EarthCube Town Hall
AGU
December 9, 2019

This work supported through the National Science Foundation award #1928208.

Office Transition Sept-October 2019

Thanks to the EarthCube Science Support Office (ESSO)!

(Welcome back, Lynne Schreiber!)

ECO: Vision & Goals

EC Office that is effective, efficient, transparent, responsive

- Promotes clear communication to stakeholders.
- Ensure goals are reached and metrics tracked.
- Identify synergies between EC and related community activities.
- Translate governance and NSF directives into actions and outcomes.

Data Access and Discovery activity

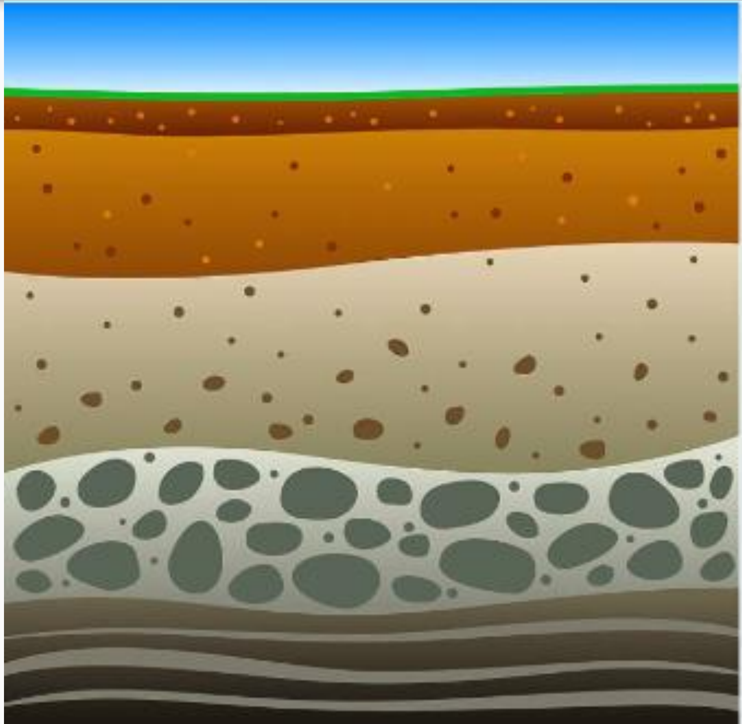
- Increased adoption and participation from the community.
- Effective facilitation of the emerging community-defined workbench.
- Increased adoption of EC outputs and integrations.
- Bridging science and technical (builders) for the community needs.

***Benefitting geoscientists** and the wider research community through*
tools, methods, standards, architectures, and community connections.

ECO Partners: ESIP, NCSA/UIUC, Ronin Institute, SDSC/UC San Diego, University of Hawaii, USGIN

ECO Team

- Strong links to Geosciences community
- Sustainable coordination mechanisms
- Trust, rapport, shared work experience
- People embedded in EarthCube, committed to vision
- Dedicated Executive Director



Core EC Office

Christine Kirkpatrick, Rebecca Koskela, Lynne Schreiber

Specific Attributes

- Full-time Executive Director
- FT Coordinator and PT support staff
- Shared staff for web, event management
- Experienced, dedicated grant and fiscal management staff

Data Community and Training

Lead: Erin Robinson

Supported by:
Megan Carter, Samantha Sands

Specific Activity Examples

- Data Fairs
- Data Help Desk at science meetings
- Data Challenges
- Training on demand, workshops
  - (GO) FAIR awareness, specific topics

2019 AGU Data FAIR, booth #1329

Planned Collaborations

Science Coordination

Cathy Constable, Ouida Meier, Karen Stocks



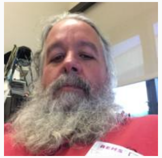
Specific Activities

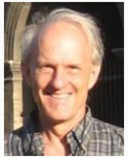
- Work with geoscientists to identify pressing technical needs and bottlenecks
- Coordinate w/ technical team; convey capab. & limit. of avail. solutions from user perspective
- Maintain a collection of use cases and sci-tech success stories across EarthCube
- Communication: Newsletter, Science features, EC & other Tool features, Workflows

Technical Team

Lead: Kenton McHenry

Supported by: Ilya Zaslavsky, David Valentine, Melissa Cragin, Steve Richard, Craig Willis




Specific Activities

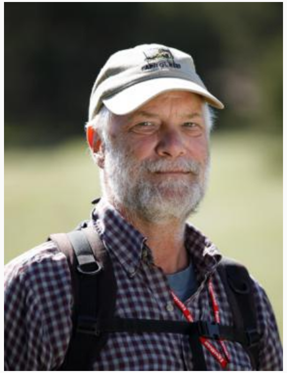
- Maintain common software support used across projects. Monitor and report EC resource re-use
- Support and extend GeoCODES, and EarthCube registries for data, components, and practices
- Integrated tool platform support, outreach, integration.
- Community standards and best practices: promote and build consensus
- Technical resource demonstration and training: demonstrate EC capabilities to geoscientists
- EC projects: identify and promote opportunities for interoperability, broker reuse of resources within and beyond EC. Review technical products for documentation and alignment with EC priorities.
- Strategic planning: work with committees, ad hoc WGs, surveys from users → technical recommendations

Leadership Council Chair Election

Vote December 9-20, 2019

Michael Daniels
Colorado State University
NCAR

Ken Rubin
Dept. of Earth Sciences
Univ. of Hawaii at Manoa



J. Douglas Walker
Department of Geology
University of Kansas

Eva E Zanzerkia - National Science Foundation

NSF Opportunities FY2020

Eva Zanzerkia GEO/EAR

Colleen Strawhacker GEO/OPP

Marc Stieglitz GEO/OPP

Amy Walton CISE/OAC

Eric DeWeaver GEO/AGS

Sean Kennan GEO/OCE

Cyberinfrastructure for the Geosciences-Opportunities

https://www.nsf.gov/geo/geo-ci/index.jsp

National Science Foundation
WHERE DISCOVERIES BEGIN

Contact | Help

Search

NSB | Research Areas | Funding | Awards | Document Library | News | About NSF

Geosciences (GEO)
- Geosciences (GEO) Home
  - About
  - Programs
  - Staff
  - Funding
  - Awards
  - News
  - Events
  - Additional Resources
- Atmospheric and Geospace

Home › Research Areas › Geosciences

Email Print Share

Cyberinfrastructure for the Geosciences – Opportunities

A variety of NSF opportunities support the development and implementation of cyberinfrastructure for the geosciences. A common goal of these opportunities is to increase public access to data generated through NSF-sponsored research. NSF's strategy for public access to research data is articulated in its Public Access Plan (NSF 15-52), and specific data policies are described in the Proposal & Award Policies & Procedures Guide (PAPPG, NSF 18-1). GEO Divisions and Offices specify additional data policies, and identify cyber resources available to support these policies (Directorate for Geosciences—Data Policies).

This website provides an overview of the range of NSF opportunities supporting cyberinfrastructure and data sharing in the geosciences. PIs are encouraged to reach out to cognizant Program Officers to learn more about relevant funding opportunities.

Upcoming Events

- EarthCube Town Hall at the AGU Fall Meeting - December 13, 2018, 12:30-1:30 PM EST, Washington, DC
- Webinar for EarthCube Office solicitation: Friday, January 4, 2019: 1 PM EST (see below for details)
- Webinar for EarthCube core solicitation: Friday, January 25, 2019: 1 PM EST (see below for details)

NSF Funding Opportunities

EarthCube NSF 20-520

Supplement Requests

- Science Adoption
  - Enhance existing science awards through adoption of tools/standards
- Data Resource Adoption
  - Adoption of data standards like GeoCODES by data facilities and resources

EarthCube RCNs

- Focus/Topic
  - community standards/data management
  - common CI & technology grand challenges
  - Areas convergent with Big Ideas
- EC Participation
- Steering Committee/Participants
- Outcomes

EarthCube NSF 20-520

Pilots

- Focus/Topic
  - Demonstration of lightweight and sustainable approaches to interoperability
  - Reuse of existing infrastructure, semantics and APIs
- Use scenarios that cross the geosciences
- Outcomes within 12-18 months

Science Enabling Data Capabilities

- Specific scientific challenge driven by geoscientists
- Reuse of existing tools/infrastructure
- Sustainability
- Metrics and assessment

Participation in EarthCube Governance

EarthCube NSF 20-520

Webinar: Monday, January 27, 2019: 1 PM EST

- EarthCube Research Coordination Networks (RCN)
  Target Date March 12, 2019

- EarthCube Science-Enabling Data Capabilities and Pilots
  Due Date March 12, 2019

Accelerator "Track A1":
HARNESSING THE DATA REVOLUTION

- Advanced science data infrastructure that is interoperable and has an open architecture (makes it easier to access and link heterogeneous data products)
- Open Knowledge Network – an open semantic information infrastructure to discover new knowledge from multiple disparate knowledge sources
- Create a nonproprietary shared knowledge infrastructure, with a particular focus on publicly available U.S. Government and similar public datasets. Challenges include underlying representation of facts, services that perform reasoning tasks, and secured access. Domains include geosciences, education, smart health, and manufacturing.

Track A1 - Clusters

Open Knowledge Network (21 projects)

Horizontal Projects

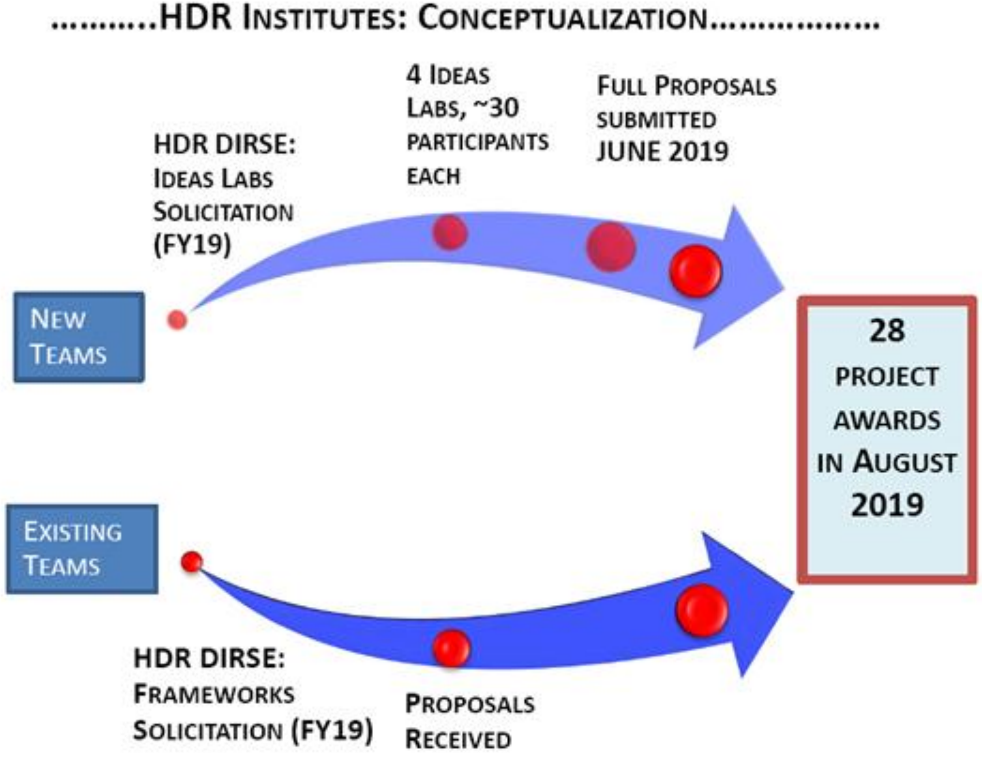
A7143 Information Credibility
A6940 Knowledge Application Development Environment
A6677 Spatial Data Models/Methods
A7136 Federated Search
A7165 Internet Structure & Security
A6731 Web Data Extraction/Integration
A7908 Spatial Decision Support

Vertical Projects

A7160 Precision Medicine
7043 Design & Manufacturing
A7099 Urban Flooding
A7152 Space Sciences
A7017 Molecular Data
A7095 Census
A7123 Court Records
A7115 Civil Infrastructure
A7137 Energy Systems
A7134 Intelligent Textbooks
A7033 Public Policy Data
A7153 Finance
A6884 Mobility
A6950 Ocean Resources

- Projects should seek "track integration"; are required to collaborate with industry; and encouraged to also collaborate / link with other relevant efforts in the community

HDR Institutes Roadmap



.....HDR Institutes: Convergence.....

<u>2021 Timeframe</u>

Open competition for 4-5 institutes that bring together multiple science and engineering communities

National Artificial Intelligence (AI) Research Institutes (NSF-20-503)

This program, a joint effort of NSF, USDA, NIFA, DHS, S&T, DOT, FHWA, and VA, seeks to enable such research through AI Research Institutes. This program solicitation describes two tracks: **Planning and Institute tracks**. Submissions to the **Planning Track** are encouraged in any areas of foundational and use-inspired research appropriate to NSF and its partner organizations. Proposals for the **Institute Track** must have a principal focus in one or more of the following themes, detailed in the Program Description under "Institute Track":

- Trustworthy AI
- Foundations of Machine Learning
- AI-Driven Innovation in Agriculture and the Food System
- AI-Augmented Learning
- AI for Accelerating Molecular Synthesis and Manufacturing
- AI for Discovery in Physics

NSF plans to make 1-6 Institute Awards, and ~ 8 Planning Grants.

