EarthCube Town Hall:

Evolution of the Geosciences / Cyberinfrastructure Community

Christine Kirkpatrick
SDSC-UCSD

December 13, 2022

This work is supported through the NSF award #1928208.

J-EDI

Justice, Equity, Diversity, and Inclusion

The heart of EarthCube is people. We put people first and do our best to recognize, appreciate, and **respect the diversity** of our global contributors. **EarthCube welcomes** contributions from **everyone** who shares our interests and wants to contribute in a **healthy and constructive** manner within our community.

We acknowledge that for millennia, Indigineous peoples have been a part of the land where this venue and our respective institutions' reside. This land has nourished, healed, and protected people for many generations in a relationship of balance and harmony.

EarthCube

NSF

Town Hall Overview

- EarthCube Retrospective - C. Kirkpatrick
- Sustainability - Mike Daniels
- NSF Update - Eva Zanzerkia
- Key Projects for Continued EC Involvement
  - DeCODER - Kenton McHenry
  - FARR - Christine Kirkpatrick
- Funded Project Lightning Talks
  - OpenMindat - Marshall Ma
  - Project Raijin - John Clyne

EarthCube Retrospective Highlights

- ***November 8-9, 2022 @AGU in DC -*** wonderful hosts
- ***Met and exceeded expectations***
  - EarthCube history by Cliff Jacobs, NSF (retired)
- ***Great impact***
  - EC-funded publications have 2x the citations as the average papers in the same journal
  - EC scholarship impactful in altmetrics, e.g.picked up by news outlets, social media, usage
- ***Vibrant and interconnected community network***
- ***Can't be all volunteer***
  - EC found just the right combination of governance, staffing, and incentives to keep work moving and the community flowing

Top 10 Altmetric scores

- full (114) Altmetric scores [csv (download)] | [markdown (.md) (download)]
- top 10 Altmetric scores [markdown (.md) (download)]

Altmetric Score	Publication Title	Realtime Score**
980.77	A global multiproxy database for temperature reconstructions of the Common Era (doi: 10.1038/sdata.2017.88)	972

Picked up by **36** news outlets
Blogged by **15**
Tweeted by **939**
On **12** Facebook pages
Referenced in **6** Wikipedia pages
Mentioned in **2** Google+ posts
Reddited by **4**
Mentioned in **1** Q&A threads

375 readers on Mendeley
1 readers on CiteULike

See more details

Impact figure source: Maull, K., and M. Mayernik, 2022: EarthCube Program Metrics Analysis 2013-2022. https://doi.org/10.5065/VMFJ-QY55.

EarthCube Retrospective Highlights ct.

- ***Contributed to positive culture change***
  - *emphasis on cooperation, collective effort, and inclusivity*
  - *more work to do*
- ***What's Next?***
  - *strong signals from the community that new opportunities like EC are needed*

Progress: The balance between cooperation and competition in the culture of your field or discipline.

Progress: Degree to which success is primarily a product of individual effort or a product of collective effort.

Progress: EarthCube initiative is inclusive in the way it operates.

Pain Point: Trusting that shared data, tools, models, notebooks, and software will be well-documented and reliable.

Pain Point: Tenure, promotion, and rewards in my organization recognize and value sharing research data.

Image source: Cutcher-Gershenfeld, Joel & Canavan, Pat, Waymark Analytics

Additional Reflections

- EarthCube contributed significantly to progress & leadership in intl FAIR data efforts
- In addition to NSF, NASA, NOAA, USGS recognize value of EC community and tools
  - opportunities for community post-EC
- EarthCube's legacy includes robust sustainability models

EarthCube Sustainability

Mike Daniels
Ronin Institute

December 13, 2022

EarthCube Sustainability Timeline

January-March 2021: Discussions with NSF Program Managers
June-July 2021: Listening Sessions at the 2021 EarthCube Annual Meeting
September 2021: EarthCube Sustainability Panel (ECSP) Convened
October-December 2021: CDF and CFP Sustainability Working Groups formed
December 2021: Draft ECSP Report Announced at Fall 2021 AGU
January 2022: Community feedback incorporated into the ECSP Report
February 2022: Final ECSP Report Released
June 2022: Sunset/Reshaping of Earthcube Governance Committees
May 2022: CDF Sustainability Working Group Report Completed
July 2022: CFP Sustainability Working Group Report Completed
July 2022: First meeting of the EarthCube Transition Leadership Team
November 2022: EarthCube Retrospective in Washington, D.C.
December 2022: Final Wrap-up of EarthCube Governance and the Office

EarthCube

https://tinyurl.com/E3-Sustainability

NSF

The EarthCube Transition Leadership Team

Mike Daniels
Ronin Institute

Stephen Kuehn
Concord
University

Christine Kirkpatrick
ECO/SDSC

Dave Fulker
OPeNDAP

Leah LeVay
Texas A&M

Kenton McHenry
ECO/NCSA

Denise Hills
Advanced
Resources
International

Lynne
Schreiber
ECO/SDSC

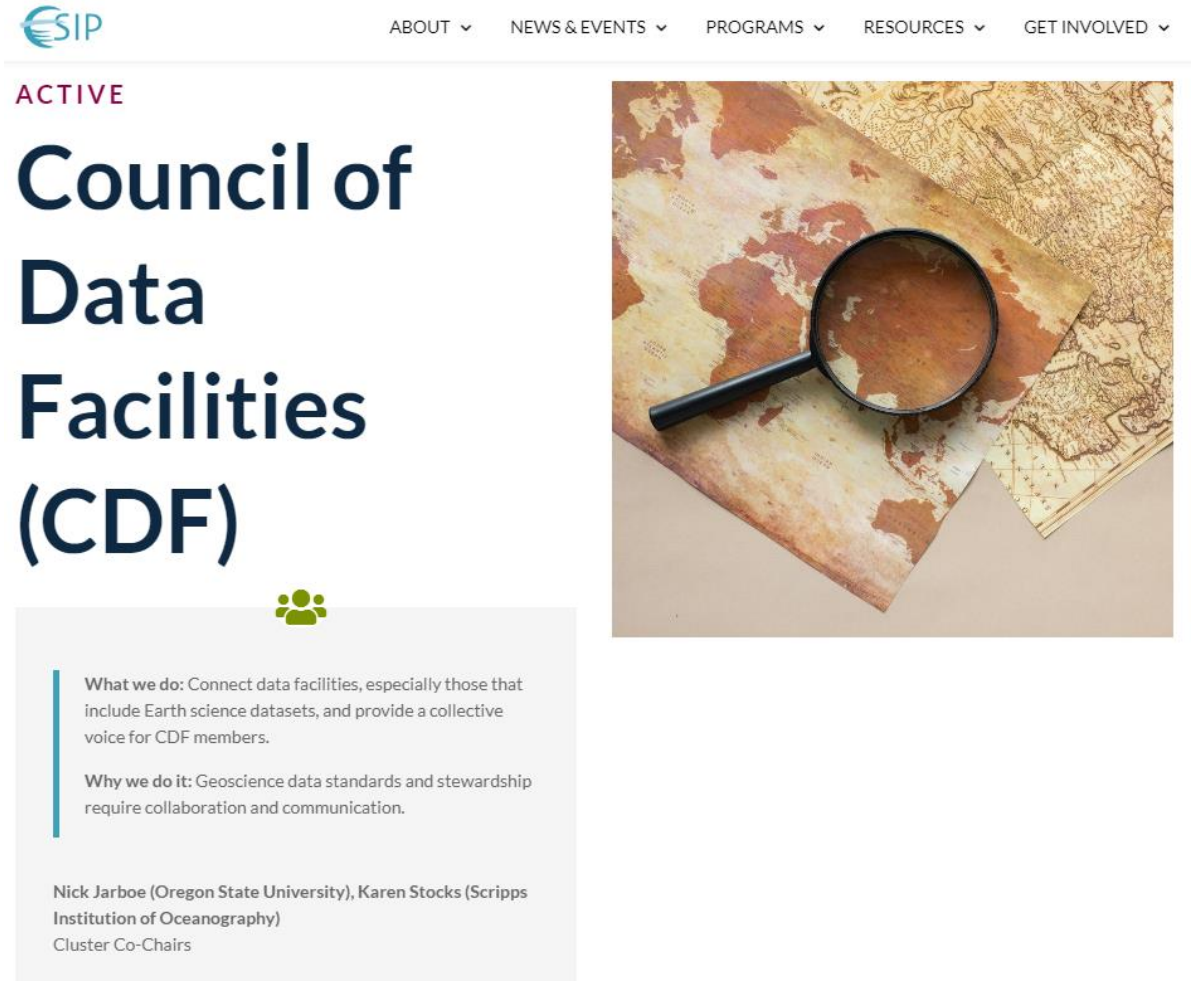
Karen Stocks
ECO/UCSD

EarthCube

https://tinyurl.com/EC-Transition

CDF is reshaping itself as an ESIP Cluster

Join CDF discussions every second Friday of the month at noon Eastern.



EarthCube

https://www.esipfed.org/get-involved/collaborate/council-of-data-facilities-cdf

NSF

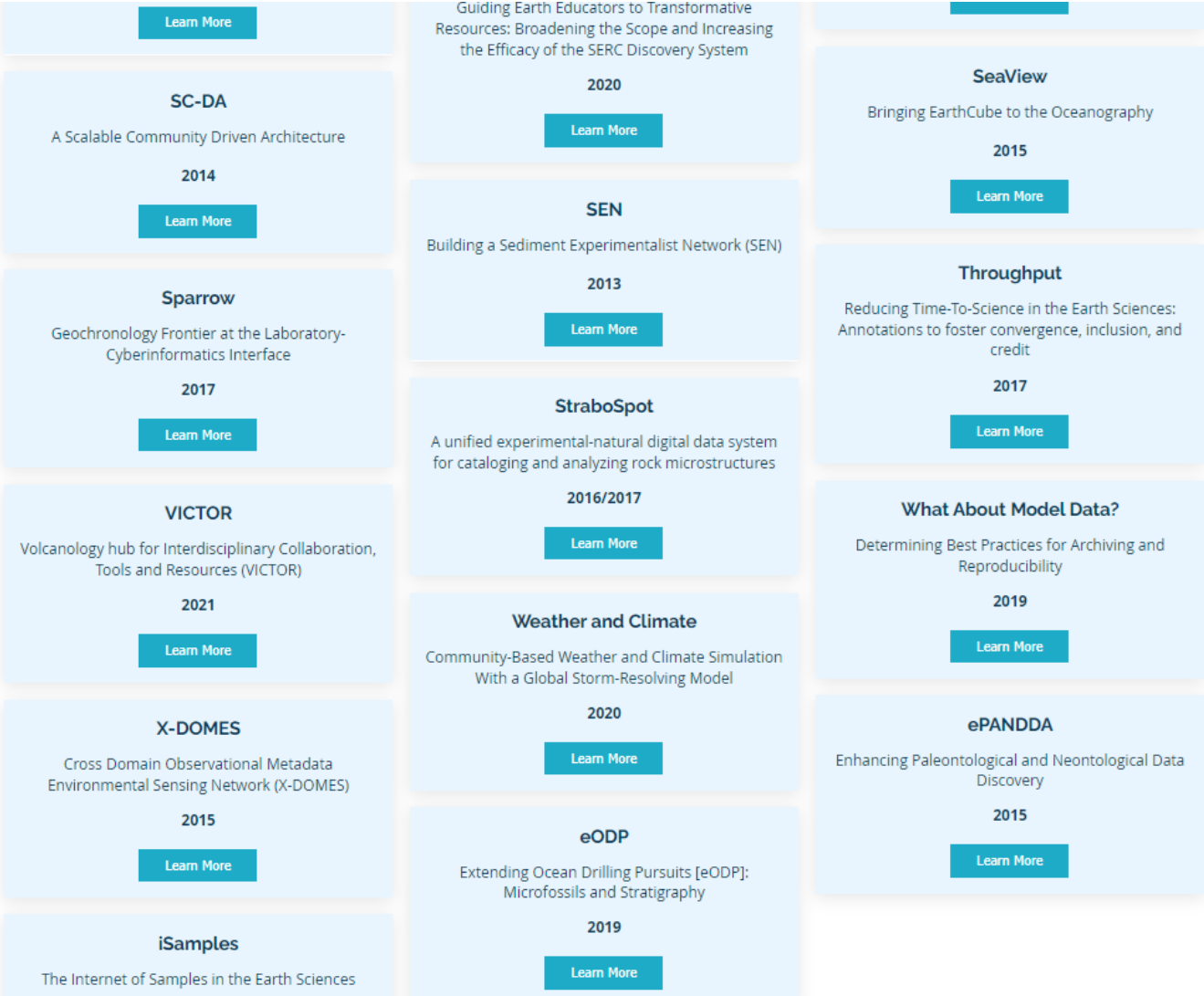
Notebooks as Scholarly Objects

https://data.agu.org/notebooks-now/

Sustainability models for integrated digital Earth Science

Virapongse, Arika; Gallagher, James; Tikoff, Basil; Cornillon, Peter; Koskela, Rebecca; Shingledecker, Susan; Trabant, Chad; Hanson, Brooks (2022). Sustainability models for integrated digital Earth Science. In EarthCube Organization Materials. UC San Diego Library Digital Collections. https://doi.org/10.6075/J0JH3MBN

