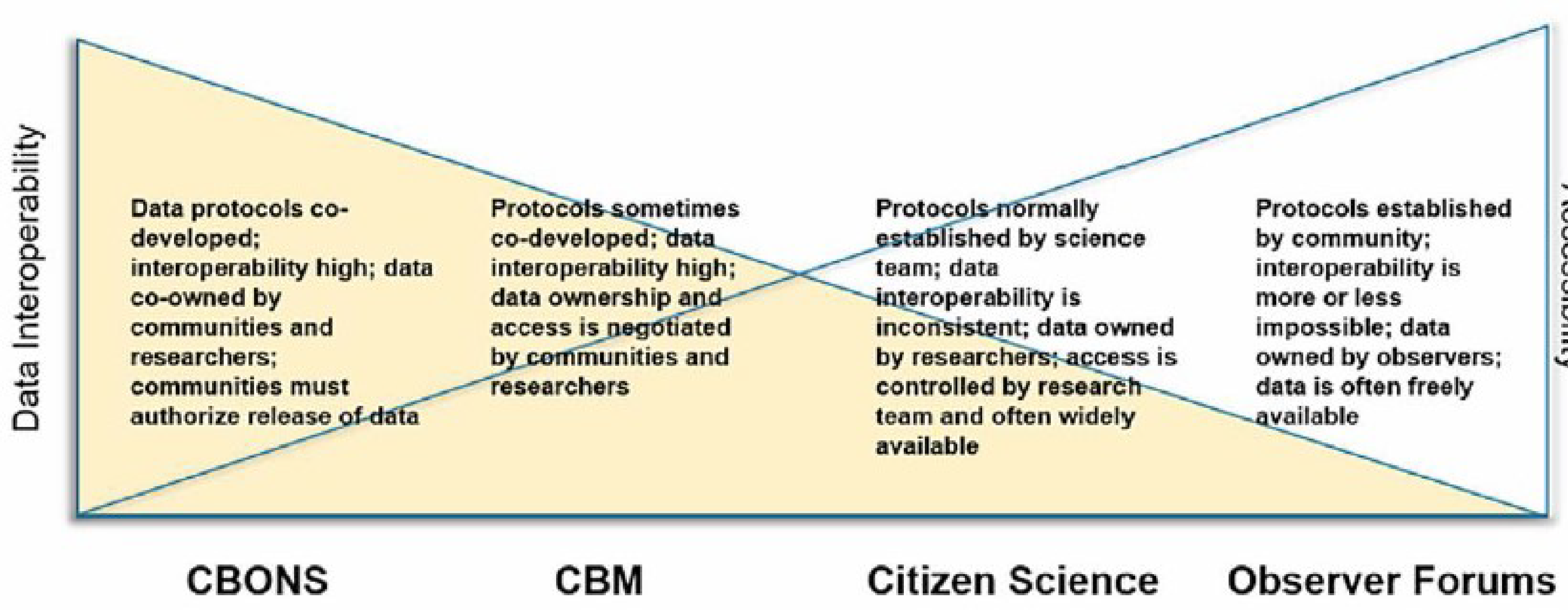
Community-based Observing Network Systems (CBONS)

- report social and environmental variables by place-based observers
- co-created by communities and researchers using interoperable formats
- global environmental change and its impacts on Indigenous communities in the Arctic
- local scale human sensor array … detecting change at the scale of daily human lives



Griffith & Alessa, 2018. Front Ecol Environ 16(S1)

Typology of Community-based Observing



Griffith & Alessa , 2017. NMIO Tech Bull

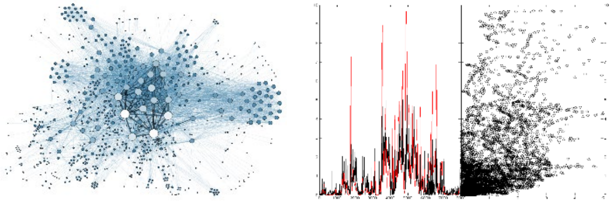
Community-based Observing Network & System for Situational Awareness (CBON-SA)