Anticipated Funding Amount: $24,000,000 to $124,000,000

Proposal Deadlines: January 28, 2020 for Institute proposals

January 30, 2020 for Planning proposals

Ken Rubin – EarthCube Leadership Council

Ken Rubin, EarthCube Leadership Council Chair

The Leadership Council and other Governance Committees have been working for a couple of years now to refocus EarthCube efforts onto:

- Enabling and Enhancing Geoscientists' Research Workflows
- Promoting science-targeted project developments
- Promoting interoperability
- Promoting collaboration and interaction between funded projects
- Promoting and Educating about Data Stewardship Practices

Ken Rubin, EarthCube Leadership Council Chair

Learn more this week about our efforts on the aforementioned initiatives

- ESIP/AGU Data FAIR, **tue 13:00-14:30** “Learn How EarthCube Plans to Help Scientists Make Full Data Pipelines FAIR & Give Input”
- Talk on **wed 10:50 AM, IN32A-03** “Geoscientific Data Integration Occurs at the Workflow Layer, Requiring FAIR for all Aspects of an Analysis Workflow and More Data Repository Integration: an EarthCube view”
- https://www.earthcube.org/info/about/earthcube-governance
- Also, see the “OneGeochemistry” Townhall (Lehnert/Wyborn) **tomorrow at 12:30PM**, to participate in a developing community of practice on these topics

Ken Rubin, EarthCube Leadership Council Chair

This year, we've opted to forgo the AGU Townhall Governance update to bring you some recent project highlights instead:

Bonnie L Hurwitz - Planet Microbe
Joseph Hamman - Pangeo
Suzanne A Pierce - Is Geo (Intelligent Systems)
Alexander G Kosovichev - Heliophysics
Samira Daneshgar Asl - ICEBERG
Basil Tikoff - StraboSpot

Bonnie L Hurwitz - Planet Microbe

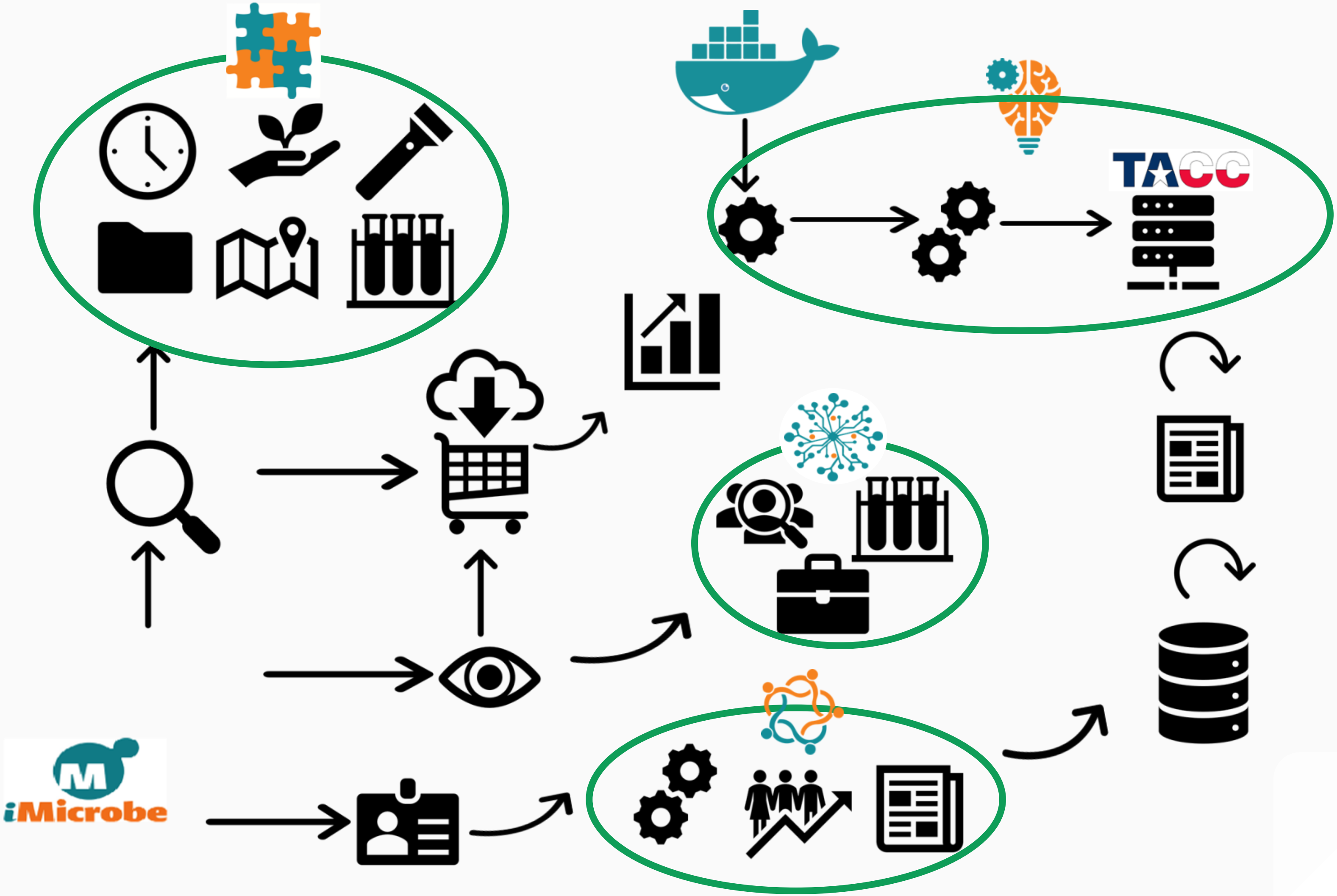
PlanetMicrobe

Bonnie Hurwitz, University of Arizona

Global ocean ‘omics datasets are spread across diverse projects, with varied metadata and semantics


PlanetMicrobe
Project Overview
Reintegration of 'omics data with their environmental context
Standardization of semantics for increased data interoperability
Providing community driven analysis and visualizations tools
iMicrobe
Data Repository

Mind Map



Data containers for dataset interoperability

Planet Microbe uses frictionless data packages to make datasets interoperable, as well as shareable between systems.

FRICTIONLESS DATA

CSV

environmental metadata from select samples

JSON

type, unit and source information about the metadata

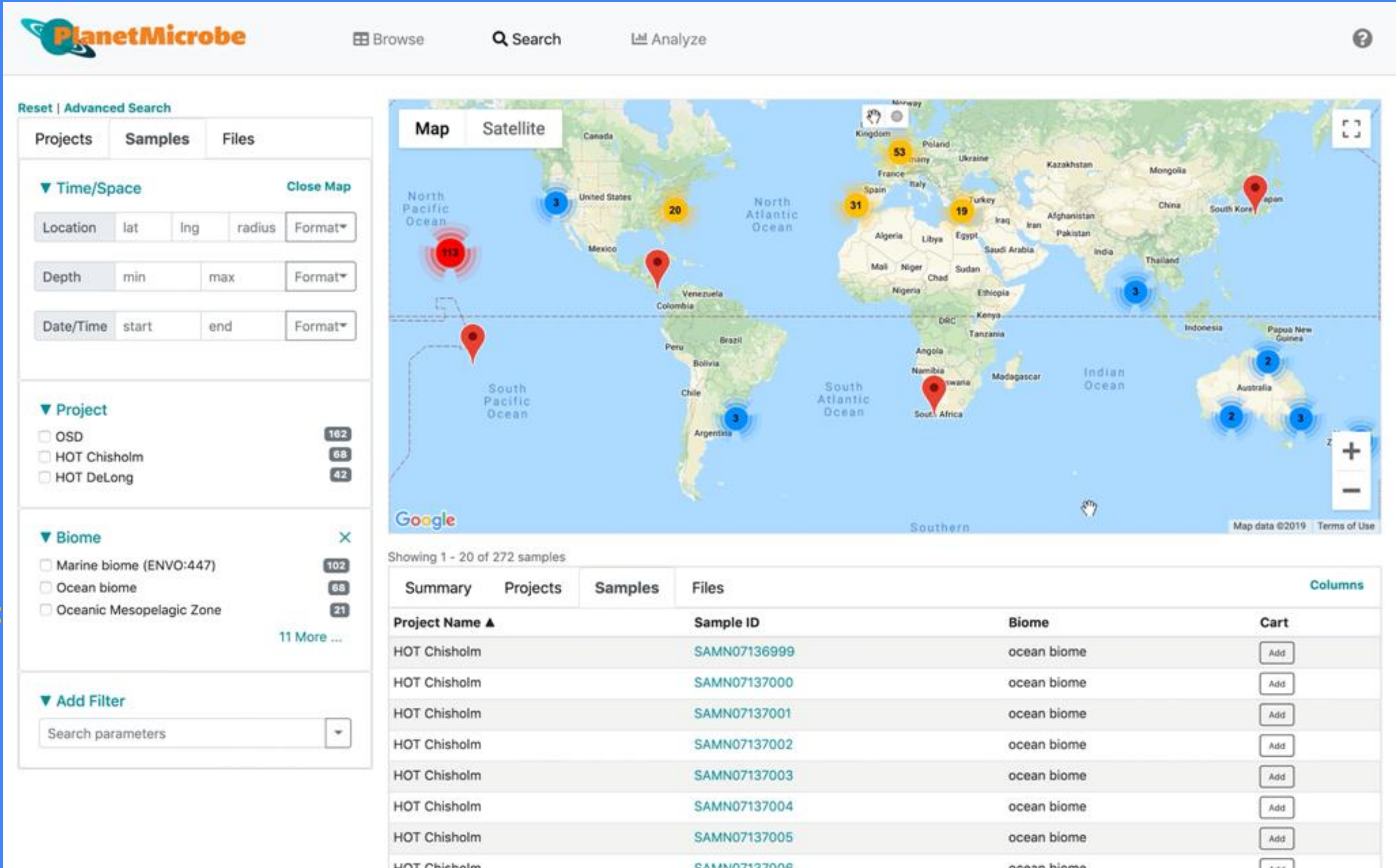
Specifications

Lightweight containerisation formats for data that provide a minimal yet powerful foundation for data publication, transport, and consumption.

Software

Apps and integrations that make it easy to integrate Frictionless Data specifications into your data publication, access, and analysis workflows.

Map & Sample Search

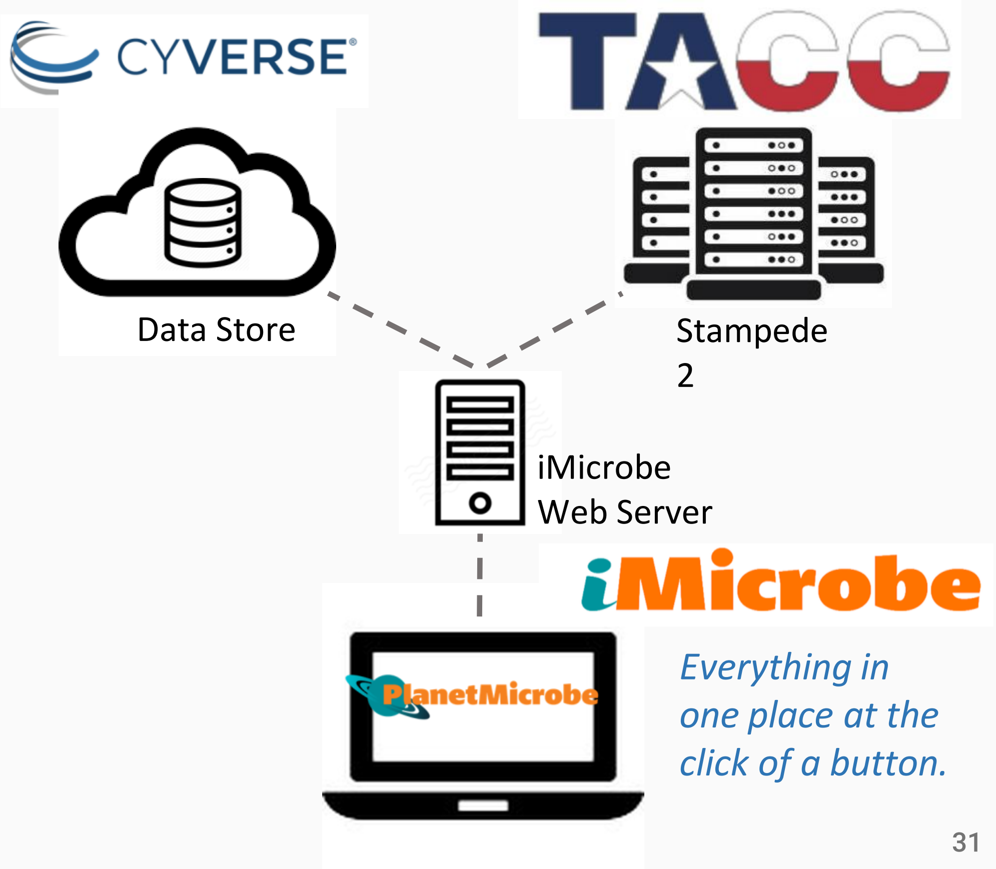


Dynamic range-based queries

Scalable Analytics

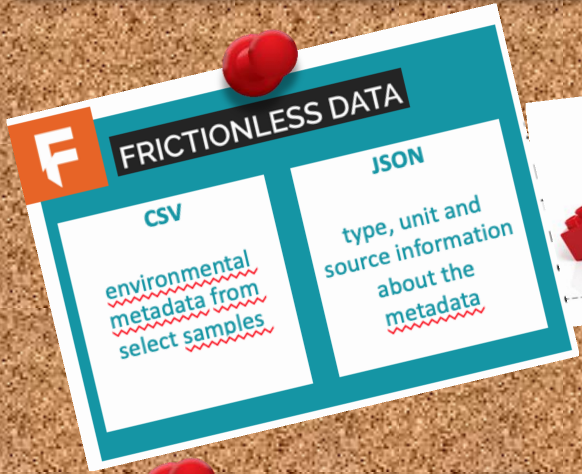
Cloud Architecture

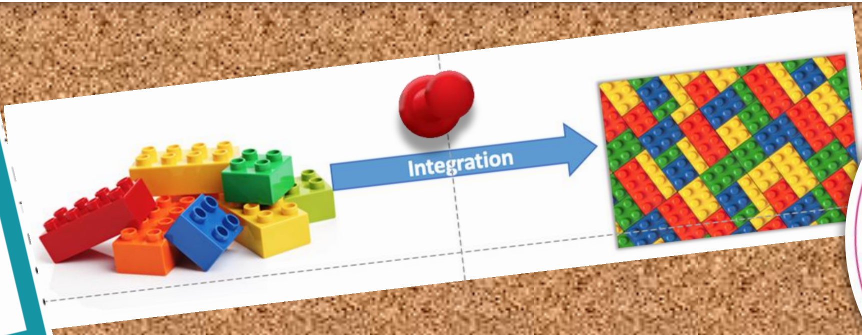
- Leverages CyVerse & TACC
  - Security
  - Scalability
  - Lower cost of development/maintenance
- Apps (tool containerization)
  - Docker & Singularity
  - Build once, run everywhere



Everything in one place at the click of a button.