EarthCube



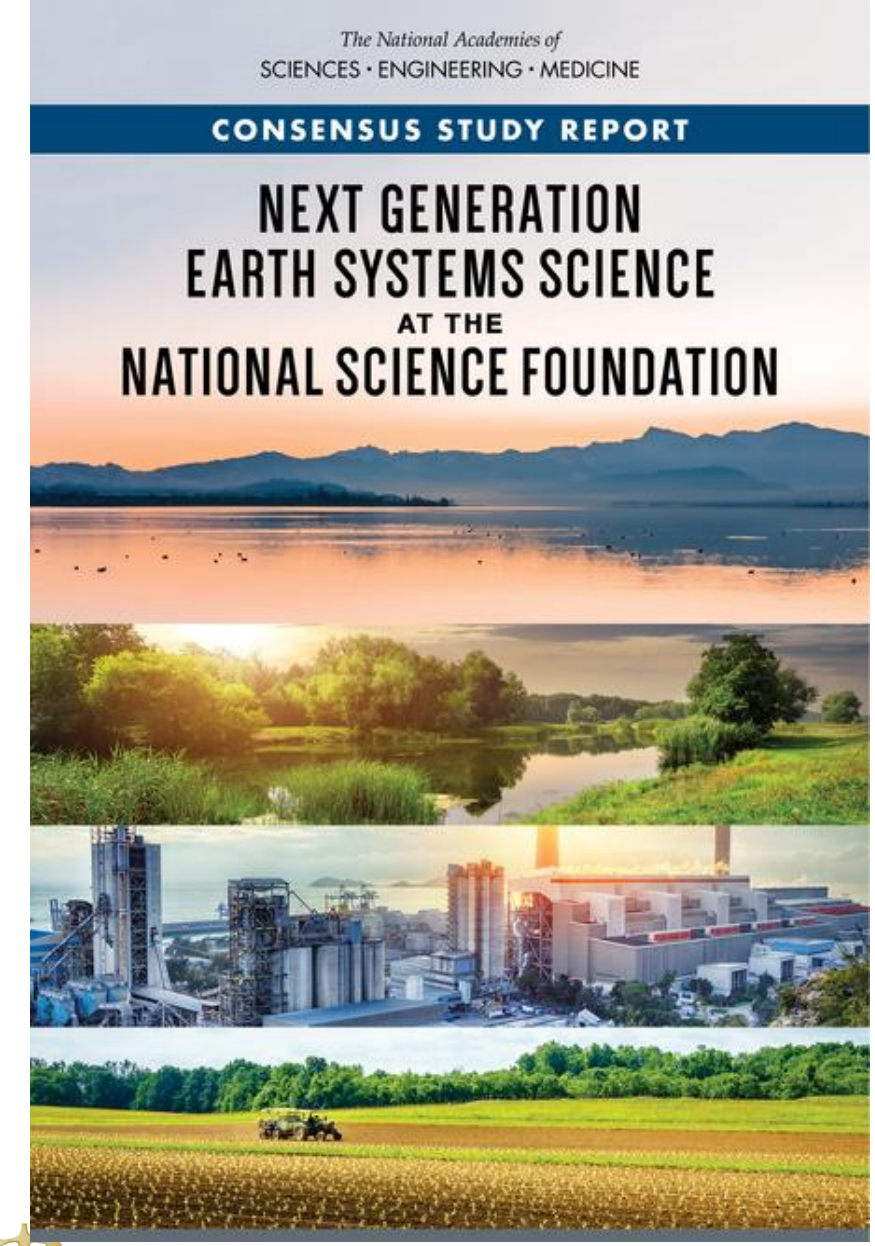
Sustainability of the EarthCube Community

Please stay tuned as discussions are underway for support of a Summer 2023 workshop which will serve as a bridge from EarthCube to new NSF initiatives for CI and Earth Science!

Thank you!

NSF Opportunities 2023

Eva Zanzerkia, NSF/GEO



National and community reports call for future R&D investments in cyberinfrastructure and open science

National Research Infrastructure plans (NITRD)

- National Discovery Cloud
- Open Science Data Commons
- Future Advanced Computing
- National AI Research Resource

Key Characteristics for NSF Earth Systems Sciences (NASEM 2022 report)

- Advance both curiosity-driven and use-inspired basic research
- Use observational, computational, and modeling capabilities synergistically to accelerate discovery and convergence.
- Educate and support a workforce with the skills and knowledge

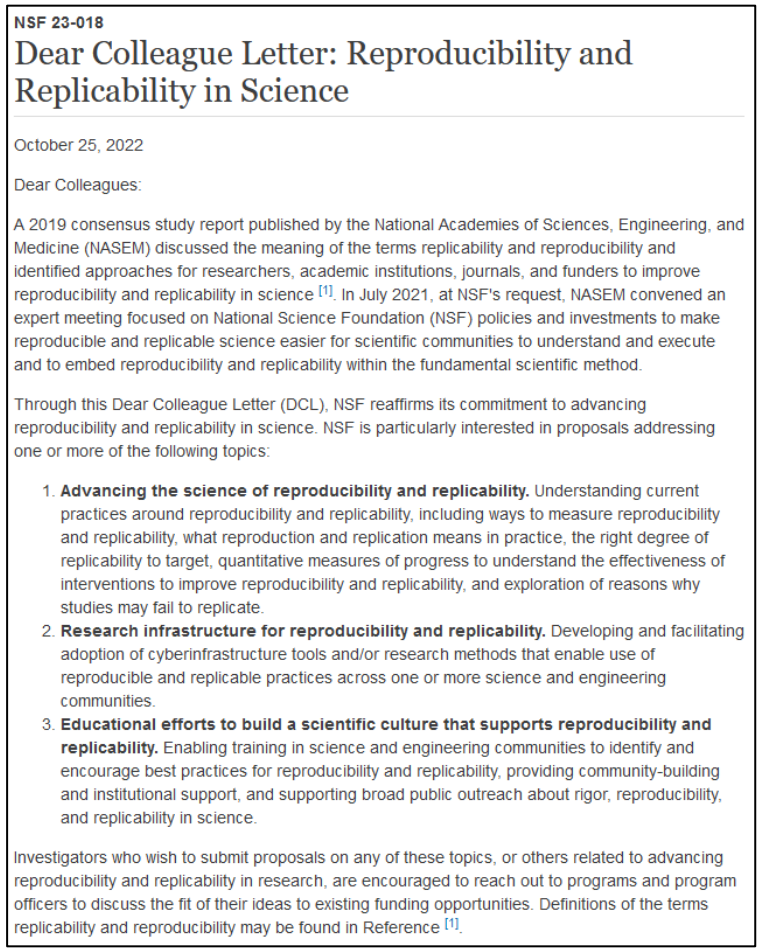
Interagency Collaborations

- NSF-NOAA partnership team on Earth System Modeling and AI/ML
- ICAMS (Interagency Council for Advancing Meteorological Services) Subcommittee on Earth Systems ML/AI and Advanced Technology

Growing importance of open, accessible, and reproducible science at NSF

- **OSTP Memo**: *Ensuring Free, Immediate, and Equitable Access to Federally Funded Research* (Aug. 25, 2022)
- **FAIROS RCNs** (NSF 22-553)
  - New NSF program supporting Research Coordination Networks (RCNs) that advance FAIR principles (findable, accessible, interoperable, reusable) and open science (OS) practices
- **Reproducibility & Replicability DCL** (NSF 23-018)

** Public Access Listen and Learn Session for NSF Stakeholders
November 30, 2022, 2 PM EST via Zoom **
https://nsf.zoomgov.com/webinar/register/WN_nKHJhfI4RJGWQnEXqFiaLQ



NSF 23-018

Dear Colleague Letter: Reproducibility and Replicability in Science

October 25, 2022

Dear Colleagues:

A 2019 consensus study report published by the National Academies of Sciences, Engineering, and Medicine (NASEM) discussed the meaning of the terms replicability and reproducibility and identified approaches for researchers, academic institutions, journals, and funders to improve reproducibility and replicability in science [1]. In July 2021, at NSF's request, NASEM convened an expert meeting focused on National Science Foundation (NSF) policies and investments to make reproducible and replicable science easier for scientific communities to understand and execute and to embed reproducibility and replicability within the fundamental scientific method.

Through this Dear Colleague Letter (DCL), NSF reaffirms its commitment to advancing reproducibility and replicability in science. NSF is particularly interested in proposals addressing one or more of the following topics:

1. **Advancing the science of reproducibility and replicability.** Understanding current practices around reproducibility and replicability, including ways to measure reproducibility and replicability, what reproduction and replication means in practice, the right degree of replicability to target, quantitative measures of progress to understand the effectiveness of interventions to improve reproducibility and replicability, and exploration of reasons why studies may fail to replicate.
2. **Research infrastructure for reproducibility and replicability.** Developing and facilitating adoption of cyberinfrastructure tools and/or research methods that enable use of reproducible and replicable practices across one or more science and engineering communities.
3. **Educational efforts to build a scientific culture that supports reproducibility and replicability.** Enabling training in science and engineering communities to identify and encourage best practices for reproducibility and replicability, providing community-building and institutional support, and supporting broad public outreach about rigor, reproducibility, and replicability in science.

Investigators who wish to submit proposals on any of these topics, or others related to advancing reproducibility and replicability in research, are encouraged to reach out to programs and program officers to discuss the fit of their ideas to existing funding opportunities. Definitions of the terms replicability and reproducibility may be found in Reference [1].