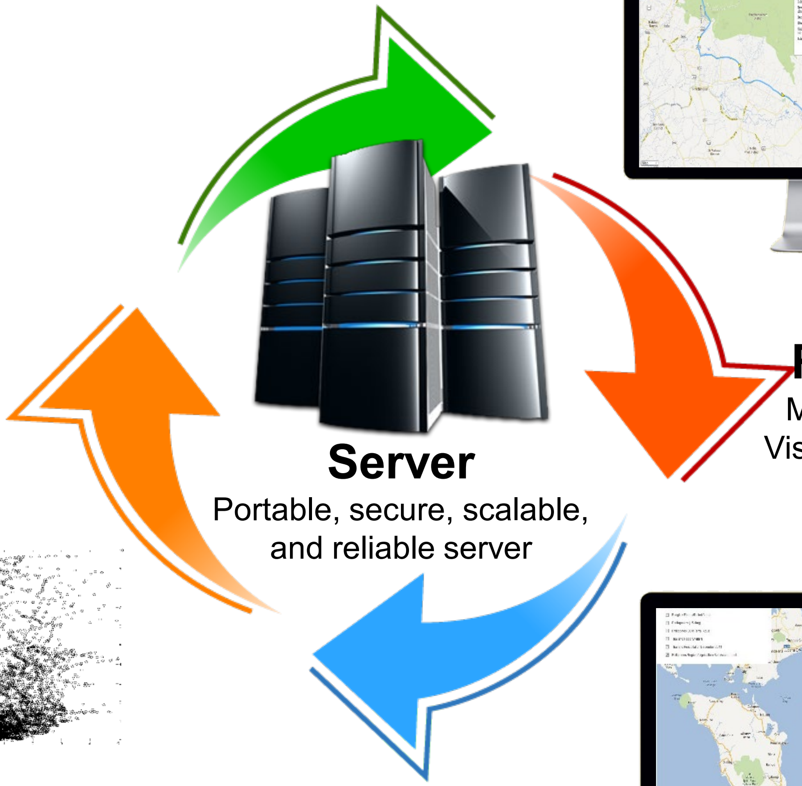
- distributed community observers to systematically observe and document Arctic environmental change
- community observers detect and place in context a range of critical variables pertinent to maritime security
- 2015-2018

CBON-SA Data Flows and Management

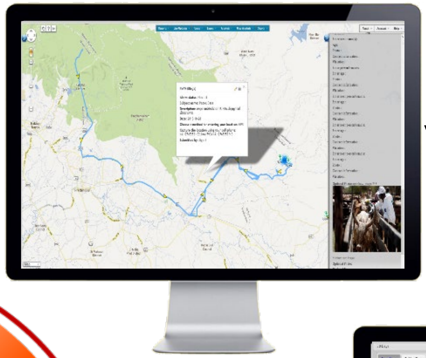
Field Report
Multi-media report in all connectivity situations

Analysis
Automated analysis options in various platforms



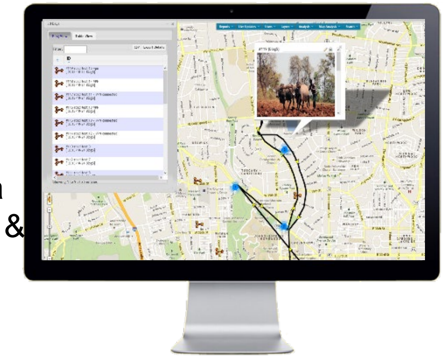


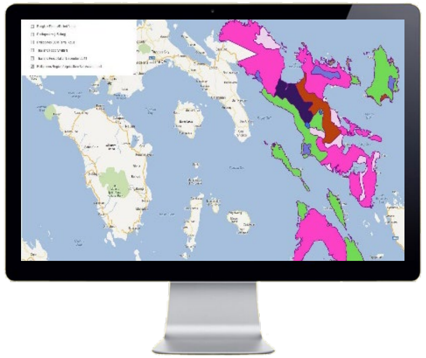
Server
Portable, secure, scalable, and reliable server



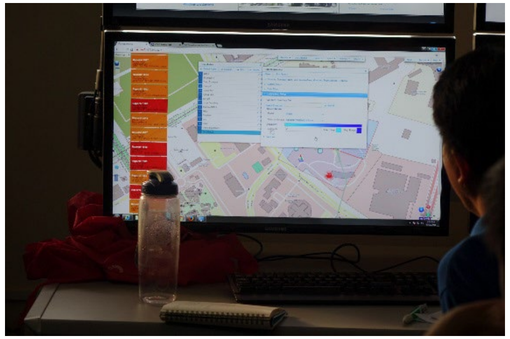
Portal Report
View report anywhere in the world

Portal Report
Multi-Media Visualization & Fusion





Portal Report
Share Information with partners & stakeholders



EyesNorth – An IPY4 initiated Research Coordination Network

- CBO as Collaborative Science for the IPY
- "science undertaken as a partnership … produces useable outcomes to the community"
- big data not always useable without social context …

A. Kliskey, 2008

EYES NORTH

EyesNorth – An IPY Research Coordination Network

- Building trust
- Building relationships
- Negotiating international / political borders
- CBO as a long-term endeavor

Aleut International Association, 2012