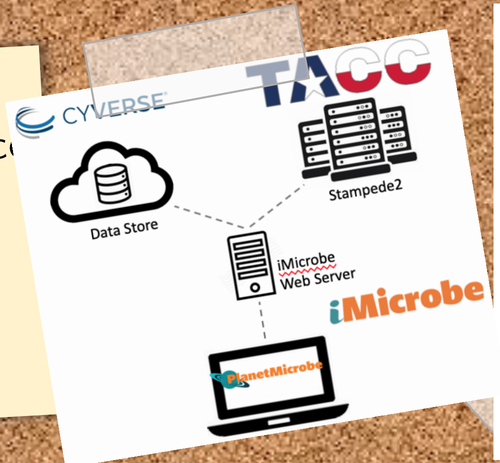
PlanetMicrobe Promoting Open Science

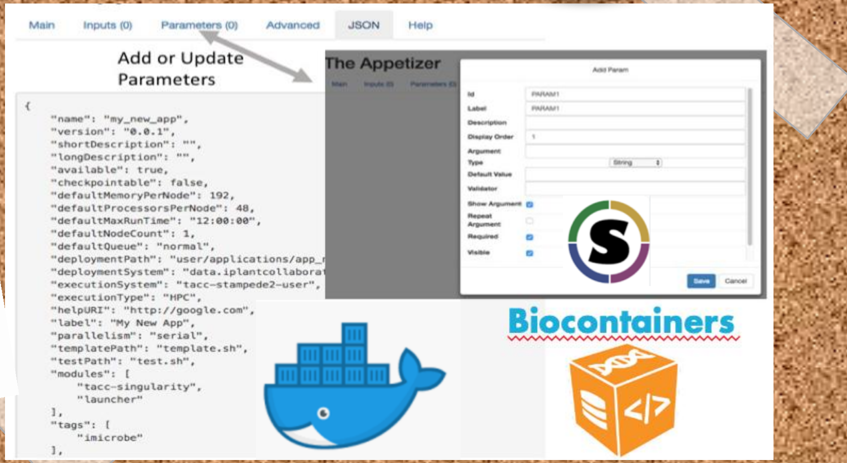




open and integrative data science systems

- ✔ Standards
- ✔ Quality
- ✔ Integration
- ✔ Access





PlanetMicrobe

Alise Ponsero, PhD, MS
Post-Doctoral Fellow

Matt Bomhoff
Sr. Developer

Ken Youens-Clark, MS
Sr. Developer

Kai Blumberg
Graduate Student

University of Hawaii
Dr. Ed Delong

Lawrence Berkeley National Lab
Dr. Elisha Wood-Charleson

@hurwitzlab @planetmicrobe

Demo: Earthcube Booth
Tuesday 2:30-3:30pm

Joe Hammam - Pangeo

Pangeo and Jupyter Meets Earth Project Spotlight

Joe Hamman, NCAR

PANGEO

AN OPEN SOURCE BIG DATA CLIMATE SCIENCE PLATFORM

RESEARCH USE CASES

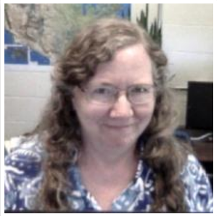
- Atmospheric Moisture Budgets
- Convective Parameters for Understanding Thunderstorms
- Climate downscaling for hydrologic modeling
- Statistical analysis of eddy-resolving ocean models

TECH DEVELOPMENT

- Integrate existing scientific python ecosystem
- Deploy system on HPC and Cloud Computing Systems
- Benchmark and tune based on real-world workflows
- Document and demonstrate application

Ryan Abernathey

Chiara Lepore



Naomi Henderson

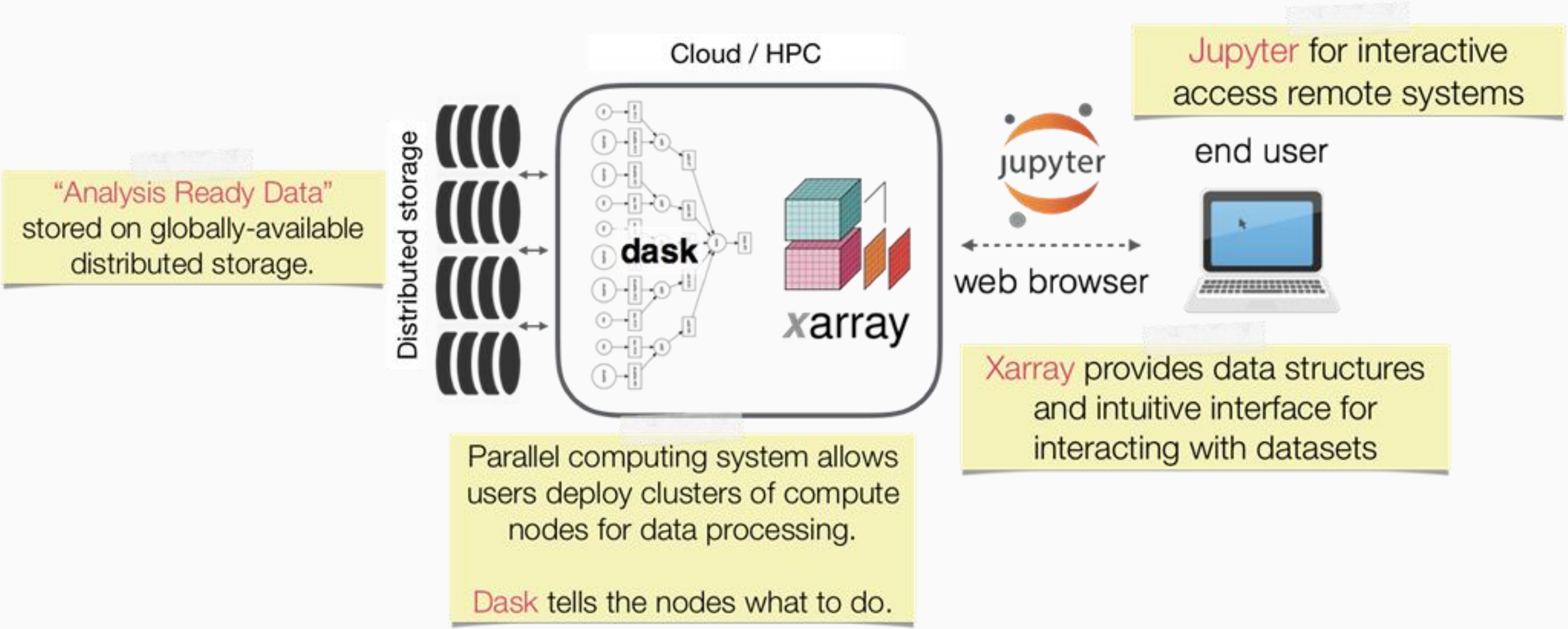
Kevin Paul

Joe Hamman

Ryan May

Matt Rocklin

PANGEO ARCHITECTURE



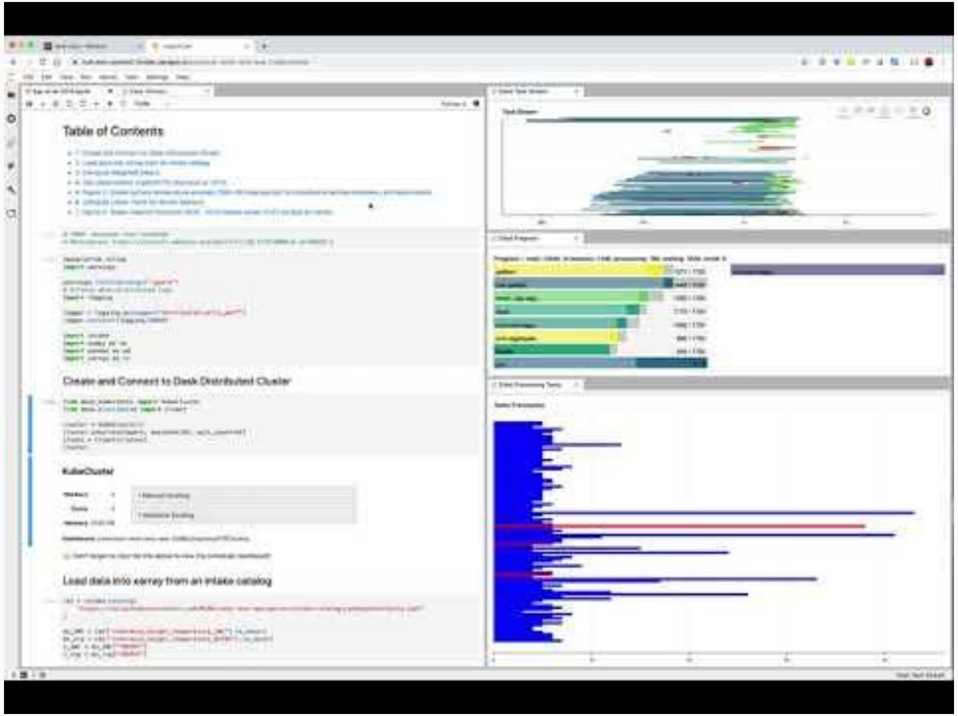
INTEGRATION OF THE SCIENTIFIC PYTHON ECOSYSTEM

HIGHLIGHT 1:
DISTRIBUTED COMPUTING WITH XARRAY AND DASK

Xarray: Labeled N-Dimensional Arrays in Python
Dask: Parallel, distributed arrays in Python

- Xarray supports using Dask arrays under the hood
- Integration with Dask-distributed for multi-process reads and write with large computer clusters

The Result: massively scalable data analysis on very large multidimensional datasets



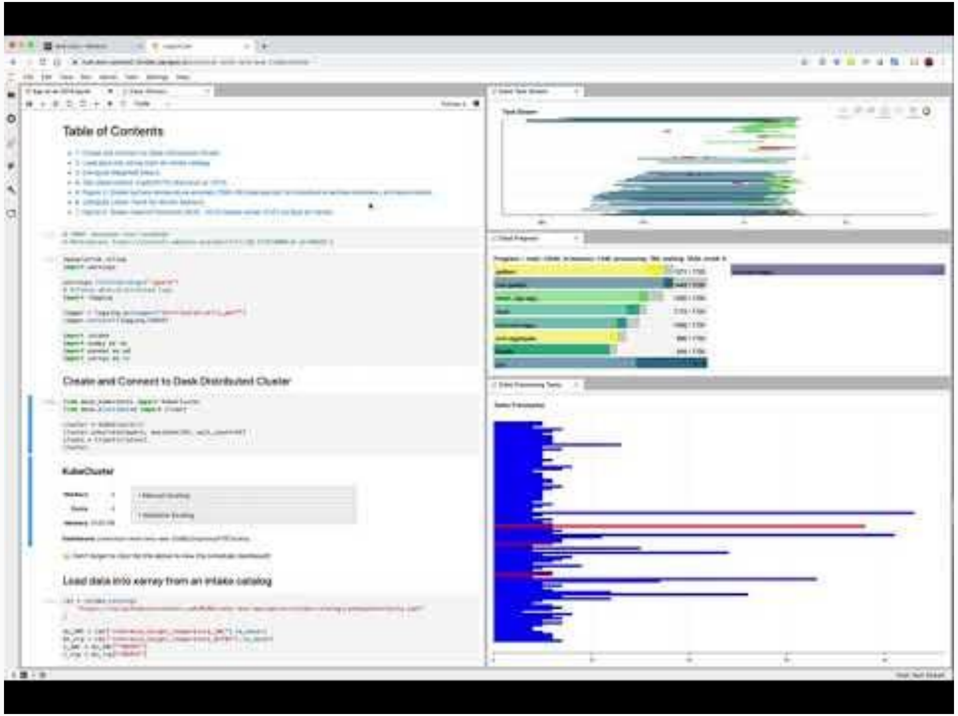
HIGHLIGHT 2:
ACTIVE DEPLOYMENTS ON HPC AND CLOUD PLATFORMS

HPC

- JupyterHub on Cheyenne Supercomputer (https://jupyterhub.ucar.edu/)

Cloud

- JupyterHub deployments on all major cloud providers
  - e.g. hub.pangeo.io, binder.pangeo.io
- New libraries for distributed computing using Dask with Kubernetes
- Integration of Jupyter and Dask software tools (e.g. dask-labextension)



HIGHLIGHT 3: NEW CLOUD OPTIMIZED STORAGE FORMAT FOR NETCDF DATA

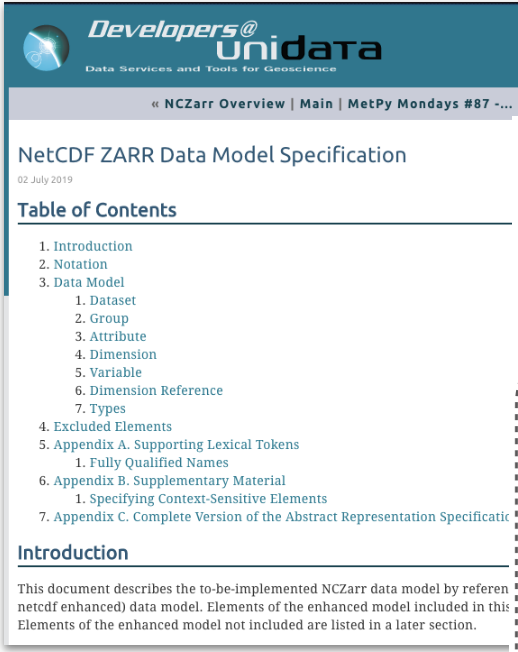
We needed a “cloud optimized NetCDF format”.