NSF's FY 2023 Budget Request:
Major GEO Investments

NSF-Wide Climate Activities: $33.67M

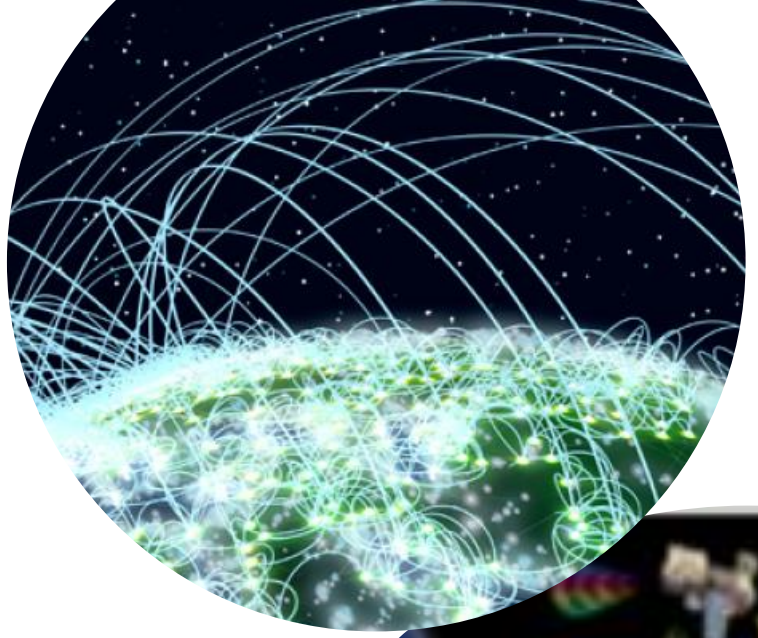
GEO Climate Incubators: $80M

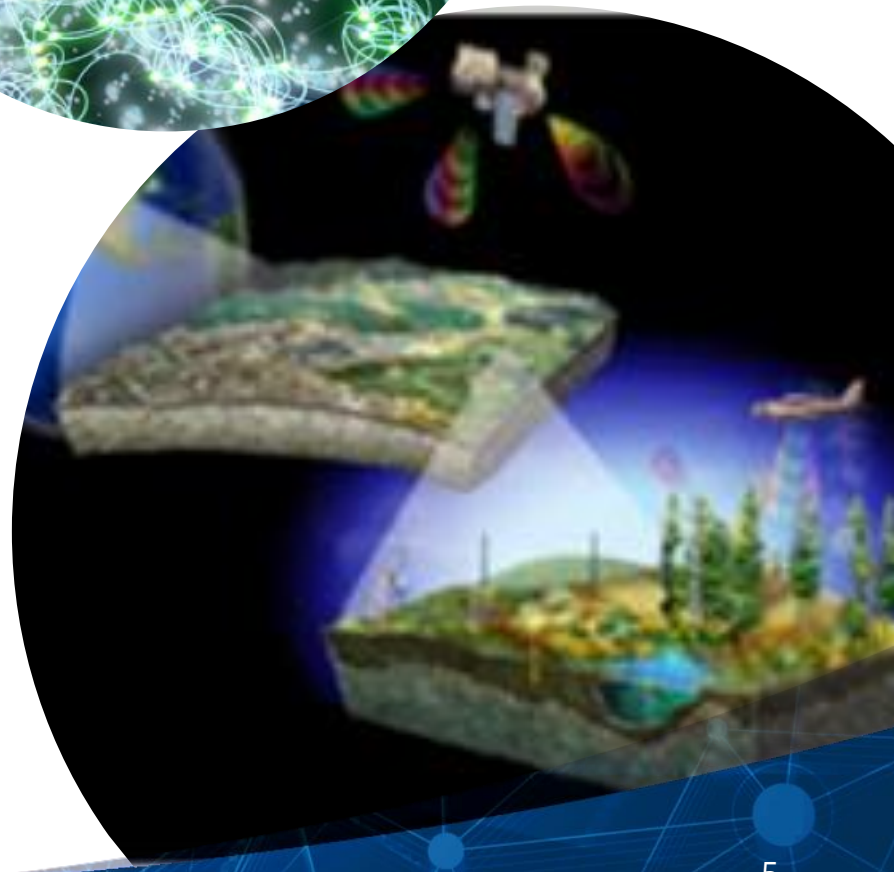
GEO Cyberinfrastructure Incubator: $20M

GEO Education, Diversity, and Equity Incubator: $7.5M

Foundational Investments

GEO Cyberinfrastructure Goals





- **Advance geoscience research**
- **Promote openness and participation through Open Science**
- **Pursue AI/ML innovation in the geosciences**

Office of Advanced Cyberinfrastructure (OAC) opportunities

- **Cyberinfrastructure for Sustained Scientific Innovation (CSSI)** (NSF 22-632): Support for robust, reliable, and sustainable data and software cyberinfrastructure (*Deadline:* December 16, 2022)
- **CyberTraining (NSF 23-520):** Supports efforts toward broad adoption of CI tools, methods, and resources; and integration of CI literacy into curriculum / instructional materials (*Deadline:* Feb. 23, 2023)
- **SCIPE (NSF 23-521):** "Strengthening the Cyberinfrastructure Professionals (CIP) Ecosystem" (*Deadline:* Feb. 23, 2023)
- *OAC CI Resource Ecosystem:*
  - **ACCESS**: Coordinates allocations for advanced computing, visualization, and data resources for researchers and educators (follow-on to XSEDE) (https://access-ci.org)
  - **CloudBank:** Enabling access to commercial cloud service resources
  - **Partnership to Advance Throughput Computing (PATh):** Pilot support for high-throughput computing (HTC) resources (see NSF 22-051)

More information here: https://www.nsf.gov/geo/geo-ci

Relevant CI-Related Programs

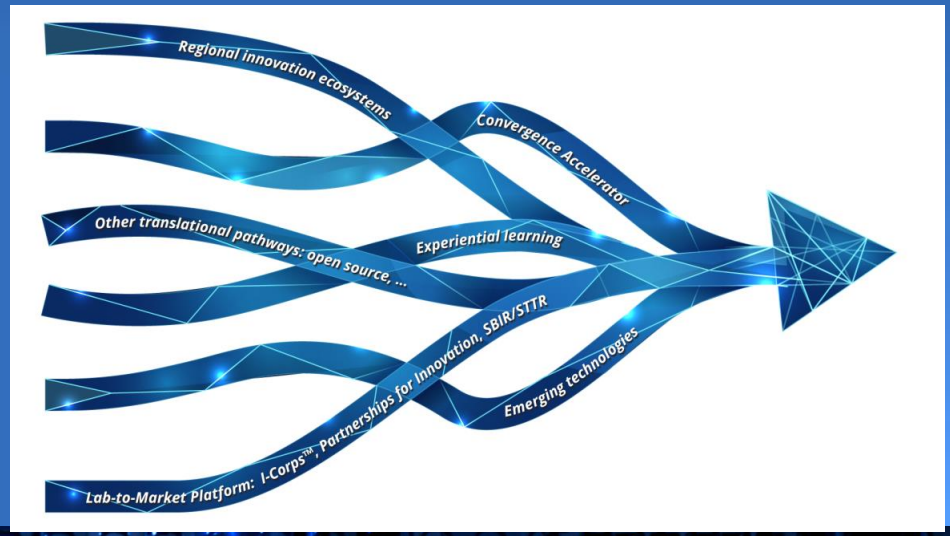
	CSSI	OAC CORE	SCIPE	CyberTraining
What will your project focus on?	Develop software or data repository	Perform research that will enable future CI	Support research that uses CI, foster CIP careers	Provide training
What gap will you fill?	Increase community CI for research or education	Increase knowledge needed for CI	Increase support for CIP career & research by underserved groups	Increase research workforce to leverage CI
What will your project deliver?	Community-sustained CI	Techniques that will enable CI and a CI prototype	Research support in CI, CIP career paths	Scalable and sustainable training program

Technology, Innovation, and Partnerships (TIP) opportunities

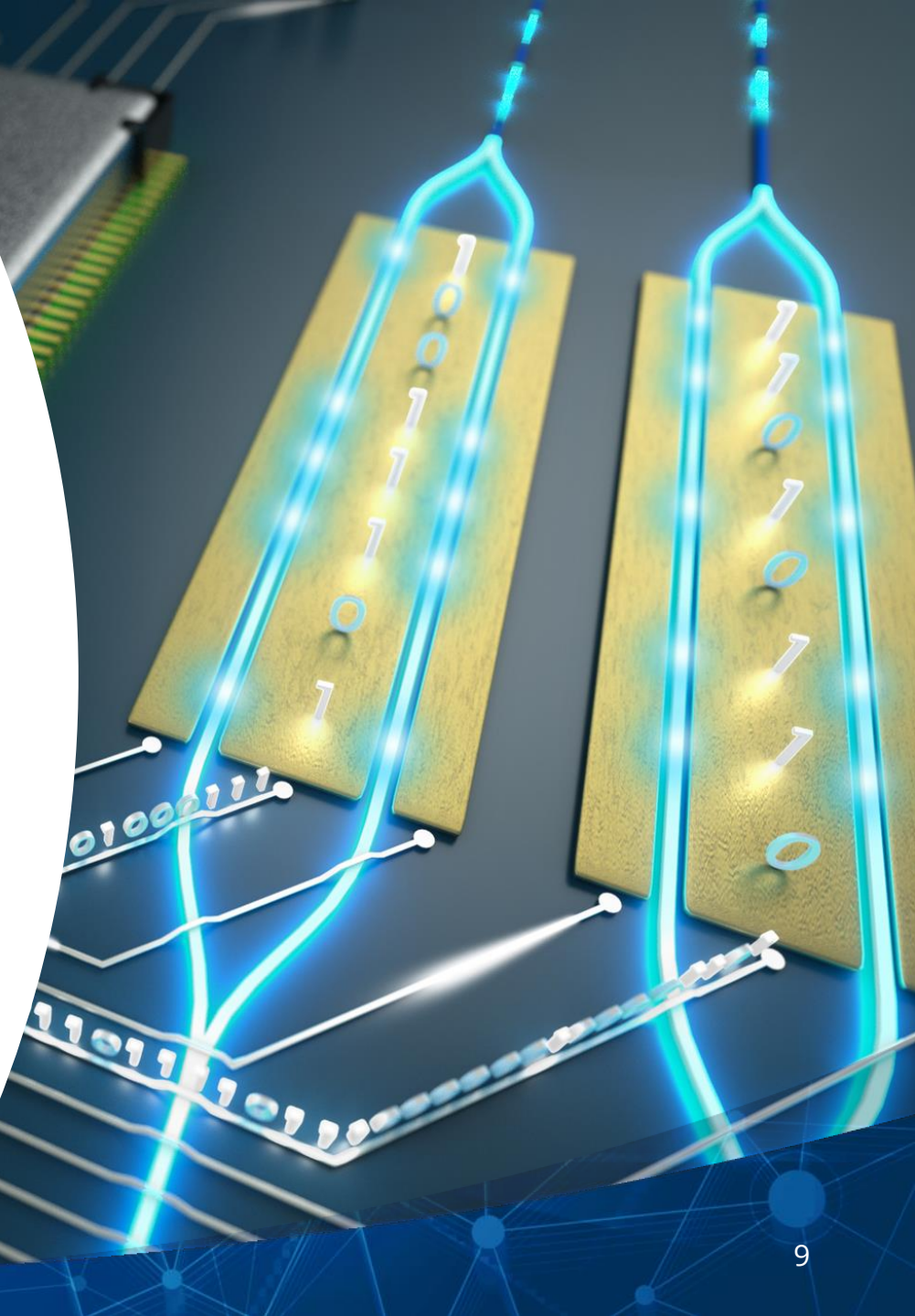
TIP is NSF's newest directorate. Supports use-inspired research, accelerating development of key technologies, and expanding the STEM workforce

- **Convergence Accelerator:** phased funding model to accelerate solutions toward societal impact (RFI for topics => Workshops => topic selection for project cohorts)
  - *2022 workshop examples:* Ethical Design of AI, Computing Solutions for Climate-Driven Extreme Events
  - *Past cohort examples:* Open Knowledge Networks, AI-driven Innovation via Data & Model Sharing

- **Pathways to Enable Open-Source Ecosystems (POSE)** (NSF 22-552)
  - Supports efforts toward harnessing open-source development approaches for new technology solutions to problems of national and societal importance
  - Applicable to a wide range of "open-source" development (including software and other products)



Student Training in AI



NSF's Convergence Accelerator is funding a use-inspired training opportunity to tackle climate challenges with AI

The Institute's theme for the 2023 CORE Fellows is "Tackling Climate-Induced Challenges with AI"

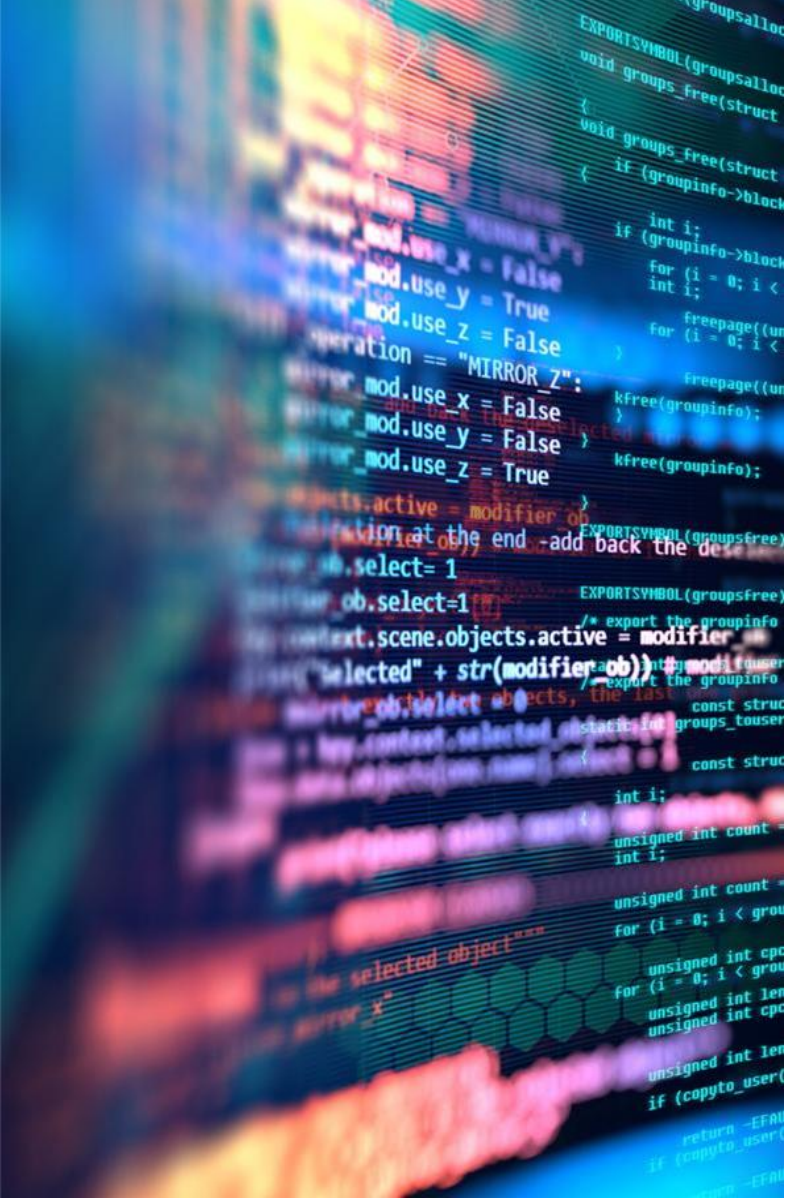
Due: January 8, 2023

https://www.sdsc.edu/education_and_training/core_institute.html?utm_medium=email&utm_source=govdelivery