- Zarr provides chunked compressed, N-dimensional arrays amenable to cloud object storage
- We added support for Zarr in Xarray, effectively creating the first version of cloud-ready NetCDF

```
import xarray as xr

ds = xr.open_dataset('data.nc')  # NetCDF
# ...
ds.to_zarr('data.zarr')  # Write Zarr
ds_zarr = ds.open_zarr('data.zarr')  # Read Zarr
```

- Unidata has adopted this approach. Beta version of NetCDF support for Zarr is imminent
- For more info, Go see Ward Fisher’s poster on Wednesday (IN31B-0800):



Developers@ Unidata
Data Services and Tools for Geoscience
Login | Register
« NCZarr Overview | Main | MetPy Mondays #87 -... »

NetCDF ZARR Data Model Specification

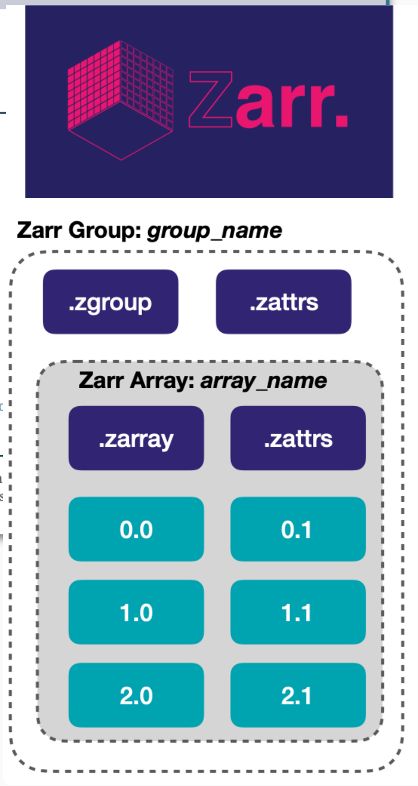
02 July 2019

Table of Contents

1. Introduction
2. Notation
3. Data Model
   1. Dataset
   2. Group
   3. Attribute
   4. Dimension
   5. Variable
   6. Dimension Reference
   7. Types
4. Excluded Elements
5. Appendix A. Supporting Lexical Tokens
   1. Fully Qualified Names
6. Appendix B. Supplementary Material
   1. Specifying Context-Sensitive Elements
7. Appendix C. Complete Version of the Abstract Representation Specificatic

Introduction

This document describes the to-be-implemented NCZarr data model by referen
netcdf enhanced) data model. Elements of the enhanced model included in this
Elements of the enhanced model not included are listed in a later section.



HIGHLIGHT 4: RAPID GROWTH OF THE PANGEO COMMUNITY

Community participation

- Individuals across academia and industry
- New science domains like astronomy and biology

Working Groups

- Data
- Cloud DevOps
- Machine Learning
- Education and Outreach

Ways to get involved

- Weekly community telecon (https://pangeo.io/meeting-notes.html)
- Workshops and conferences (https://pangeo.io/meetings/index.html)
- Online Discourse forum (https://discourse.pangeo.io/)

Left: 2019 Community Meeting & Hackathon (Seattle)

Right: 2018 Developers Meeting (Boulder)

https://medium.com/pangeo/pangeo2019-17f1a2a016e0

jupyter MEETS THE EARTH

ENABLING DISCOVERY IN GEOSCIENCE THROUGH INTERACTIVE COMPUTING AT SCALE

RESEARCH USE CASES

- CMIP6 Climate data analysis
- Large scale hydrological modelling
- Geophysical simulations and inversions

jupyter TECH DEVELOPMENT

- Data discovery through JupyterLab
- Interactivity: Widgets & Dashboards
- JupyterHub: Using and managing shared computational infrastructure

Fernando Perez

Joe Hamman

Laurel Larsen

Kevin Paul

Lindsey Heagy

Chris Holdgraf

Yuvi Panda

Suzanne A Pierce - IS Geo

What is the IS-GEO RCN?

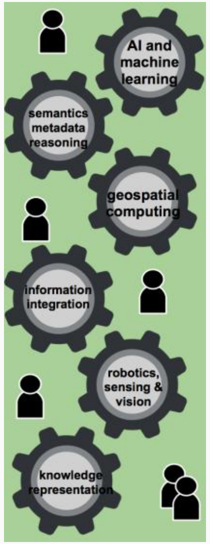
Integrating AI with Earth Sciences to drive discoveries

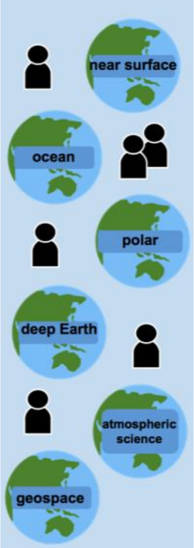
EARTHCUBE
TRANSFORMING GEOSCIENCES RESEARCH

@isgeo.community | spierce@tacc.utexas.edu | www.is-geo.org

(Kaplan, 2016)

What is the IS-GEO RCN?





Integrating AI with Earth Sciences to drive discoveries

EarthCube
TRANSFORMING GEOSCIENCES RESEARCH

@isgeo.community | spierce@tacc.utexas.edu | www.is-geo.org

(Kaplan, 2016)

What is Artificial Intelligence?

It’s much more than Machine Learning

EarthCube
TRANSFORMING GEOSCIENCES RESEARCH

What is Artificial Intelligence?

Because this is the American Geophysical Union…

"The essence of AI

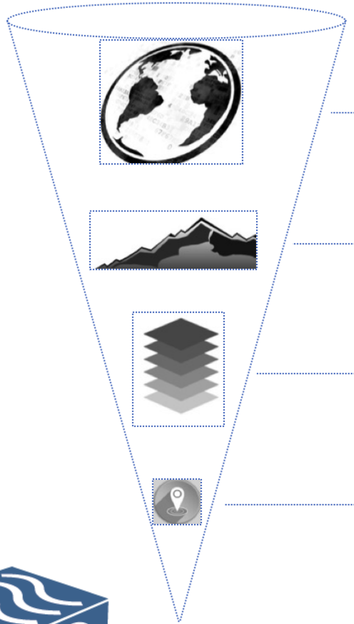
– indeed the essence of Intelligence –

Is the ability to make appropriate generalizations in a timely fashion based on limited data."

(Kaplan, 2016)

EARTHCUBE
TRANSFORMING GEOSCIENCES RESEARCH

Beyond ML, meaningful AI in GEO includes…



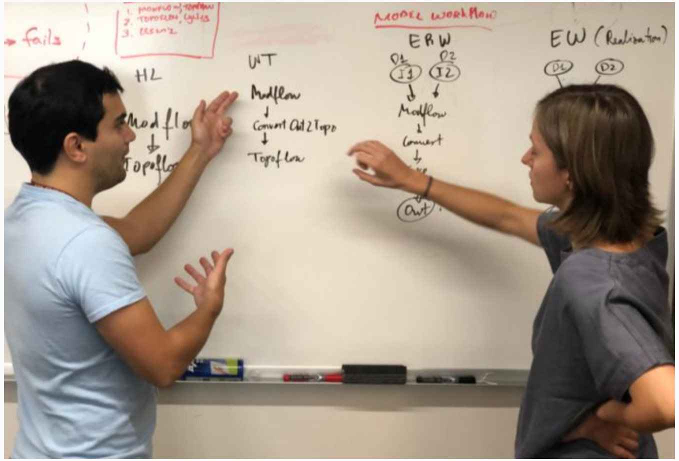
Interactive Workspaces
- Synthesis studies
- Wholistic approaches
- High-level research questions

Theory-Driven Learning
- Large but minuscule datasets
- Extreme events
- Multiple scales

Information Ecosystems
- Many locations & data types
- Georeferenced & curated data
- Discover interlinked events

Model-Driven Sensing
- Selective data collection
- Inaccessible locations
- Optimize experimental design

EARTHCUBE
TRANSFORMING GEOSCIENCES RESEARCH



- **Integrated Intelligence,**
- **Meaningful Interaction,**
- **Self-Aware Learning**

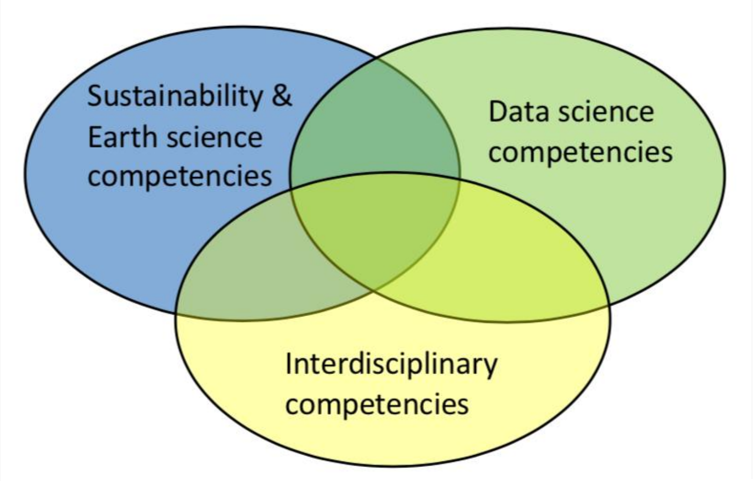
(Gil et al, 2019 ; AAAI, 2019; Gil and Pierce (Eds), 2015)

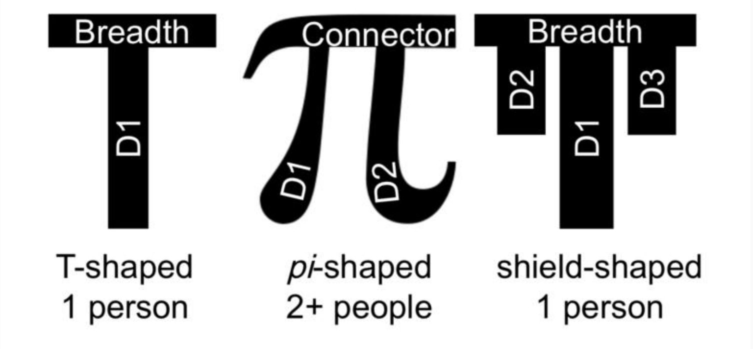
What enables IS-GEO?

Hands-On Workshops | SME Lectures | Digital Tools

To advance IS-GEO the community needs:

- Access to Data needed - Open Repositories
- SME knowledge for deep reasoning
- Digital Sandboxes to test solutions
- AI and GEO research platforms
- Well-posed science questions





(from Pennington et al, 2019)

IS-GEO offers hands-on collaborations

Team: Hack Fast – Fake News, Boulder, CO

We are a community exploring & implementing end-to-end applications!

EARTHCUBE
TRANSFORMING GEOSCIENCES RESEARCH

What enables IS-GEO?

'All the instances of scientific development and practice . . . are as much embedded in politics and cultures as they are creations of the researchers, practitioners, and industries.'

EarthCube
TRANSFORMING GEOSCIENCES RESEARCH

(Paraphrased from Heymann, 2010)

What enables IS-GEO?

The emerging community of researchers in intelligent systems (IS) and Geosciences (GEO).

(just a few of the cats and dogs ☺ engaging in IS-GEO discussions)

Intelligent Systems and Geosciences Research Coordination Network

National Science Foundation [award number 1632211]

References

AAAI, 2019, A 20-Year Roadmap for AI Research, Honolulu, Hawaii from Townhall, https://aaai.org/Conferences/AAAI-19/townhall-a-20-year-roadmap-for-ai-research/

Kaplan, J., 2016, Artificial Intelligence: What Everyone Needs to Know, Oxford University Press, 165 p.

Gil, Y., Pierce, S.A. (Eds), 2015, *Final Report on the 2015 NSF Workshop on Information and Intelligent Systems for Geosciences*. National Science Foundation Workshop Report, October 2015.

Pennington, D, Ebert-Uphoff, I, Martin, J., Freed, N., Pierce, S.A., 2019, Data Science Education, Sustainability Sciences Journal,

EarthCube
TRANSFORMING GEOSCIENCES RESEARCH

Thank you!

@isgeo.community | spierce@tacc.utexas.edu | www.is-geo.org

Alexander G Kosovichev - Heliophysics

EarthCube Data Infrastructure: Intelligent Databases and Analysis Tools for Geospace Data

Alexander Kosovichev
Gelu Nita, Vincent Oria, Viacheslav Sadykov
Center for Computational Heliophysics
New Jersey Institute of Technology

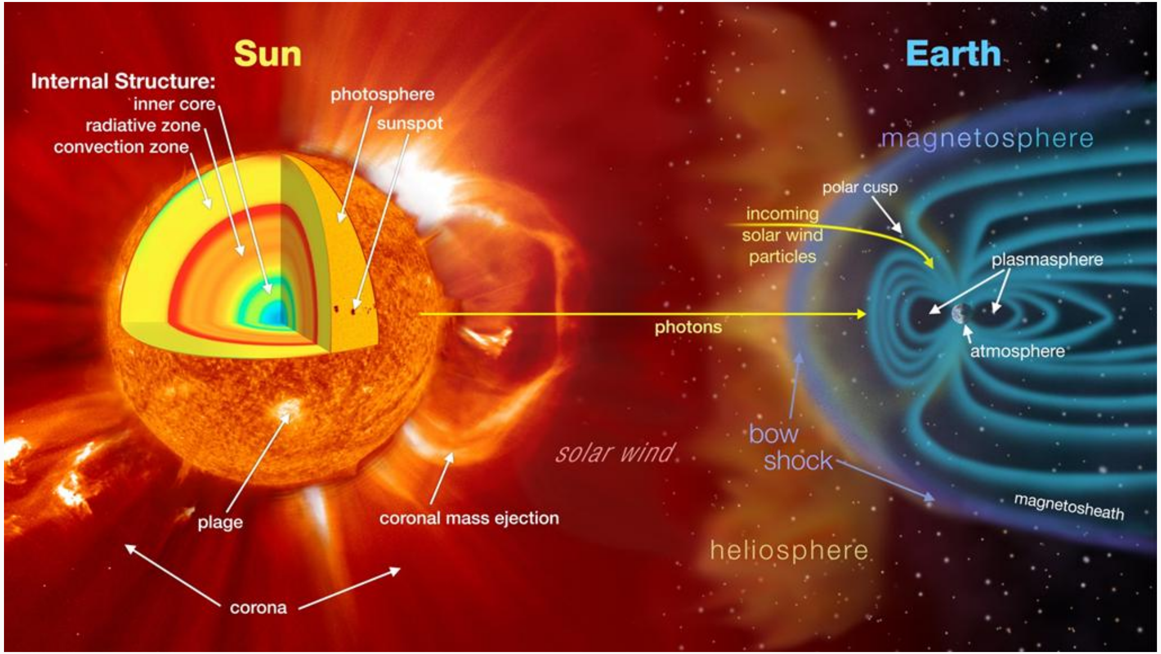
https://www.earthcube.org/group/intelligent-databases-analysis-tools-geospace-data

The project goals:

1) improve access of geoscientists to the geospace data and modeling products that are mostly relevant to studying the impacts of solar activity on Earth systems and to developing predictive capabilities
2) develop approaches, provisional tools, and working examples of database of extreme solar events by integrating observational data from several spacecraft and ground-based
3) establish close collaboration and partnership between solar physicists and computer scientists in the framework of the New Jersey Institute of Technology (NJIT) Center for Computational Heliophysics (CCH), the NASA Earth Exchange (NEX), and the Heliophysics Modeling and Simulations group at the NASA Ames Research Center; use the open NJIT and NASA facilities for on-line access and sustainability of the developed databases and tools
3) promote collaboration between geophysics and geospace scientists
4) involve students and young scientists in EarthCube Geospace activities.

Motivation

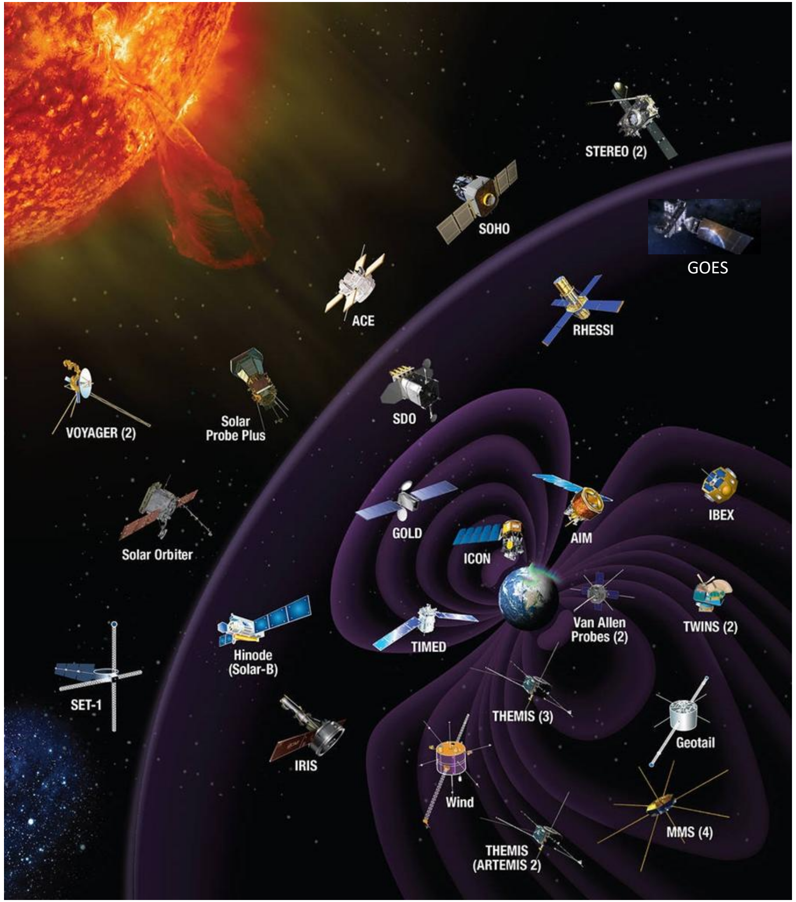
Solar activity plays especially important role in the Solar-Terrestrial interactions: it a primary source of global changes of the Earth atmosphere, powerful geomagnetic storms, high-energy radiation and particles that affect the Earth’s space environment, technological and biological systems.



Focus on Understanding Multi-Wavelength Nature of Solar Flares and Eruptions

The solar flares and accompanying events (CMEs and SEPs) greatly impact the Space Weather and terrestrial environment:

- Destroy satellites equipment and disrupt power grids
- Affect radio communications and GPS navigation
- Provide potential danger for space exploration (especially in the perspective of NASA's Moon-by-2024 program)
- Nowadays the solar flares are observed in multiple wavelengths by many ground-based and space-based facilities.
- For understanding of the flares it is necessary to perform a combined multi-wavelength analysis and classify large amounts of scientific data.



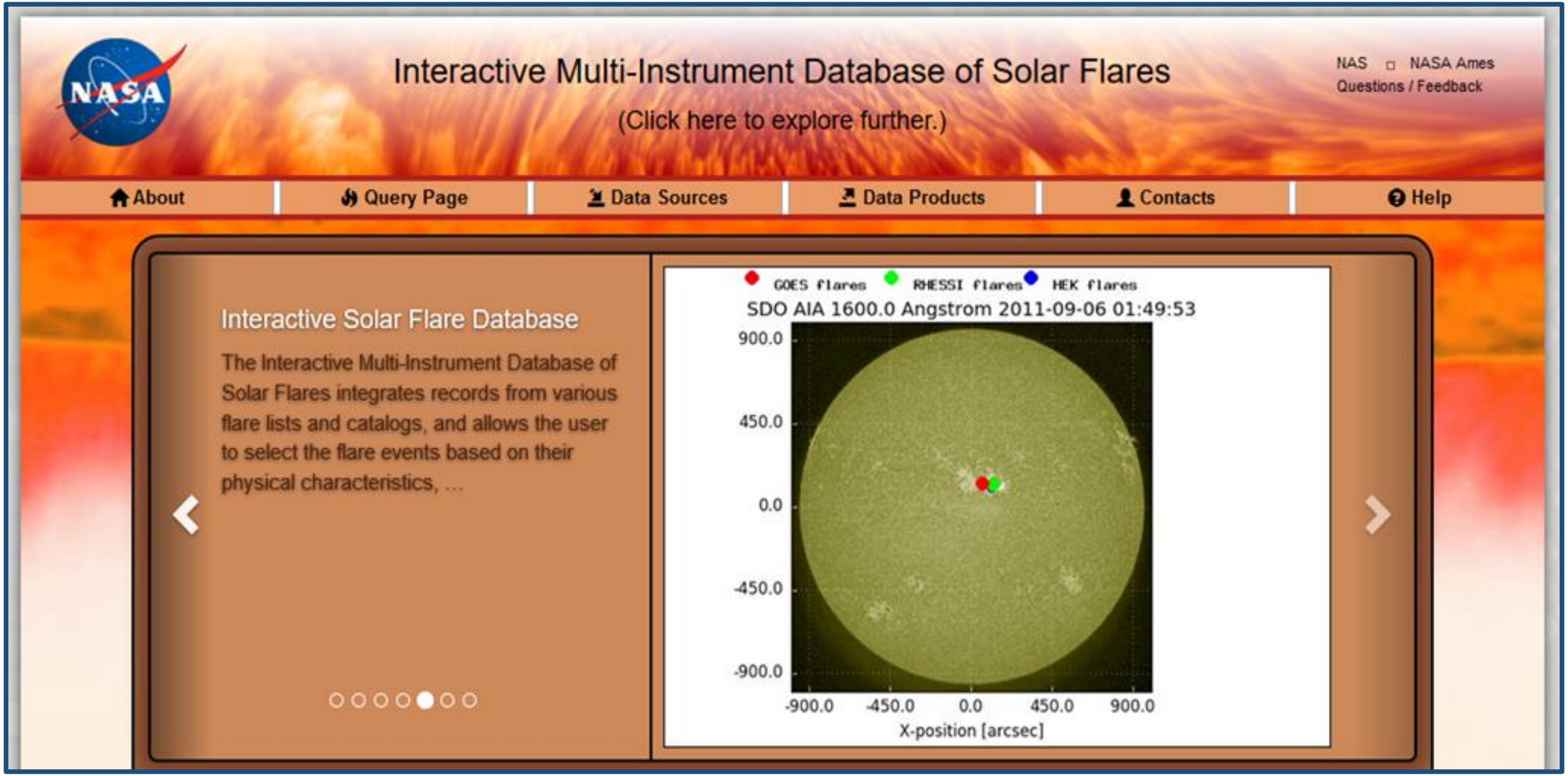
Heliophysics Systems Observatory. Credits:

Research outcome:

- **Developed and implemented the Interactive Multi-Instrument Database of Solar Flares, which integrates data from 15 repositories (including NASA space missions and ground-based observatories)**
  - Integration of flare lists, flare-related event catalogs and observing logs
  - Identification of the uniquely-matched flare events based on time and position information
  - Search for the flare events based on their physical descriptors (automatically calculated)
  - Automatic identification of similar events
  - Automatically updated on the daily basis
- **The Database initiated statistical and machine learning studies that led to discoveries in Space Weather and possibilities for enhancement of Space Weather Operational Forecast**

Interactive Multi-Instrument Database of Solar Flares

(IMIDSF, https://helioportal.nas.nasa.gov/, https://solarflare.njit.edu/)



Discoveries in Space Weather via Statistical Studies

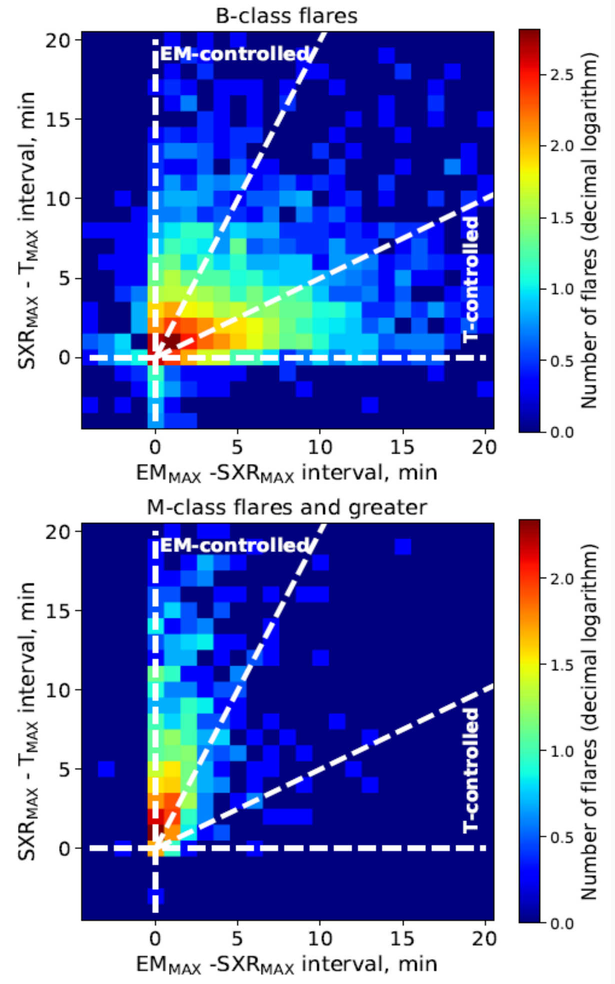
Statistical Study of Chromospheric Evaporation in Solar Flares. Comparison with Models.

- To connect energy fluxes deposited in solar flares and the properties of the responding solar plasma and compare results with the RHD chromospheric evaporation simulations, **the dataset of flares simultaneously observed by IRIS** (here in the fast-scanning regime) **and RHESSI is required.** Such dataset was obtained using Helioportal.

Discovery of two types of flares from GOES Observations.

- The results of the application of TEBBS algorithm (T and EM calculations) for GOES flares detected from 2002 until today **are available as a data product at Helioportal**. The detailed look at timescales between T, SXR, and EM peaks resulted in finding "T-controlled" and "EM-controlled" flares, and determining physical differences between them.