At AGU

- **NSF Booth** (#1313 in Exhibit Hall): lightning talks (times are tentative)
  - *Data Management & Public Access:* 12/12 at 4:50 PM CST and 12/14 at 11:50 AM CST
  - *Antarctic & Arctic Sciences - New Data Policy:* 12/13 at 3:00 PM CST, 12/14 at 3:00 PM CST
  - *The Evolving Landscape of Open Science in Federally funded Research:* 12/14 at 4:50 PM CST
  - *Mid-Scale Infrastructure Funding Opportunities*: 12/15 10:20 AM CST

GEO Cyberinfrastructure Opportunities



DeCODER

Democratized Cyberinfrastructure for Open Discovery to Enable Research

Kenton McHenry
University of Illinois Urbana-Champaign

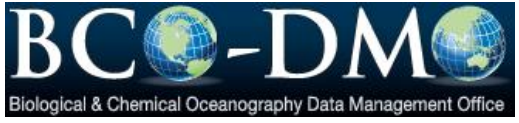
December 13, 2022

This work is supported through the NSF award #2209863.

NSF EarthCube Program

- NSF Program started around 2012 to advance cyberinfrastructure in support of geoscience research
  - Develop tools, data resources, infrastructure, standards
    - Data sharing, data interface standards, data wrangling, workflows, analytics, reproducibility
  - Funded roughly 95 efforts to date

Council of Data Facilities



Science on Schema

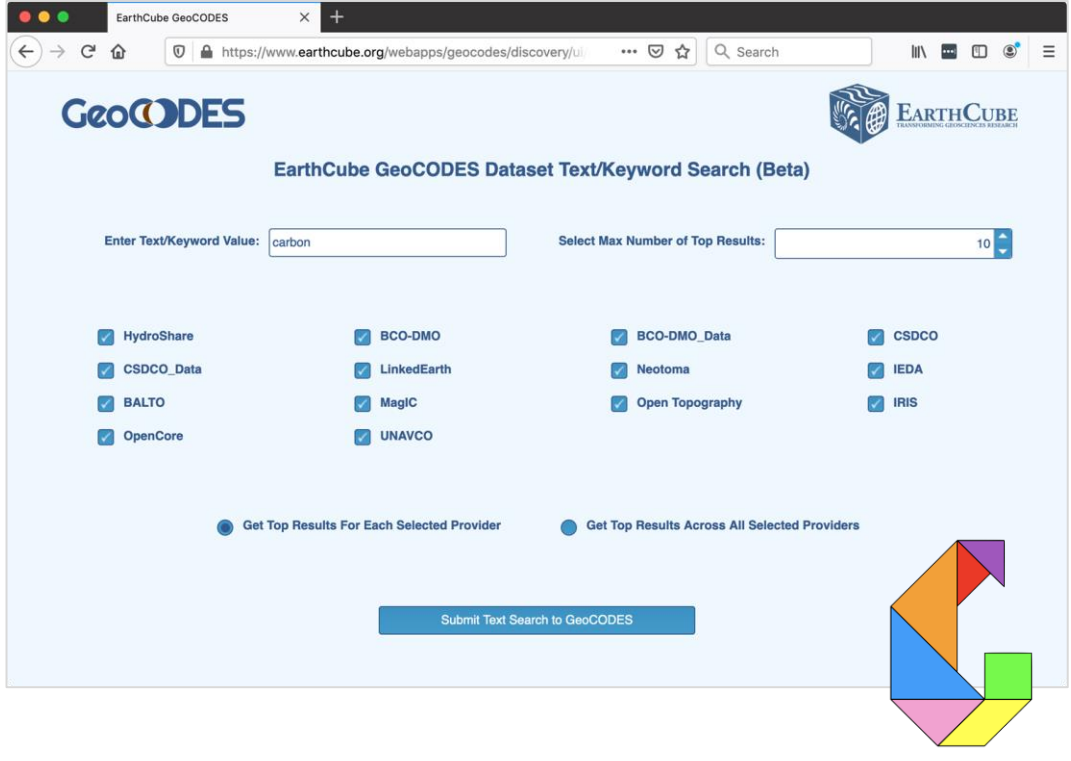
```
<html>
<head>
  <title>NCDC Storm Events Database</title>
  <script type="application/ld+json">
  {
    "@context":"https://schema.org/",
    "@type":"Dataset",
    "name":"NCDC Storm Events Database",
    "description":"Storm Data is provided by the National Weather Service (NWS) and contain statist
    "url":"https://catalog.data.gov/dataset/ncdc-storm-events-database",
    "sameAs":"https://gis.ncdc.noaa.gov/geoportal/catalog/search/resource/details.page?id=gov.noaa.
    "identifier": ["https://doi.org/10.1000/182",
                   "https://identifiers.org/ark:/12345/fk1234"],
    "keywords":[
       "ATMOSPHERE > ATMOSPHERIC PHENOMENA > CYCLONES",
       "ATMOSPHERE > ATMOSPHERIC PHENOMENA > DROUGHT",
       "ATMOSPHERE > ATMOSPHERIC PHENOMENA > FOG",
       "ATMOSPHERE > ATMOSPHERIC PHENOMENA > FREEZE"
    ],
    "license" : "https://creativecommons.org/publicdomain/zero/1.0/",
    "hasPart" : [
      {
        "@type": "Dataset",
        "name": "Sub dataset 01",
        "description": "Informative description of the first subdataset...",
        "license" : "https://creativecommons.org/publicdomain/zero/1.0/"
      },
      {
        "@type": "Dataset",
```

https://github.com/ESIPFed/science-on-schema.org

GeoCODES

- Work with repositories to adopt standard means of sharing metadata
- Pilot crawler and demonstration portal for geoscience data

Council of Data Facilities

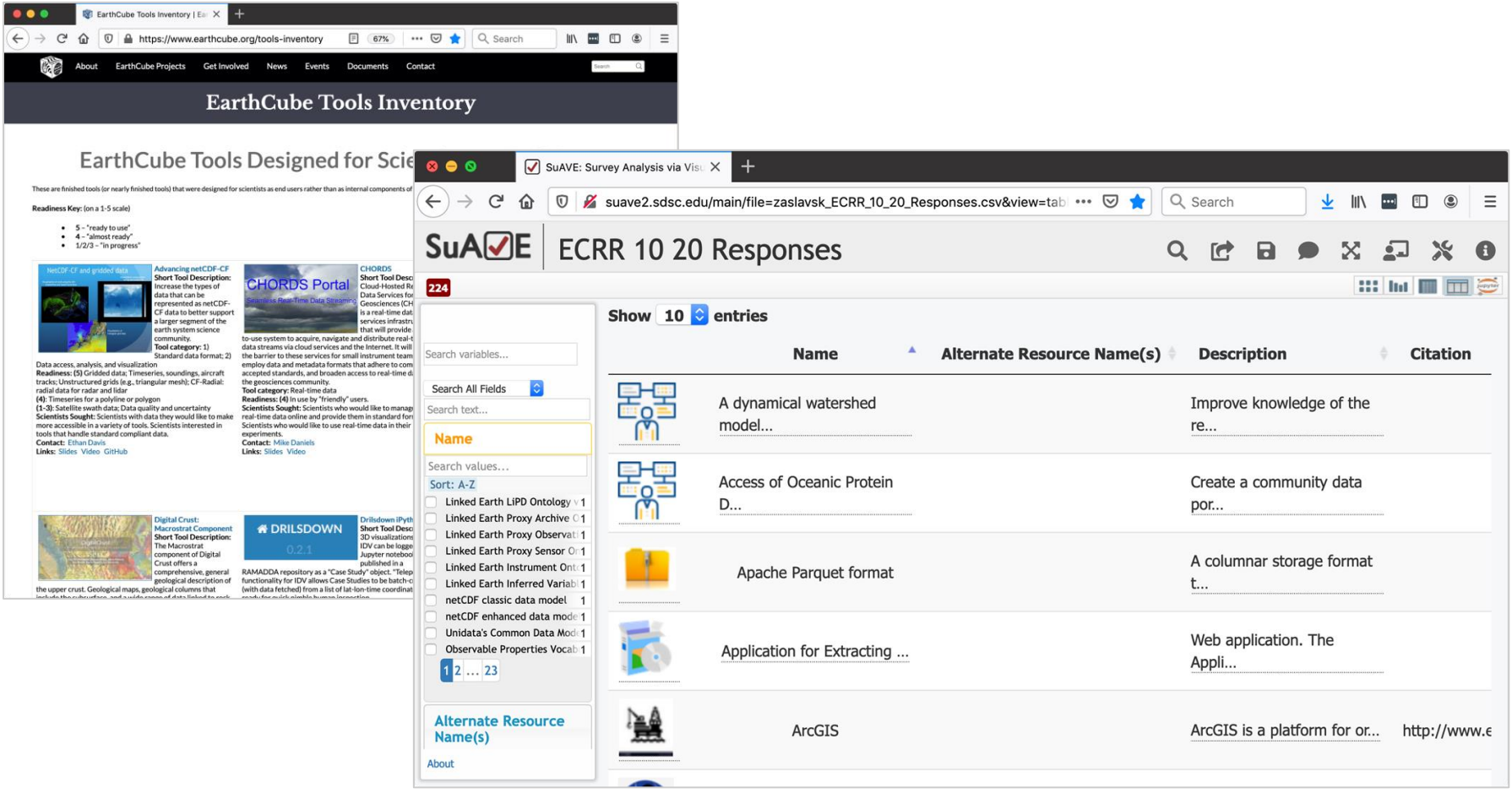


EarthCube

NSF

Repositories Using Science on Schema

- 23 repositories: AquaDocs, OpenTopography, IRIS, Environmental Data Initiative, Biological and Chemical Oceanography Data Management Office, Consortium of Universities for the Advancement of Hydrologic Science, Inc.(CUAHSI), IEDA (Integrated Earth Data Applications), UNAVCO, IODP Site Survey Databank, Balto, Linked Earth, Linked PaleoData, IRIS, UCAR, opencoredata, Magnetics Information Consortium (MagIC), Neotoma, earthchem, xdomes, National Ecological Observatory Network (NEON), Resource Registry, UNIDATA, Rolling Deck to Repository Program (R2R), Geocodes Demo Datasets, U.S. Antarctic Program Data Center
- Currently working with 3 others: Decade, CReSIS, and DesignSafe
- Engaging with Google to support analytics within Data Commons



https://www.earthcube.org/tools-inventory

http://resourceregistry.earthcube.org

Call for Notebooks - Abstracts Due April 15

As scientific studies become more data intensive and software dependent, reproducibility principles and other factors increase the importance of citable publications that include reusable workflows, software, and data-access procedures. This importance is reflected in new academic journals, such as the Journal of Open Source Software, whose peer reviewed articles highlight the software itself, and often can include executable notebooks (Jupyter , R Studio, etc.). In this spirit, EarthCube is issuing its first call for Notebooks as primary, peer-reviewed submissions to a digital proceedings for this year's EarthCube Annual Meeting. Submitted notebooks should highlight a tool (i.e. software, service, library, dataset, standard), explaining—and demonstrating interactively-—how the tool may be used to address a significant problem in geoscience.