SOLID	GOES class	RHESSI energy range, keV	IRIS mode
SOL2014-02-13T01:32:00	M1.8	6-12	Coarse raster, 8 slits
SOL2014-02-13T02:41:00	M1.0	3-6	Coarse raster, 8 slits
SOL2014-03-29T17:35:00	X1.0	100-300	Coarse raster, 8 slits
SOL2014-06-12T18:03:00	M1.3	25-50	Coarse raster, 8 slits
SOL2014-06-13T00:30:00	C8.5	12-25	Coarse raster, 8 slits
SOL2015-03-11T11:21:00	C5.8	12-25	Coarse raster, 8 slits
SOL2015-11-04T13:31:00	M3.7	50-100	Coarse raster, 16 slits



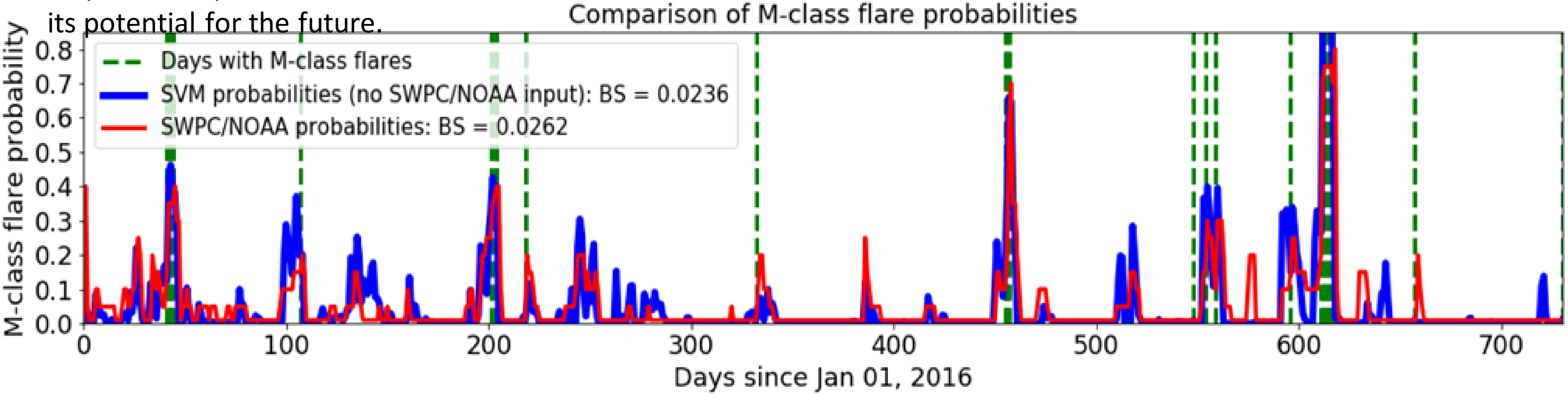
Possibility to Enhance Space Weather Operational Forecasts

- Helioportal allows the users to request **the statistics of flares for each Active Region (AR)** in one click. Integration of the AR magnetic field descriptors (SHARP and PIL parameters) with flare events is started. It is also possible to request **not only the GOES class but other physical characteristics of solar flares**, and work on the prediction of these characteristics

$$BS = \frac{1}{n}\sum_{i=1}^{n}(P_i - Q_i)^2 \qquad BSS = 1 - \frac{BS}{BS_{clim}}$$

- An attempt to build a probabilistic forecast of M-class flares and close comparison with SWPC/NOAA probabilities reveals higher Brier Skill Scores (BSS) for ML-based forecasts with respect to expert-based ones and demonstrates its potential for the future.

BSS, M flares	ES threshold	SVM RBF	SVM Sigmoid
Expert scores (ES)	**0.14±0.04**	-	-
ALL + ES	-	0.30±0.03	**0.32±0.03**
ALL – ES	-	0.30±0.03	**0.32±0.03**



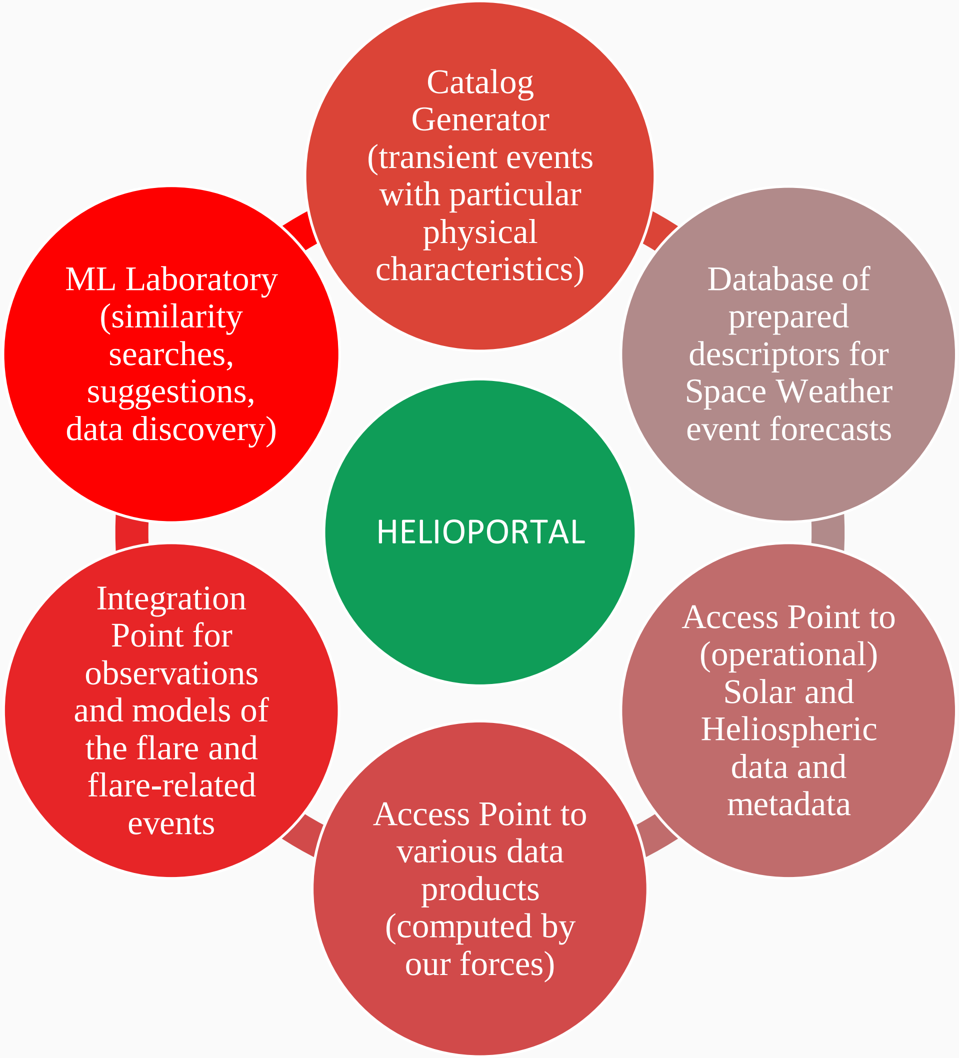
Broader Impacts: outcome

- Contributed to two PhD dissertations:
  - Viacheslav Sadykov (Applied Physics)
  - Sheetal Rajgure (Computer Science)
- Stimulated new EarthCube projects:
  - EarthCube RCN: Towards Integration of Heliophysics Data, Modeling, and Analysis Tools (PI G. Nita) – funded 2018
  - EarthCube Data Capabilities: Development and Application of Advanced Machine Learning Tools and Database to Understand the Onset of Solar Eruptions and Advance the Prediction of Space Weather (PI H. Wang) – funded 2019
- Contributed to NSF Roadmap:
  - Roadmap for Reliable Ensemble Forecasting of the Sun-Earth System
- Contributed to new data analysis and visualization tool:
  - GX Simulator of Active Regions – fully automatic pipeline for 3D data reconstruction and analysis
- Planned integration with:
  - NASA-NSF Coordinated Community Modeling Center
  - Lockheed-Martin Heliophysics Event Knowledgebase
- Contribution to Research and Educational Infrastructure:
  - NJIT Institute for Space Weather Sciences
- Technology transfer: platform for new NASA projects:
  - Interactive Databases of Atmospheric Radiation Dose Rate –funded 2019
  - Employing Machine Learning Methods to Forecast Energetic Proton Events (NASA O2R program, in

Future Plans

Our near-term priorities:

- Processing, integration and visualization of the Solar Energetic Particle (SEP) event data and metadata. Labeling and preparation of these data for the machine learning stage according to current operational forecast needs and requirements
- Reconstruction of 3D structure of solar magnetic fields and subsurface flows, and integration of the descriptors of the magnetic and flow fields in Active Regions (PIL, SHARP, NLFFF, other) into the Helioportal, to enhance physics-guided machine-learning capabilities.



Samira Daneshgar Asl - ICEBERG

ICEBERG

Imagery Cyber-infrastructure and Extensible Building blocks to Enhance Research in the Geosciences

Stony Brook University

RUTGERS
THE STATE UNIVERSITY OF NEW JERSEY

University of Colorado Boulder

UCSB
UNIVERSITY OF CALIFORNIA SANTA BARBARA

NORTHERN ARIZONA UNIVERSITY

https://iceberg-project.github.io/

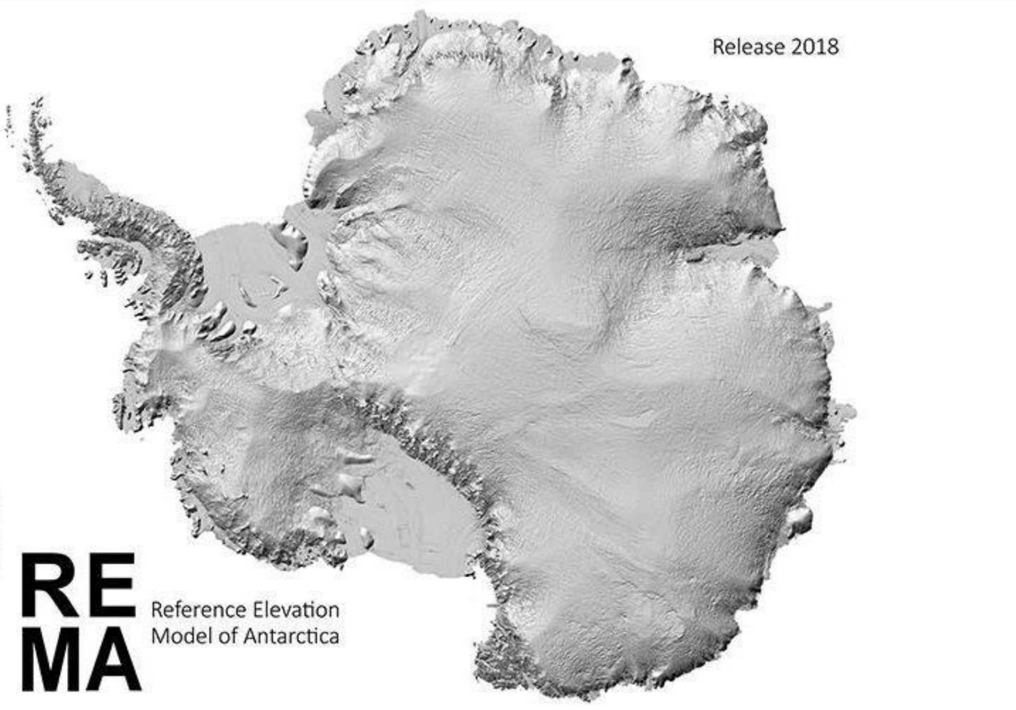
Science ~~Challenge~~ ^Opportunity:

Petabytes of high-resolution satellite imagery + New analytical tools (e.g., machine learning) + High performance and distributed computing resources

= Revolution in imagery-enabled science

- Rapid rise in the use of sub-meter spatial resolution imagery for polar science
- Many pilot demonstrations but few "at scale"
- What's the barrier to scaling up early successes in imagery-enabled science? Barriers to the flow of data, limited experience with HPDC systems, reliant on commercial software that is difficult to parallelize due to licensing issue

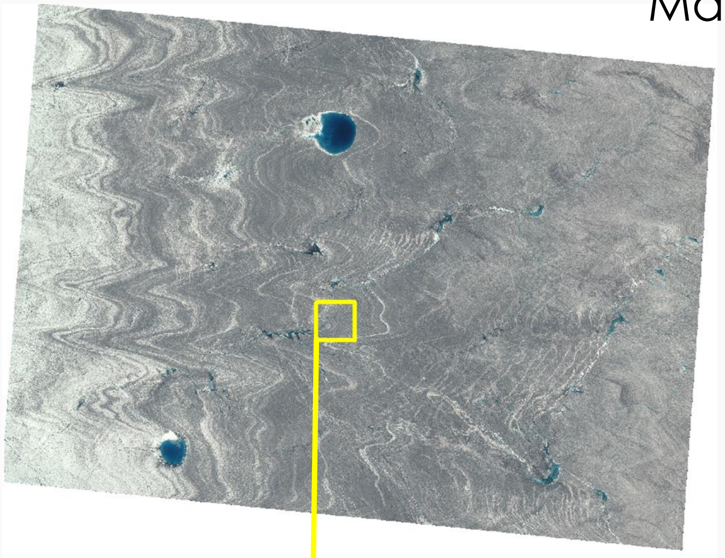
The research objective of this project is to understand the biological, geological, and hydrological functioning of the polar regions **at spatial scales heretofore beyond the reach of individual PIs**, and to **develop tools for imagery-enabled science that can be applied globally**.



Creating large-scale products, such as the new elevation maps for the Arctic and Antarctic, require new computing tools that are tailor made for efficient deployment on high-performance computers.

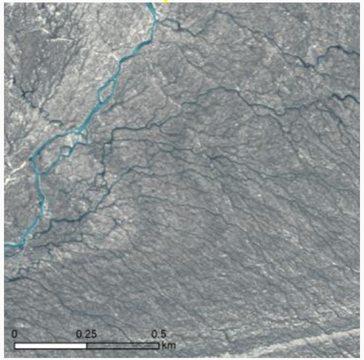
Images and figures courtesy of the Polar Geospatial Center

Manual hand annotation of imagery doesn't scale



AI & 'deep learning' present new opportunities for imagery interpretation

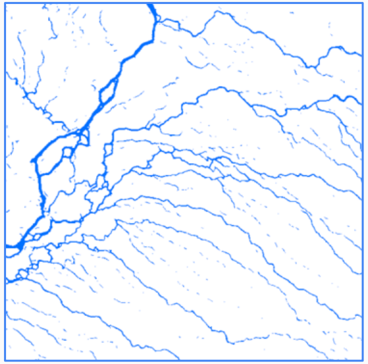
Input Image



Ground Truth

Current available data (Smith et al., 2015)

Prediction



ICEBERG Development Status

- ICEBERG is distributed as a set of Open Source packages:
  - Available through Github (https://github.com/iceberg-project/)
  - Released as packages via PyPI for Python, either combined with middleware or standalone.