- Drew Camron (UNIDATA)
- Julien Chastang (UNIDATA)
- Jeff Dozier (UCSB)
- David Fulker (OPeNDAP)
- Ryan Gooch (NREL)
- Joseph Hardin (PNNL)
- Keith Maull (NCAR)
- Kenton McHenry (NCSA)
- Chris Olson (UCSD)
- Stephen Richard (U.S. Geoscience Information Network)
- Bradley Spitzbart (Stony Brook University)
- Lisa Tauxe (Scripps Institution of Oceanography)
- Carol Willing (Project Jupyter)
- Ilya Zaslavsky (UCSD)

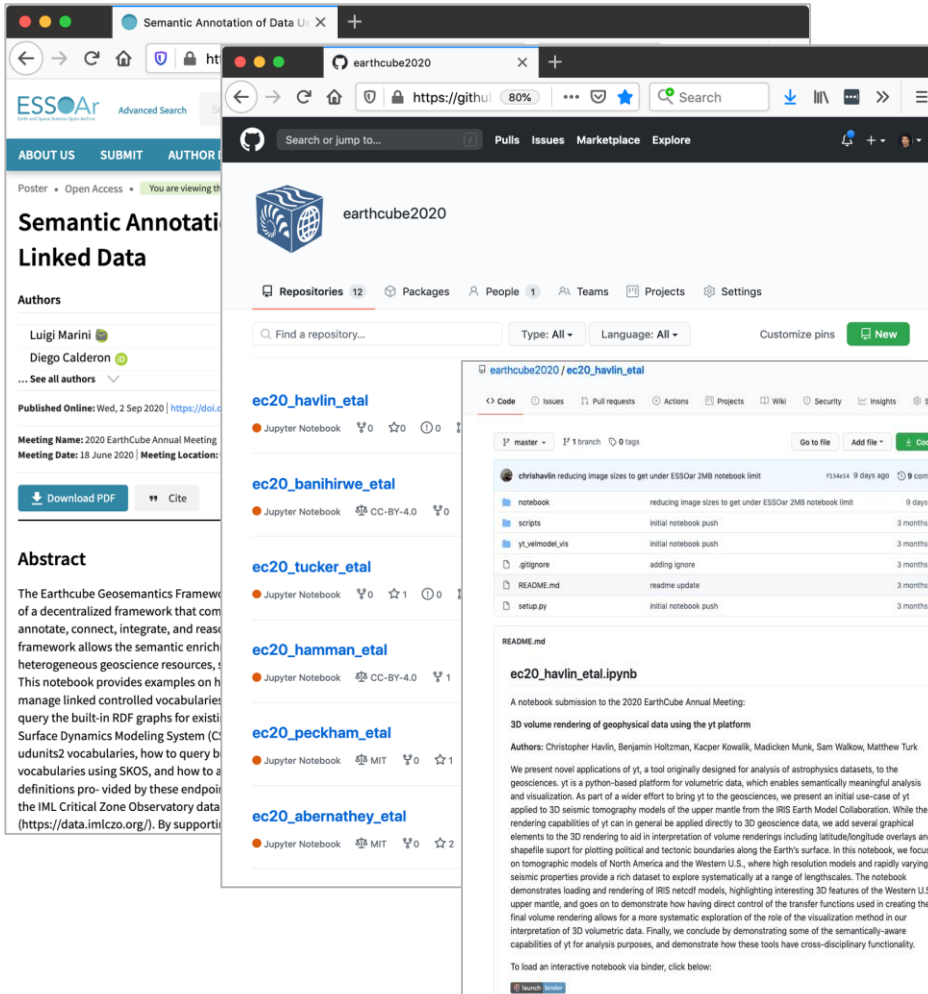



JOSS

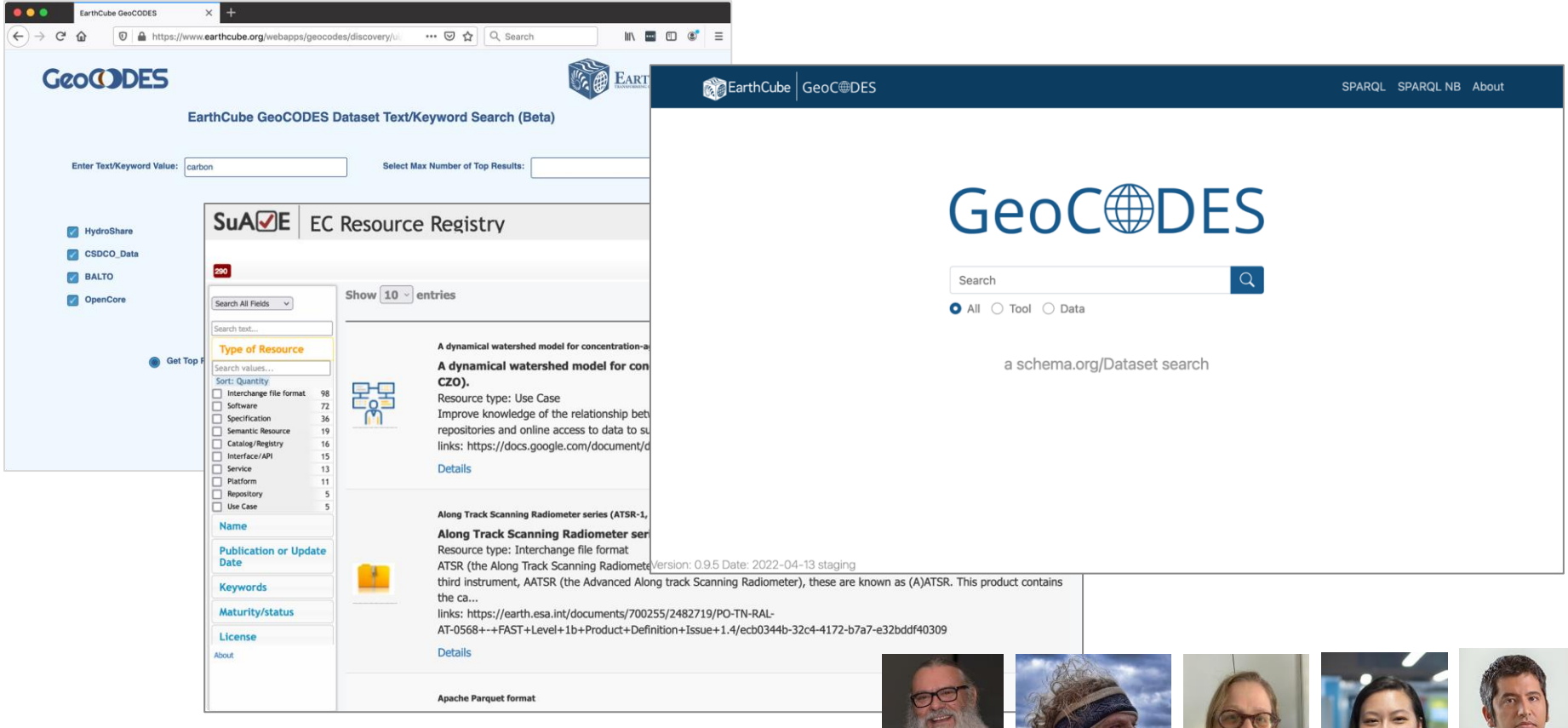
EarthCube

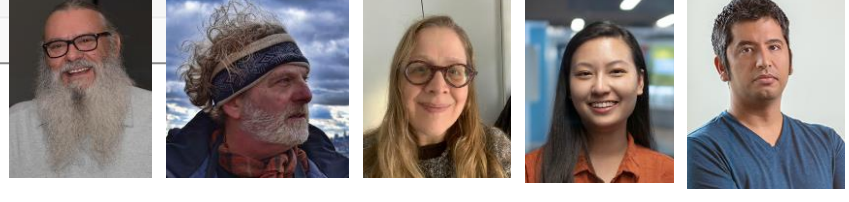
CfN’20

- https://github.com/earthcube2020
  https://www.essoar.org
- 21 submissions
  - 2-3 reviewers per notebook
  - 12 accepted, 5 as oral presentations
  - Published through AGU ESSOAr system
- Insights
  - Types of notebooks
  - Runnability
  - An emphasis on testing
  - Science Tool focused vs Science focused submissions
  - Formatting Guidelines

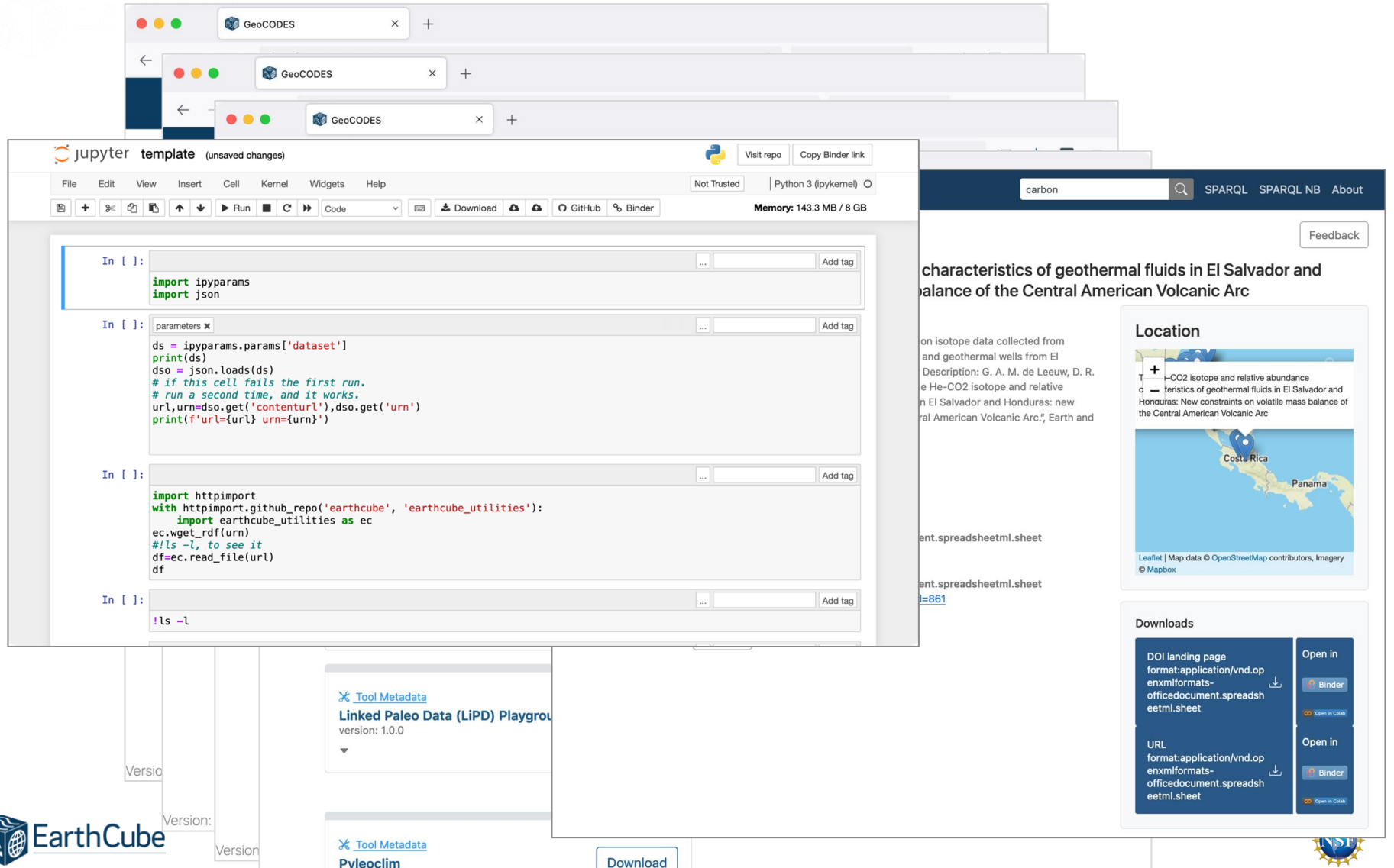


GeoCODES Data + Tools



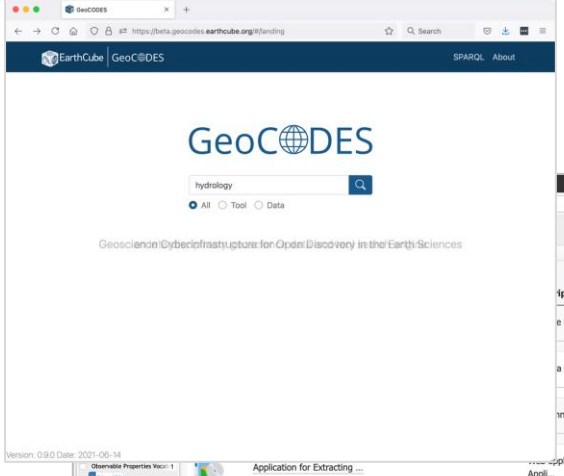


EarthCube https://geocodes.earthcube.org


GeoCODES
GeoCODES
GeoCODES
jupyter template (unsaved changes)
Visit repo
Copy Binder link
File Edit View Insert Cell Kernel Widgets Help
Not Trusted
Python 3 (ipykernel)
Run
Code
Download
GitHub
Binder
Memory: 143.3 MB / 8 GB
In [ ]:
Add tag
import ipyparams
import json
In [ ]:
parameters
ds = ipyparams.params['dataset']
print(ds)
dso = json.loads(ds)
# if this cell fails the first run.
# run a second time, and it works.
url,urn=dso.get('contenturl'),dso.get('urn')
print(f'url={url} urn={urn}')
In [ ]:
import httpimport
with httpimport.github_repo('earthcube', 'earthcube_utilities'):
import earthcube_utilities as ec
ec.wget_rdf(urn)
#!ls -l, to see it
df=ec.read_file(url)
df
In [ ]:
!ls -l
carbon
SPARQL
SPARQL NB
About
Feedback
characteristics of geothermal fluids in El Salvador and
balance of the Central American Volcanic Arc
Location
CO2 isotope and relative abundance
teristics of geothermal fluids in El Salvador and
Honduras: New constraints on volatile mass balance of
the Central American Volcanic Arc
Costa Rica
Panama
Leaflet | Map data © OpenStreetMap contributors, Imagery
© Mapbox
Downloads
DOI landing page
format:application/vnd.openxmlformats-officedocument.spreadsheetml.sheet
Open in
Binder
URL
format:application/vnd.openxmlformats-officedocument.spreadsheetml.sheet
Open in
Binder
Tool Metadata
Linked Paleo Data (LiPD) Playgrou
version: 1.0.0
Tool Metadata
Pyleoclim
Download
Version:
Version
EarthCube