- Current State of development:
  - Two of four original use cases are fully supported, released, and executed on XSEDE/Bridges (penguins, seals)
  - Rivers detection executes on XSEDE/Bridges. We are refining the training model
  - Continuing code development for other use cases (including applications outside of polar regions)

Basil Tikoff - StraboSpot

THE STRABOSPOT DIGITAL SYSTEM FOR GEOLOGICAL FIELD-BASED (STRUCTURAL GEOLOGY, SEDIMENTOLOGY, PETROLOGY), MICROANALYSIS, AND EXPERIMENTAL DEFORMATION DATA

Original PIs: **Basil Tikoff** (UWisc), **Julie Newman** (TAMU), **Doug Walker** (KU)

Programmers: Jason Ash, Jessica Novak, Nathan Novak (KU)

Postdocs & grad students: H. Cunningham & A. Snell (TAMU); R. Williams, Z. Michels, M. Kahn, N. Roberts (UWisc); E. Bunse & C. Rufledt (KU); C. Duncan (Utah)

Sedimentology PIs: M. Chan (Utah), E. Hajek (Penn State), D. Kamola (KU)

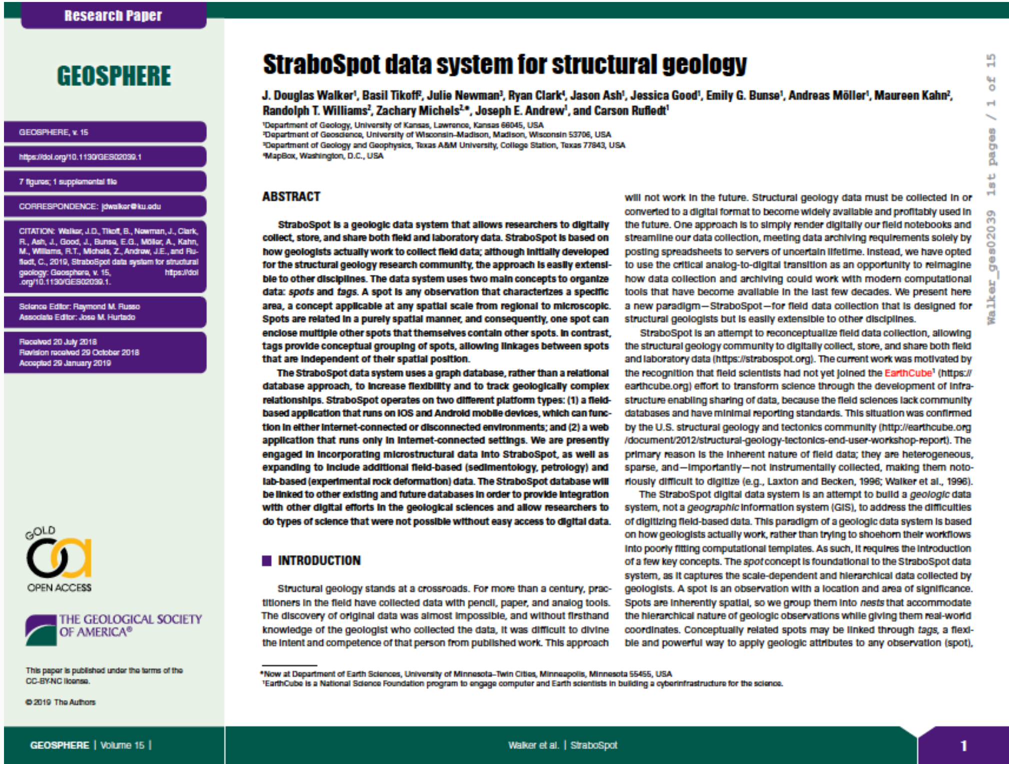
Petrology PIs: A. Glazner (UNC), B. Schoene (Princeton), F. Spear (RPI)

Experimental PIs: A. Kronenberg (TAMU), M. Pec (MIT), U. Mok (MIT), P. Skemer (Washington), C. Marone (Penn State)

Check it out at strabospot.org.

StraboSpot data system is:
-A geologic *and* geographic information system

Reviewer #1: “At last we have a <u>geologic data system</u>, not a geographic information system, with an application that might actually be a huge aid in solving problems.”



Research Paper

GEOSPHERE

GEOSPHERE, v. 15

https://doi.org/10.1130/GES02039.1

7 figures; 1 supplemental file

CORRESPONDENCE: jdwalker@ku.edu

CITATION: Walker, J.D., Tikoff, B., Newman, J., Clark, R., Ash, J., Good, J., Bunse, E.G., Möller, A., Kahn, M., Williams, R.T., Michels, Z., Andrew, J.E., and Rufledt, C., 2019, StraboSpot data system for structural geology: Geosphere, v. 15, https://doi.org/10.1130/GES02039.1.

Science Editor: Raymond M. Russo
Associate Editor: Jose M. Hurtado

Received 20 July 2018
Revision received 29 October 2018
Accepted 29 January 2019

GOLD OPEN ACCESS

THE GEOLOGICAL SOCIETY OF AMERICA®

StraboSpot data system for structural geology

J. Douglas Walker[1], Basil Tikoff[2], Julie Newman[3], Ryan Clark[4], Jason Ash[1], Jessica Good[1], Emily G. Bunse[1], Andreas Möller[1], Maureen Kahn[2], Randolph T. Williams[2], Zachary Michels[2,*], Joseph E. Andrew[1], and Carson Rufledt[1]

[1]Department of Geology, University of Kansas, Lawrence, Kansas 66045, USA
[2]Department of Geoscience, University of Wisconsin–Madison, Madison, Wisconsin 53706, USA
[3]Department of Geology and Geophysics, Texas A&M University, College Station, Texas 77843, USA
[4]MapBox, Washington, D.C., USA

ABSTRACT

StraboSpot is a geologic data system that allows researchers to digitally collect, store, and share both field and laboratory data. StraboSpot is based on how geologists actually work to collect field data; although initially developed for the structural geology research community, the approach is easily extensible to other disciplines. The data system uses two main concepts to organize data: *spots* and *tags*. A spot is any observation that characterizes a specific area, a concept applicable at any spatial scale from regional to microscopic. Spots are related in a purely spatial manner, and consequently, one spot can enclose multiple other spots that themselves contain other spots. In contrast, tags provide conceptual grouping of spots, allowing linkages between spots that are independent of their spatial position.

The StraboSpot data system uses a graph database, rather than a relational database approach, to increase flexibility and to track geologically complex relationships. StraboSpot operates on two different platform types: (1) a field-based application that runs on iOS and Android mobile devices, which can function in either internet-connected or disconnected environments; and (2) a web application that runs only in internet-connected settings. We are presently engaged in incorporating microstructural data into StraboSpot, as well as expanding to include additional field-based (sedimentology, petrology) and lab-based (experimental rock deformation) data. The StraboSpot database will be linked to other existing and future databases in order to provide integration with other digital efforts in the geological sciences and allow researchers to do types of science that were not possible without easy access to digital data.

INTRODUCTION

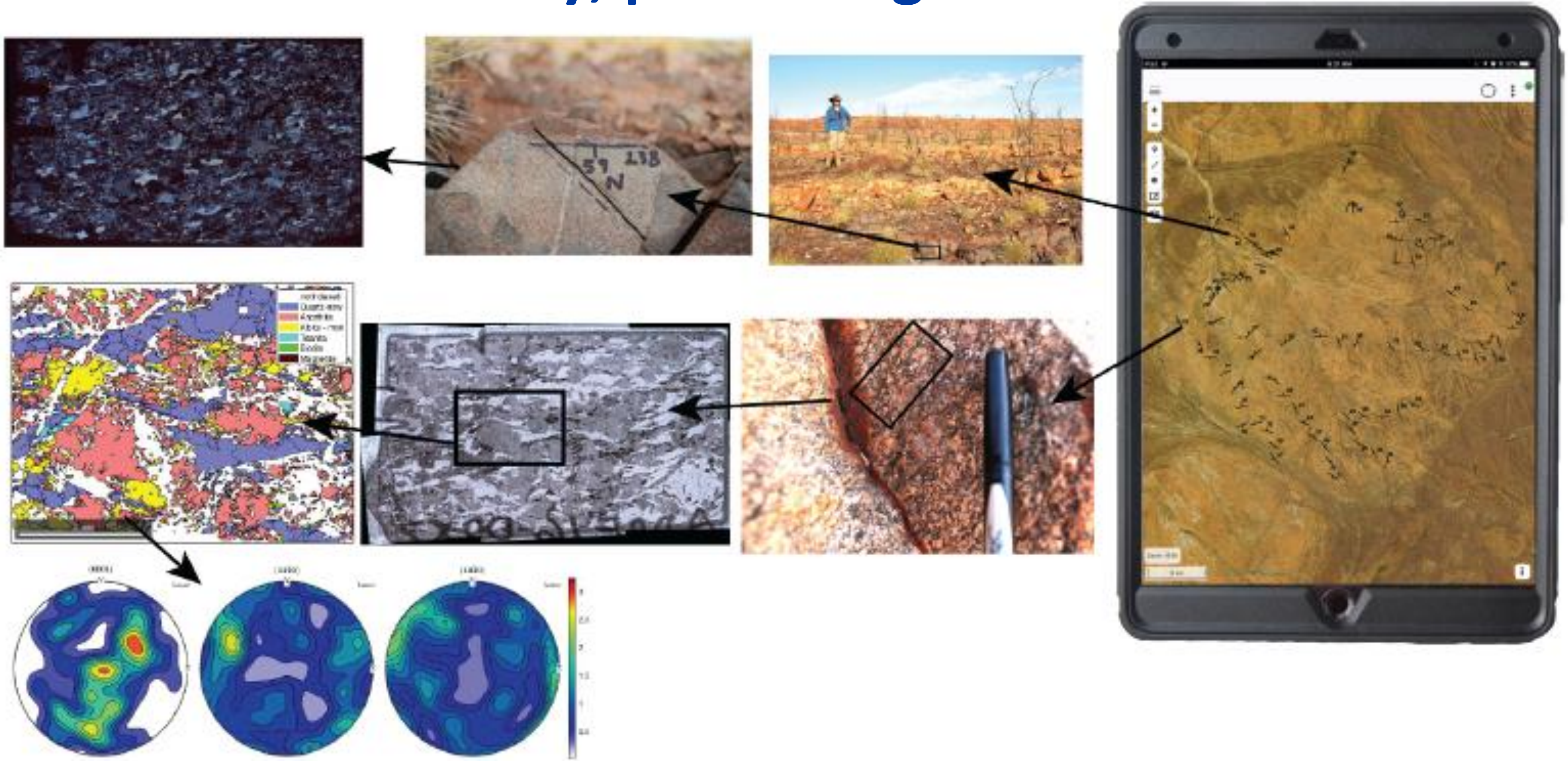
Structural geology stands at a crossroads. For more than a century, practitioners in the field have collected data with pencil, paper, and analog tools. The discovery of original data was almost impossible, and without firsthand knowledge of the geologist who collected the data, it was difficult to divine the intent and competence of that person from published work. This approach will not work in the future. Structural geology data must be collected in or converted to a digital format to become widely available and profitably used in the future. One approach is to simply render digitally our field notebooks and streamline our data collection, meeting data archiving requirements solely by posting spreadsheets to servers of uncertain lifetime. Instead, we have opted to use the critical analog-to-digital transition as an opportunity to reimagine how data collection and archiving could work with modern computational tools that have become available in the last few decades. We present here a new paradigm—StraboSpot—for field data collection that is designed for structural geologists but is easily extensible to other disciplines.

StraboSpot is an attempt to reconceptualize field data collection, allowing the structural geology community to digitally collect, store, and share both field and laboratory data (https://strabospot.org). The current work was motivated by the recognition that field scientists had not yet joined the EarthCube[1] (https://earthcube.org) effort to transform science through the development of infrastructure enabling sharing of data, because the field sciences lack community databases and have minimal reporting standards. This situation was confirmed by the U.S. structural geology and tectonics community (http://earthcube.org/document/2012/structural-geology-tectonics-end-user-workshop-report). The primary reason is the inherent nature of field data; they are heterogeneous, sparse, and—importantly—not instrumentally collected, making them notoriously difficult to digitize (e.g., Laxton and Becken, 1996; Walker et al., 1996).

The StraboSpot digital data system is an attempt to build a *geologic* data system, not a *geographic* information system (GIS), to address the difficulties of digitizing field-based data. This paradigm of a geologic data system is based on how geologists actually work, rather than trying to shoehorn their workflows into poorly fitting computational templates. As such, it requires the introduction of a few key concepts. The *spot* concept is foundational to the StraboSpot data system, as it captures the scale-dependent and hierarchical data collected by geologists. A spot is an observation with a location and area of significance. Spots are inherently spatial, so we group them into *nests* that accommodate the hierarchical nature of geologic observations while giving them real-world coordinates. Conceptually related spots may be linked through *tags*, a flexible and powerful way to apply geologic attributes to any observation (spot),

*Now at Department of Earth Sciences, University of Minnesota–Twin Cities, Minneapolis, Minnesota 55455, USA
[1]EarthCube is a National Science Foundation program to engage computer and Earth scientists in building a cyberinfrastructure for the science.