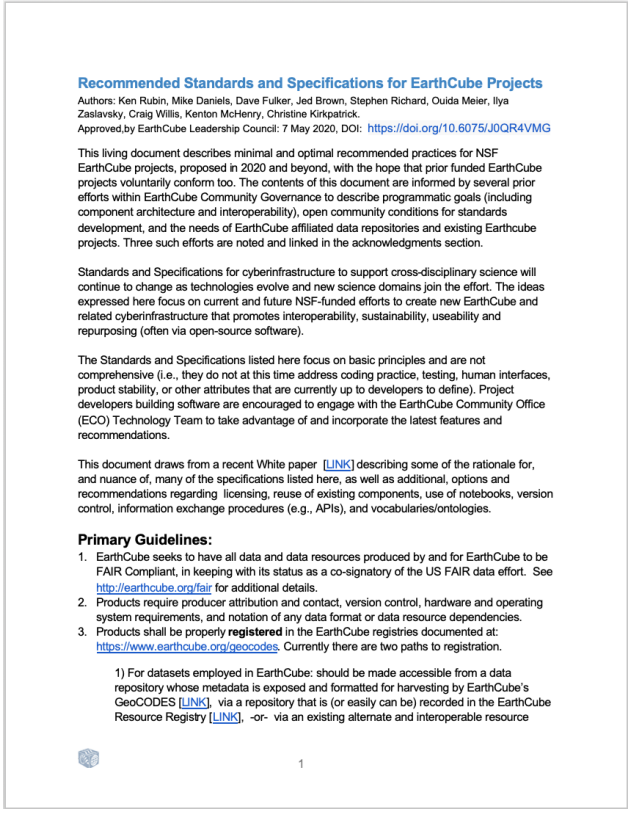
Recommended Standards and Specifications for EarthCube Projects

K. Rubin, M. Daniels, D. Fulker, J. Brown, S. Richard, O. Meier, I. Zaslavsky, C. Willis, K. McHenry, C. Kirkpatrick, "**Recommended Standards and Specifications for EarthCube Projects**", 2020
https://doi.org/10.6075/J0QR4VMG



⟶ ***Best Practices!***
F.A.I.R.

- Schema.org
- GitHub,
- Markdown,
- Docker,
- Binder,
- Jupyter,
- ...



Recommended Standards and Specifications for EarthCube Projects

Authors: Ken Rubin, Mike Daniels, Dave Fulker, Jed Brown, Stephen Richard, Ouida Meier, Ilya Zaslavsky, Craig Willis, Kenton McHenry, Christine Kirkpatrick.
Approved by EarthCube Leadership Council: 7 May 2020, DOI: https://doi.org/10.6075/J0QR4VMG

This living document describes minimal and optimal recommended practices for NSF EarthCube projects, proposed in 2020 and beyond, with the hope that prior funded EarthCube projects voluntarily conform too. The contents of this document are informed by several prior efforts within EarthCube Community Governance to describe programmatic goals (including component architecture and interoperability), open community conditions for standards development, and the needs of EarthCube affiliated data repositories and existing Earthcube projects. Three such efforts are noted and linked in the acknowledgments section.

Standards and Specifications for cyberinfrastructure to support cross-disciplinary science will continue to change as technologies evolve and new science domains join the effort. The ideas expressed here focus on current and future NSF-funded efforts to create new EarthCube and related cyberinfrastructure that promotes interoperability, sustainability, useability and repurposing (often via open-source software).

The Standards and Specifications listed here focus on basic principles and are not comprehensive (i.e., they do not at this time address coding practice, testing, human interfaces, product stability, or other attributes that are currently up to developers to define). Project developers building software are encouraged to engage with the EarthCube Community Office (ECO) Technology Team to take advantage of and incorporate the latest features and recommendations.

This document draws from a recent White paper [LINK] describing some of the rationale for, and nuance of, many of the specifications listed here, as well as additional, options and recommendations regarding licensing, reuse of existing components, use of notebooks, version control, information exchange procedures (e.g., APIs), and vocabularies/ontologies.

Primary Guidelines:

1. EarthCube seeks to have all data and data resources produced by and for EarthCube to be FAIR Compliant, in keeping with its status as a co-signatory of the US FAIR data effort. See http://earthcube.org/fair for additional details.
2. Products require producer attribution and contact, version control, hardware and operating system requirements, and notation of any data format or data resource dependencies.
3. Products shall be properly **registered** in the EarthCube registries documented at: https://www.earthcube.org/geocodes. Currently there are two paths to registration.

    1) For datasets employed in EarthCube: should be made accessible from a data repository whose metadata is exposed and formatted for harvesting by EarthCube's GeoCODES [LINK], via a repository that is (or easily can be) recorded in the EarthCube Resource Registry [LINK], -or- via an existing alternate and interoperable resource

1

EarthCube

Potential of GeoCODES

Ocean InfoHub Project

Repositories crawled and indexed

Environmental Data Initiative

28248 records

POLDER: Polar Data Discover

Polder (Dutch verb): to work collabor

UNAVCO promotes research by providing access to data that our community of geodetic scientists uses for quantifying the motions of rock, ice and water that are monitored by a variety of sensor types at or near the Earth… data enable millimeter-scale surface motion detection and monitoring at discrete points, and high-resolution strain imagery over areas of tens of square meters to hundreds of square kilometers. The data types… such as from SAR and TLS, strain and seismic borehole data, and meteorological data. Most of these can be accessed via web services. In addition, GPS/GNSS datasets, TLS datasets, and InSAR products are assigned digital object identifiers.

4136 records

Magnetics Information Consortium…

Council of Data Facilities

CCADI

Neotoma

11955 records

BCO-DMO

Internet of Water

Version: 0.9.2 Date: 2021-09-23

EarthCube

Computers & Geosciences 157 (2021) 104933

Contents lists available at ScienceDirect

Computers and Geosciences

journal homepage: www.elsevier.com/locate/cageo

The future low-temperature geochemical data-scape as envisioned by the U. S. geochemical community

Susan L. Brantley [a,o,*], Tao Wen [b], Deborah A. Agarwal [c], Jeffrey G. Catalano [d], Paul A. Schroeder [e], Kerstin Lehnert [f], Charuleka Varadharajan [g], Julie Pett-Ridge [h], Mark Engle [i], Anthony M. Castronova [j], Richard P. Hooper [k], Xiaogang Ma [l], Lixin Jin [j], Kenton McHenry [m], Emma Aronson [n], Andrew R. Shaughnessy [o], Louis A. Derry [p], Justin Richardson [q], Jerad Bales [j], Eric M. Pierce [r]

[a] Earth and Environmental Systems Institute, The Pennsylvania State University, University Park, PA, USA
[b] Department of Earth and Environmental Sciences, Syracuse University, Syracuse, NY, USA
[c] Advanced Computing for Science Department, Lawrence Berkeley National Laboratory, Berkeley, CA, USA
[d] Department of Earth and Planetary Sciences, Washington University, St. Louis, MO, USA
[e] Department of Geology, University of Georgia, Athens, GA, USA
[f] Lamont-Doherty Earth Observatory, Columbia University, Palisades, NY, USA
[g] Earth and Environmental Sciences Area, Lawrence Berkeley National Laboratory, Berkeley, CA, USA
[h] Department of Crop and Soil Science, Oregon State University, Corvallis, OR, USA
[i] Department of Geological Sciences, The University of Texas at El Paso, El Paso, TX, USA
[j] Consortium of Universities for the Advancement of Hydrological Science, Inc, Cambridge, MA, USA
[k] Department of Civil and Environmental Engineering, Tufts University, Medford, MA, USA
[l] Department of Computer Science, University of Idaho, Moscow, ID, USA
[m] National Center for Supercomputing Applications, University of Illinois, Urbana, IL, USA
[n] Department of Microbiology and Plant Pathology, University of California, Riverside, USA
[o] Department of Geosciences, The Pennsylvania State University, University Park, PA, USA
[p] Department of Earth and Atmospheric Sciences, Cornell University, Ithaca, NY, USA
[q] Department of Geosciences, University of Massachusetts Amherst, Amherst, MA, USA
[r] Environmental Sciences Division, Oak Ridge National Laboratory, Oak Ridge, TN, USA

ARTICLE INFO

Keywords:
Data management
Data repositories
Geochemistry
Metadata
Data sharing
Open science

ABSTRACT

Data sharing benefits the researcher, the scientific community, and the public by allowing the impact of data to be generalized beyond one project and by making science more transparent. However, many scientific communities have not developed protocols or standards for publishing, citing, and versioning datasets. One community that lags in data management is that of low-temperature geochemistry (LTG). This paper resulted from an initiative from 2018 through 2020 to convene LTG and data scientists in the U.S. to strategize future management of LTG data. Through webinars, a workshop, a preprint, a townhall, and a community survey, the group of U.S. scientists discussed the landscape of data management for LTG – the data-scape. Currently this data-scape includes a "street bazaar" of data repositories. This was deemed appropriate in the same way that LTG scientists publish articles in many journals. The variety of data repositories and journals reflect that LTG scientists target many different scientific questions, produce data with extremely different structures and volumes, and utilize copious and complex metadata. Nonetheless, the group agreed that publication of LTG science must be accompanied by sharing of data in publicly accessible repositories, and, for sample-based data, registration of samples with globally unique persistent identifiers. LTG scientists should use certified data repositories that are either highly structured databases designed for specialized types of data, or unstructured generalized data systems. Recognizing the need for tools to enable search and cross-referencing across the proliferating data repositories, the group proposed that the overall data informatics paradigm in LTG should shift from "build data repository, data will come" to "publish data online, cybertools will find". Funding agencies could also provide portals for LTG scientists to register funded projects and datasets, and forge approaches that cross national

* Corresponding author. Earth and Environmental Systems Institute and Department of Geosciences, The Pennsylvania State University, University Park, PA, USA.
E-mail address: sxb7@psu.edu (S.L. Brantley).

https://doi.org/10.1016/j.cageo.2021.104933
Received 31 January 2021; Received in revised form 26 July 2021; Accepted 3 September 2021
Available online 7 September 2021
0098-3004/

NSF CSSI Frameworks: DeCODER

Kenton McHenry
Cyberinfrastructure
ILLINOIS
UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN

Christine Kirkpatrick
Cyberinfrastructure
UC San Diego

Karen Stocks
Ocean Science
UC San Diego

Tao Wen
Hydrochemistry
Syracuse University

Shuang Zhang
Biochemical Cycles
TEXAS A&M UNIVERSITY

Thomas Quinn
Ecology
VIRGINIA TECH

Carl Boettiger
Ecology
Berkeley
UNIVERSITY OF CALIFORNIA

Lynne Schreiber
Coordinator
UC San Diego

EarthCube

Geoscience Cyberinfrastructure for Open Discovery in the Earth Sciences (GeoCODES)

↓

Democratized Cyberinfrastructure for Open Discovery to Enable Research (DeCODER)

NSF

DeCODER

CSSI: Democratized Cyberinfrastructure for Open Discovery to Enable Research - $3,199,896

https://nsf.gov/awardsearch/showAward?AWD_ID=2209863 - $900,640 (UIUC)

https://nsf.gov/awardsearch/showAward?AWD_ID=2209864 - $460,281 (Syracuse)

https://nsf.gov/awardsearch/showAward?AWD_ID=2209865 - $1,303,971 (UCSD)

https://nsf.gov/awardsearch/showAward?AWD_ID=2209866 - $535,004 (Virginia)

Oct 1, 2022 - Sept. 30, 2026 (4 years)

RCN: Disciplinary Improvements: AI Readiness, Reproducibility, and FAIR: Connecting Computing and Domain Communities Across the ML Lifecycle

https://nsf.gov/awardsearch/showAward?AWD_ID=2226453 - $1,260,000 (UCSD)

August 15, 2022 - July 31, 2025 (3 years)

Tasks

- Gleaner community
- Extending Schema.org
  - Depth, horizon, …
- Portal customization and deployments
- Graph Search
  - Data Integration
- Linking tools with data
  - Notebooks as scholarly objects
  - Crawling tools
- Community Support & Engagement
  - Support for smaller repositories

Opportunities for EarthCube Projects & Community

- Advice and jump start help on adopting schema.org for repositories
- Register your geo CI resources and associate with data sets and notebooks
- Provide use cases for our continual UI/UX process
  - test DeCODER against your own scientific questions and use cases
- Get credit for other types of scholarship with our annual notebook competition
  - encourage your students to take part!
  - be a reviewer

FARR:

FAIR in ML, AI Readiness, & Reproducibility Research Coordination Network

FARR
FAIR, AI Readiness & Reproducibility

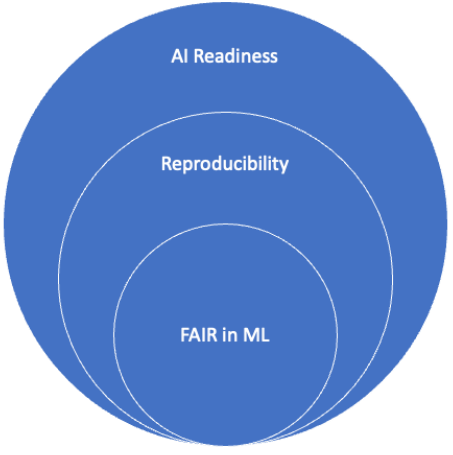
Christine Kirkpatrick
SDSC, UC San Diego

Date: December 13, 2022

SDSC SAN DIEGO SUPERCOMPUTER CENTER
NC STATE UNIVERSITY
ILLINOIS NCSA | National Center for Supercomputing Applications
UC San Diego
SCRIPPS INSTITUTION OF OCEANOGRAPHY

This work is supported through the NSF award #2226453.

Motivation



- If 80% of time with data is wrangling, can FAIR principles increase efficiency of people and machines?
- What does it mean to be AI ready as a repository or organization?
- What roles can repositories play in AI readiness?
- What are best practices in AI reproducibility and what are the gaps in current knowledge?

FARR Goals and Activities

We welcome individual researchers, institutions/organizations, CI providers, repositories/facilities, and networks of facilities in Computer Science, Geosciences, and the 'Research Data' Community. FARR provides a neutral and novel meeting place for bridging multiple networks.

- Building communities to
  - promote better practices for AI
  - harness community efforts
  - improve efficiency and reproducibility
  - stimulate and enhance new research
- Activities will include
  - workshops
  - assessing community needs
  - fostering new collaborations (proposals)
  - setting research agendas
  - community-led reports

Incorporating EarthCube's sustainability lessons from the beginning.

EarthCube

NSF

FARR Team

PI, Christine Kirkpatrick

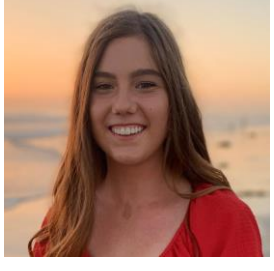
Co-PI, Karen Stocks

Co-PI, Yuhan (Douglas) Rao

Co-PI, Daniel Katz

Project Director,
Lynne Schreiber

Sr. Personnel,
Kevin Coakley



Project Manager,
Julie Christopher

Communications,
Kim Mann Bruch

SDSC SAN DIEGO SUPERCOMPUTER CENTER

NC STATE UNIVERSITY

ILLINOIS
NCSA | National Center for Supercomputing Applications

UC San Diego

SCRIPPS INSTITUTION OF OCEANOGRAPHY

EarthCube

NSF

Opportunities for EarthCube Projects & Community

- **Interviews** with CDF members and other data repositories on AI readiness topics
- **Input** on community needs, gaps, and roadmap
- **FARR network meetings** and **early career travel scholarships**
- Connect with FARR via
  - ESIP's Data readiness and ML **clusters** - ESIP Winter (virtual)
  - Research Data Alliance (RDA) FAIR4ML **interest group** - RDA P20 (Sweden/virtual)
  - FARR hosted **webinars**
  - Sign up for our **newsletter**
- Suggest **use cases** and let us **promote your project's use of AI** and FARR-related practices
- Let us feature you in a **science story**

For more info, contact us at community@farr-rcn.org

EarthCube

NSF

Funded Project Lightning Talks

- John Clyne (Project Raijin)
- Marshall Ma (OpenMindat)

PROJECT raijin

Community Geoscience Analysis Tools for Unstructured Grids

John Clyne[1], Orhan Eroglu[1], Brian P. Medeiros[1], Colin Zarzycki[2]
[1] National Center for Atmospheric Research (NCAR)
[2] Pennsylvania State University

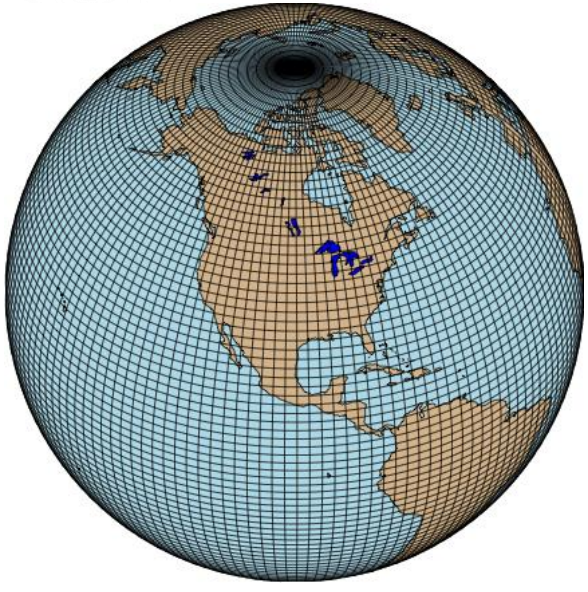
NCAR PennState

NSF Earth Cube Town Hall
December 13, 2022

EarthCube

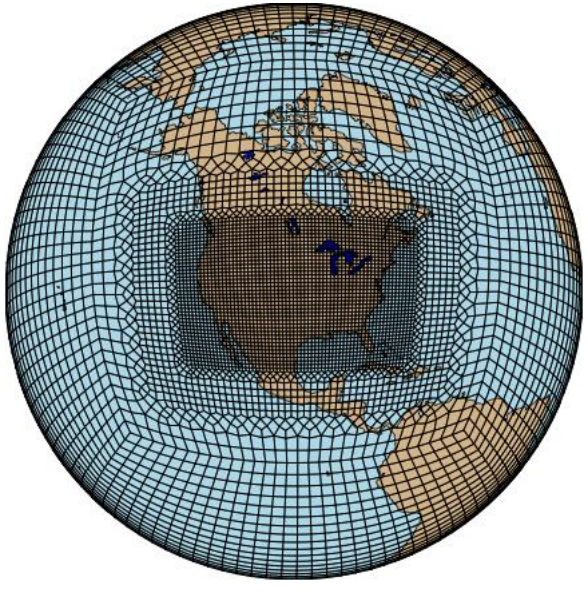
This poster presentation is based upon work supported by the National Science Foundation under Grant No.2126458

NSF



“Lat-lon” structured grid

Icosahedral grid
(MPAS-A)



Variable resolution, cube sphere grid
(CAM-SE)

After nearly two decades of development and evaluation, the climate and global weather modeling communities are transitioning from more simple structured grids to more complex, but scalable unstructured grids upon which governing equations of state are solved.

Problem?

1. No widely used convention for the storage of unstructured grid data
   - UGRID conventions: https://ugrid-conventions.github.io

1. Few analysis tools capable of working directly with unstructured data
   - Resampling to structured grids has numerous pitfalls

1. Global storm resolving resolution models are capable of generating LOTS of data
   - Further exacerbating problems with limited set of tools that operate directly on unstructured meshes

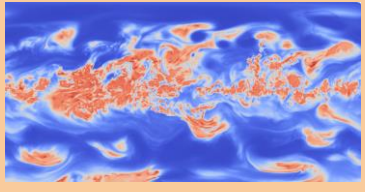
Project Raijin Goals

Community developed software for analysis on unstructured grids

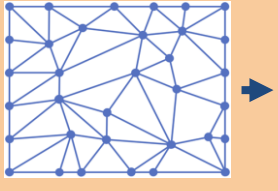
Computational operators

$$\mathcal{F}(x) \quad \nabla \quad \iint \quad \frac{\partial x}{\partial y} \quad \mathcal{L}(x)$$

Plotting



Regridding



+

Core technologies

Data ingest, internal representation, compatibility

Scalability

Versioning, continuous integration & deployment, collaboration

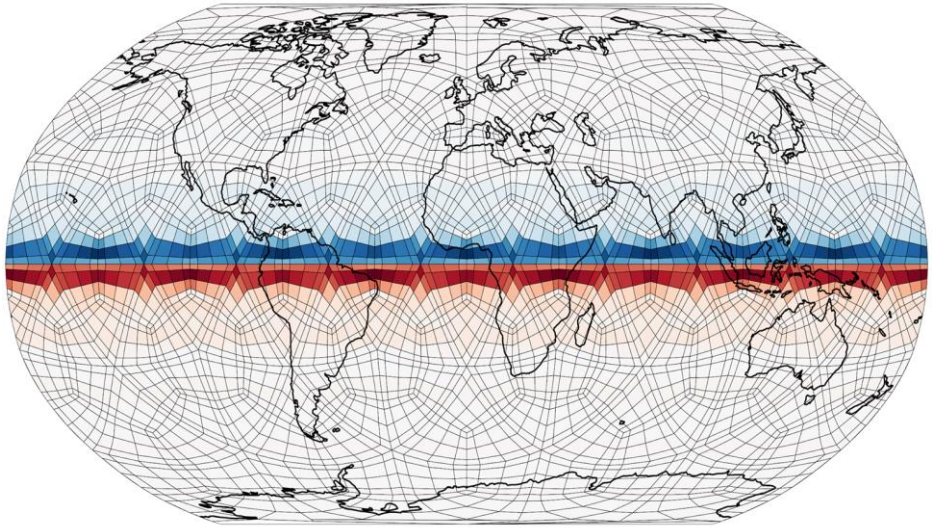
Documentation and examples

Packaging, distribution, testing

User APIs

Community Geoscience Analysis Tools for Unstructured Grids

UXarray for visualization



Get involved!