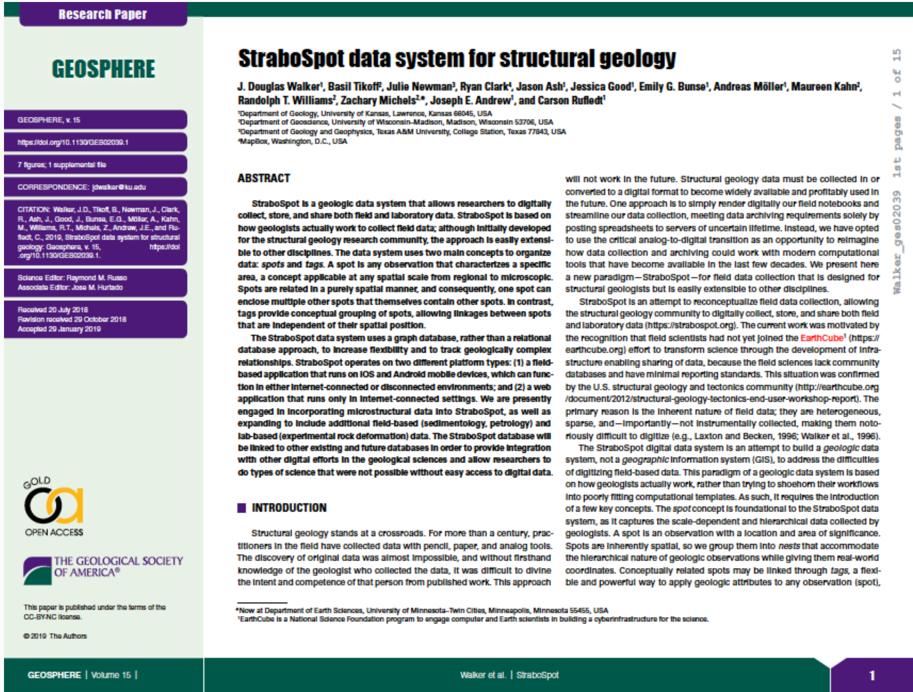
The goal: To track geological information from the field to laboratory, preserving orientation and scale



StraboSpot data system is:
-A geologic and geographic information system
-A community-directed data system

Reviewer #2: Your introduction stinks.
Here, let me rewrite it for you…

"…we have opted to use (the) momentous transition to digital as an opportunity to reimagine how data collection and archiving… We present here a new paradigm for field data collection that is designed for structural geologists but is easily extensible to other disciplines. This paradigm is based on how geologists actually work, rather than trying to shoehorn their workflows into poorly fitting computational templates..."


Research Paper

GEOSPHERE

GEOSPHERE, v. 15

https://doi.org/10.1130/GES02039.1

7 figures; 1 supplemental file

CORRESPONDENCE: jdwalker@ku.edu

CITATION: Walker, J.D., Tikoff, B., Newman, J., Clark, R., Ash, J., Good, J., Bunse, E.G., Möller, A., Kahn, M., Williams, R.T., Michels, Z., Andrew, J.E., and Rufledt, C., 2019, StraboSpot data system for structural geology: Geosphere, v. 15, https://doi.org/10.1130/GES02039.1.

Science Editor: Raymond M. Russo
Associate Editor: Jose M. Hurtado

Received 20 July 2018
Revision received 29 October 2018
Accepted 29 January 2019

StraboSpot data system for structural geology

J. Douglas Walker[1], Basil Tikoff[2], Julie Newman[3], Ryan Clark[4], Jason Ash[1], Jessica Good[1], Emily G. Bunse[1], Andreas Möller[1], Maureen Kahn[2], Randolph T. Williams[2], Zachary Michels[2,*], Joseph E. Andrew[1], and Carson Rufledt[1]
[1]Department of Geology, University of Kansas, Lawrence, Kansas 66045, USA
[2]Department of Geoscience, University of Wisconsin–Madison, Madison, Wisconsin 53706, USA
[3]Department of Geology and Geophysics, Texas A&M University, College Station, Texas 77843, USA
[4]MapBox, Washington, D.C., USA

ABSTRACT

StraboSpot is a geologic data system that allows researchers to digitally collect, store, and share both field and laboratory data. StraboSpot is based on how geologists actually work to collect field data; although initially developed for the structural geology research community, the approach is easily extensible to other disciplines. The data system uses two main concepts to organize data: *spots* and *tags*. A spot is any observation that characterizes a specific area, a concept applicable at any spatial scale from regional to microscopic. Spots are related in a purely spatial manner, and consequently, one spot can enclose multiple other spots that themselves contain other spots. In contrast, tags provide conceptual grouping of spots, allowing linkages between spots that are independent of their spatial position.

The StraboSpot data system uses a graph database, rather than a relational database approach, to increase flexibility and to track geologically complex relationships. StraboSpot operates on two different platform types: (1) a field-based application that runs on iOS and Android mobile devices, which can function in either internet-connected or disconnected environments; and (2) a web application that runs only in internet-connected settings. We are presently engaged in incorporating microstructural data into StraboSpot, as well as expanding to include additional field-based (sedimentology, petrology) and lab-based (experimental rock deformation) data. The StraboSpot database will be linked to other existing and future databases in order to provide integration with other digital efforts in the geological sciences and allow researchers to do types of science that were not possible without easy access to digital data.

INTRODUCTION

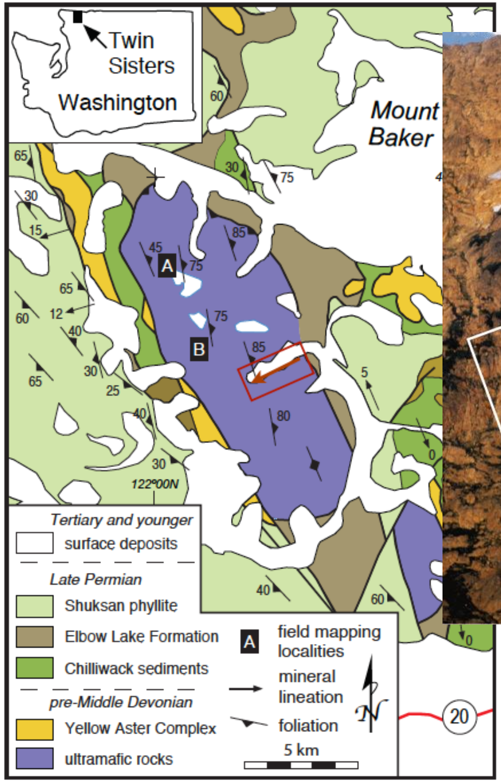
Structural geology stands at a crossroads. For more than a century, practitioners in the field have collected data with pencil, paper, and analog tools. The discovery of original data was almost impossible, and without firsthand knowledge of the geologist who collected the data, it was difficult to divine the intent and competence of that person from published work. This approach will not work in the future. Structural geology data must be collected in or converted to a digital format to become widely available and profitably used in the future. One approach is to simply render digitally our field notebooks and streamline our data collection, meeting data archiving requirements solely by posting spreadsheets to servers of uncertain lifetime. Instead, we have opted to use the critical analog-to-digital transition as an opportunity to reimagine how data collection and archiving could work with modern computational tools that have become available in the last few decades. We present here a new paradigm—StraboSpot—for field data collection that is designed for structural geologists but is easily extensible to other disciplines.

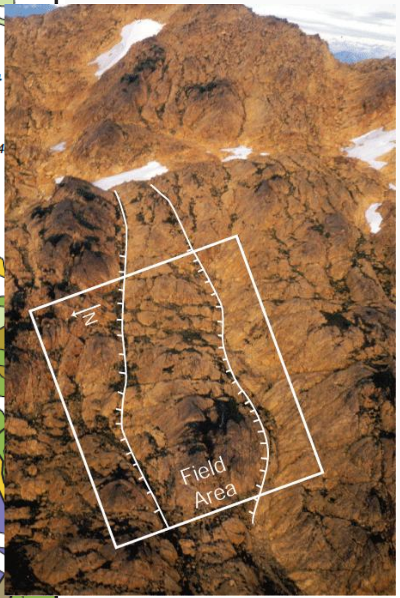
StraboSpot is an attempt to reconceptualize field data collection, allowing the structural geology community to digitally collect, store, and share both field and laboratory data (https://strabospot.org). The current work was motivated by the recognition that field scientists had not yet joined the EarthCube[1] (https://earthcube.org) effort to transform science through the development of infrastructure enabling sharing of data, because the field sciences lack community databases and have minimal reporting standards. This situation was confirmed by the U.S. structural geology and tectonics community (http://earthcube.org/document/2012/structural-geology-tectonics-end-user-workshop-report). The primary reason is the inherent nature of field data; they are heterogeneous, sparse, and—importantly—not instrumentally collected, making them notoriously difficult to digitize (e.g., Laxton and Becken, 1996; Walker et al., 1996).

The StraboSpot digital data system is an attempt to build a *geologic* data system, not a *geographic* information system (GIS), to address the difficulties of digitizing field-based data. This paradigm of a geologic data system is based on how geologists actually work, rather than trying to shoehorn their workflows into poorly fitting computational templates. As such, it requires the introduction of a few key concepts. The *spot* concept is foundational to the StraboSpot data system, as it captures the scale-dependent and hierarchical data collected by geologists. A spot is an observation with a location and area of significance. Spots are inherently spatial, so we group them into *nests* that accommodate the hierarchical nature of geologic observations while giving them real-world coordinates. Conceptually related spots may be linked through *tags*, a flexible and powerful way to apply geologic attributes to any observation (spot),

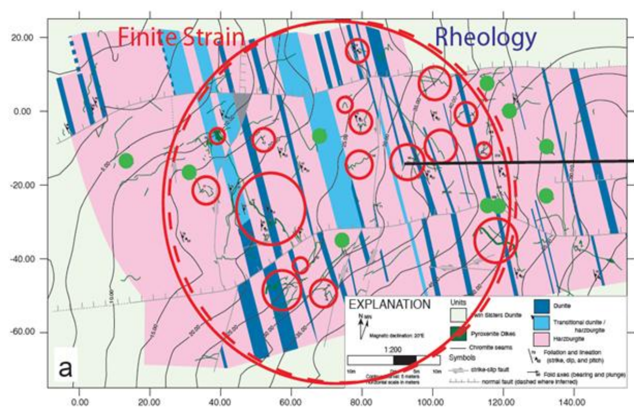
*Now at Department of Earth Sciences, University of Minnesota–Twin Cities, Minneapolis, Minnesota 55455, USA
[1]EarthCube is a National Science Foundation program to engage computer and Earth scientists in building a cyberinfrastructure for the science.

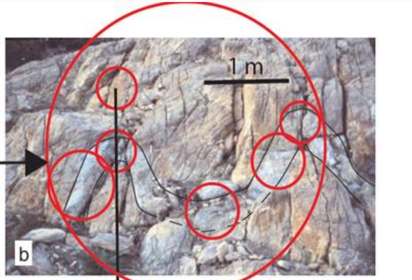
The approach: Use “spots” as an organizational principle to allow both spatial and conceptual grouping

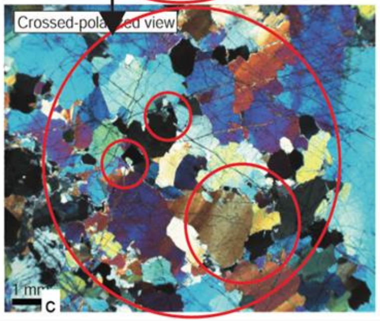




A spot is an area which can be characterized by a particular attribute. It is a way or organizing data collected at different spatial scales



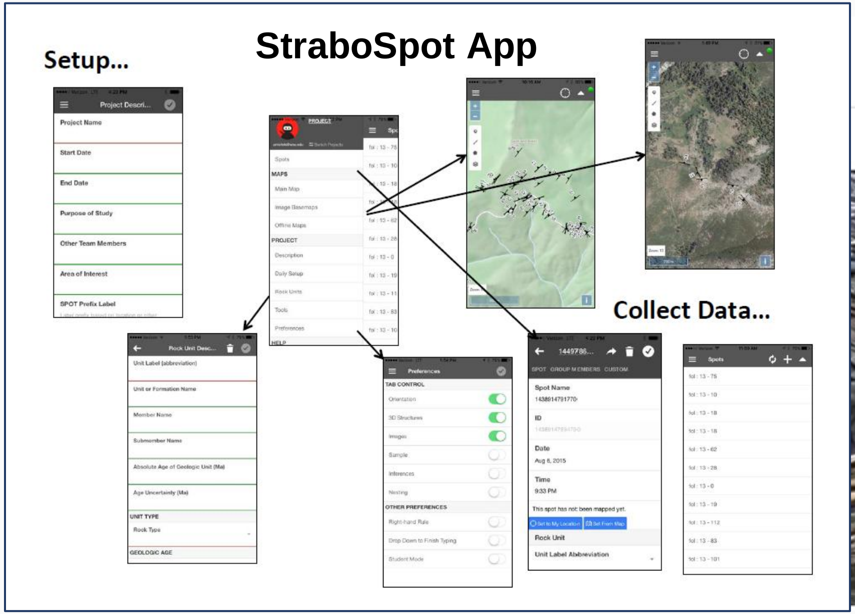




Spots within spots inherently tracked

StraboSpot data system is:
-A geologic and geographic information system
-A community-driven data system
-Is expanding through the geological community

-Currently includes structural geology, sedimentology, petrology, experimental rock deformation, microanalysis
-Working to include geomorphology, active tectonics, tephra stratigraphy



Mobile App:

- iOS or Android
- Location services
- Used at fieldcamps

Open Source Codebase and Public API

- Anyone can write external modules to leverage Strabo
- Tailor new applications to query, upload, and download data
- Design plug-ins for GIS platform

Desktop:

- Flexible, web-based interface
- Import/Export data for your favorite analysis software
- Choose which data you want to share publicly

Strabospot.org