Send us email
projectraijin@googlegroups.com

Start a discussion
https://github.com/UXARRAY/uxarray/discussions

Find out more
https://raijin.ucar.edu

EarthCube

EarthCube
Transforming Geosciences Research

Acknowledgements

NSF Earth Cube program (award #2126458)

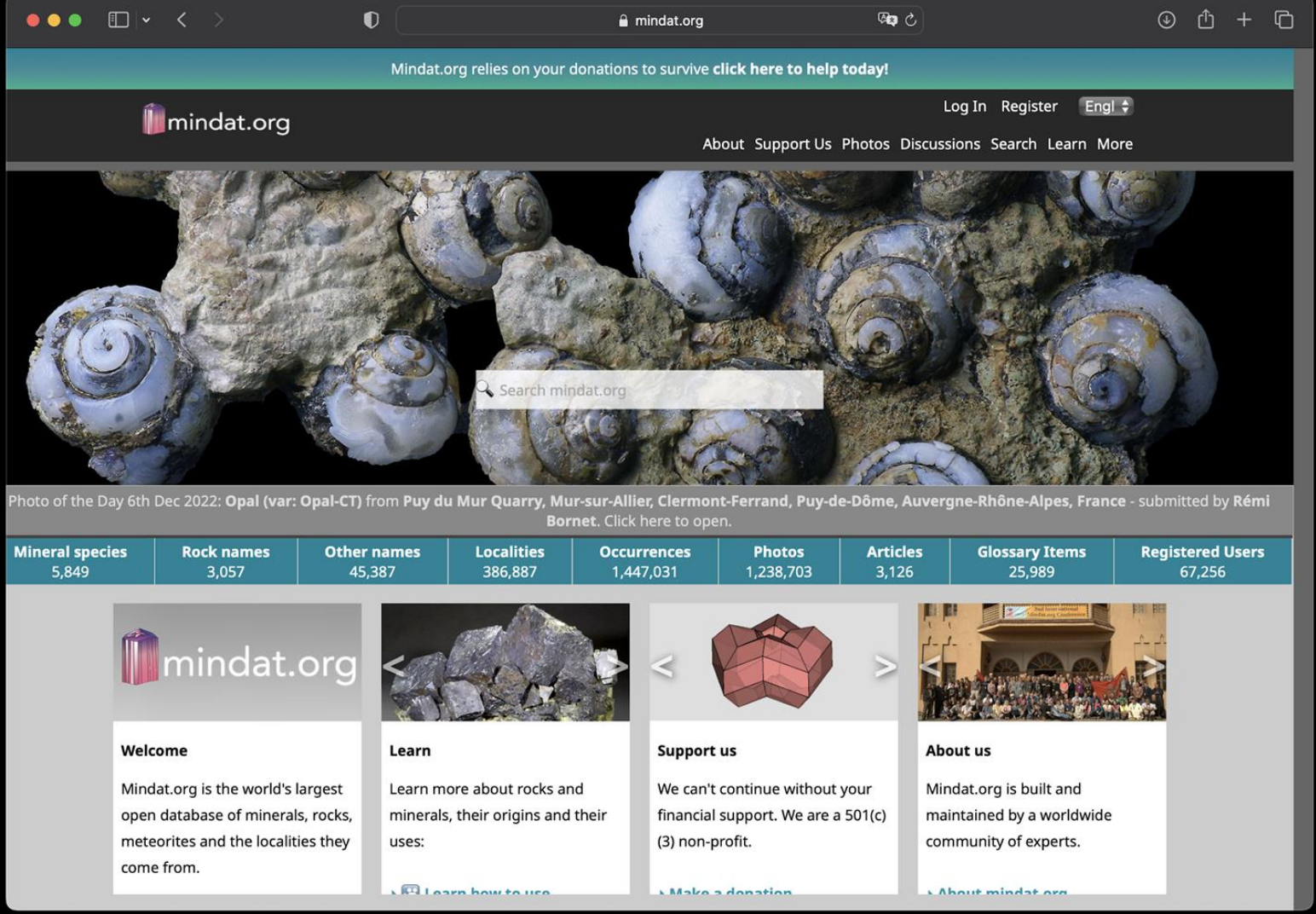
Collaborators: Ryan Abernathy, Falko Judt, David Randall, Niklas Röber, and Bjorn Stevens

Pangeo community

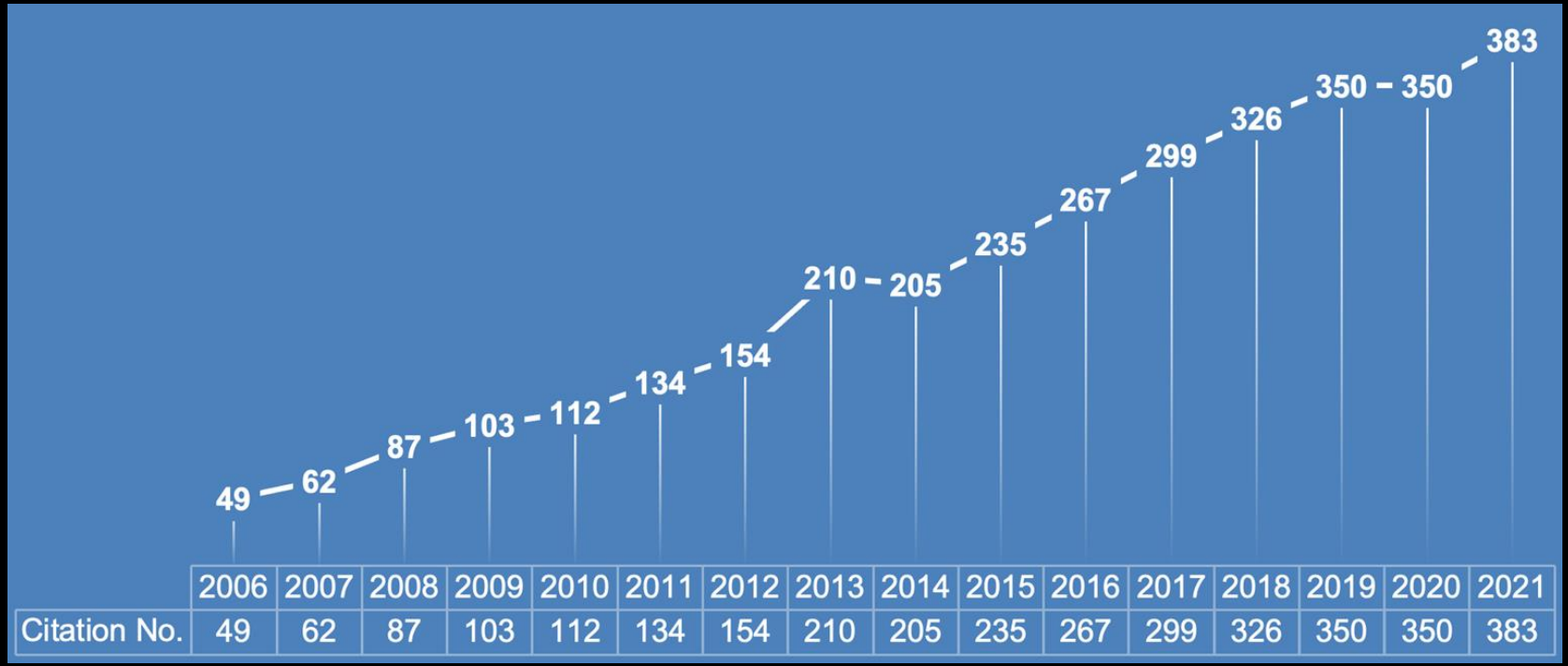
Our growing list of contributors on GitHub!

OpenMindat: machine interface to the world's largest database in mineralogy

Xiaogang (Marshall) Ma
Associate Professor, University of Idaho
Visiting Scientist, Carnegie Institution for Science
max@uidaho.edu | @MarshallXMa


mindat.org
Mindat.org relies on your donations to survive click here to help today!
mindat.org
Log In Register Engl
About Support Us Photos Discussions Search Learn More
Search mindat.org
Photo of the Day 6th Dec 2022: Opal (var: Opal-CT) from Puy du Mur Quarry, Mur-sur-Allier, Clermont-Ferrand, Puy-de-Dôme, Auvergne-Rhône-Alpes, France - submitted by Rémi Bornet. Click here to open.
Mineral species 5,849
Rock names 3,057
Other names 45,387
Localities 386,887
Occurrences 1,447,031
Photos 1,238,703
Articles 3,126
Glossary Items 25,989
Registered Users 67,256
mindat.org
Welcome
Mindat.org is the world's largest open database of minerals, rocks, meteorites and the localities they come from.
Learn
Learn more about rocks and minerals, their origins and their uses:
Support us
We can't continue without your financial support. We are a 501(c)(3) non-profit.
About us
Mindat.org is built and maintained by a worldwide community of experts.

Citations to “mindat.org” on Google Scholar



	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021
Citation No.	49	62	87	103	112	134	154	210	205	235	267	299	326	350	350	383

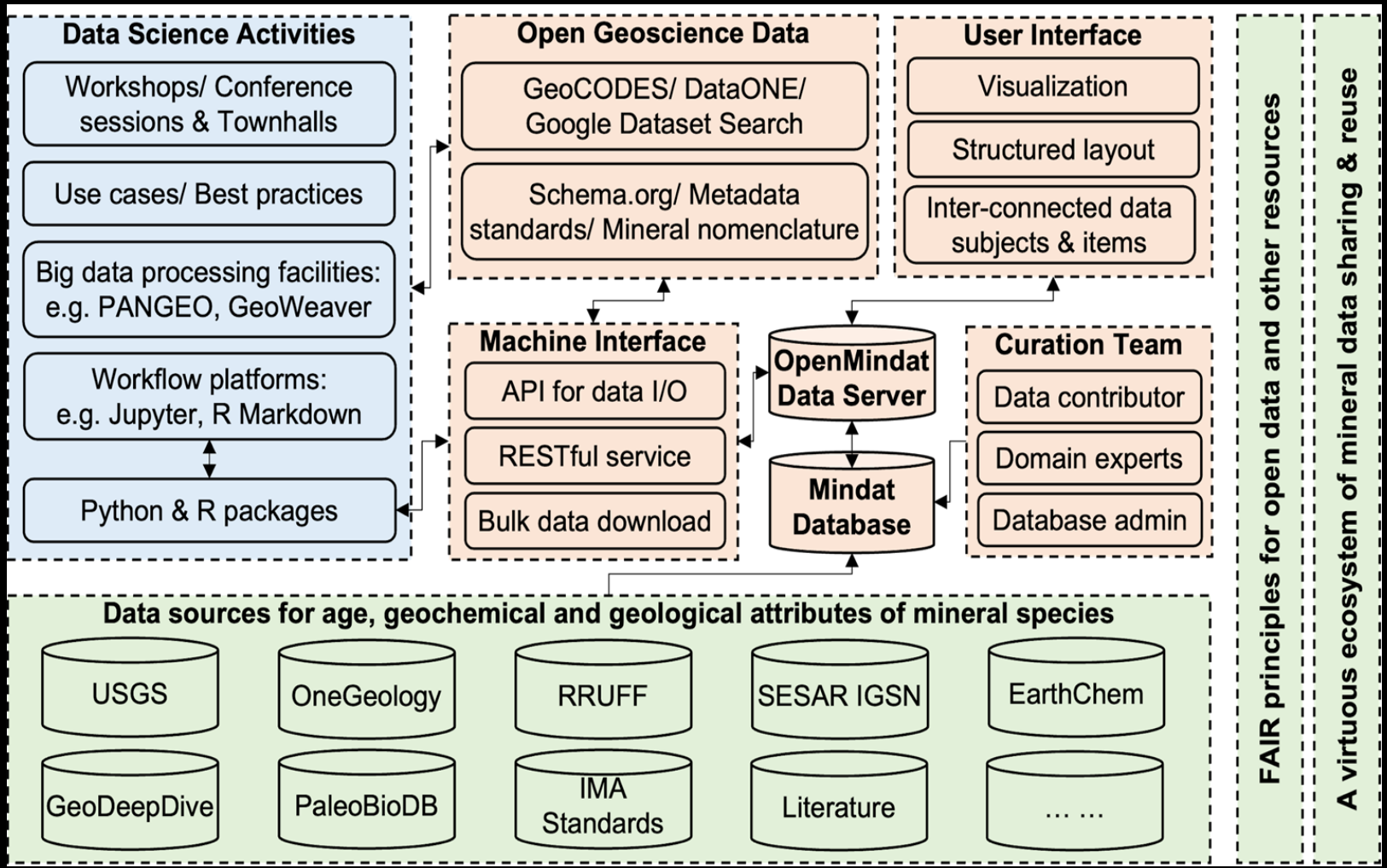
Challenges

Massive needs on data sharing

vs

No machine interface for data access

Designed Structure of OpenMindat



Work Towards Semantic Harmonization

- **Semantics of General Subjects:** For spatial and temporal information and metadata elements for dataset, we refer to existing standards such as those developed by OGC, W3C, and DataCite
- **Community-level Standards/Guidelines**: For geoscience subjects, such as name list and attributes of mineral species and rock classification, we refer to scientific societies' guidelines and widely used standards and vocabularies
- **Leverage Schema.org**: We are designing a metadata schema for annotation of dataset webpages, by following the Schema.org guidelines and best practices shared by ESIP and EarthCube
- **Persistent Identifiers**: We are discussing the possibility of deploying persistent and resolvable identifiers, such as the International Generic Sample Number, for mineral species
- **Collaboration amongst Stakeholders**: We are working together with scientific societies and data users for the design of the API to increase its utility