https://www.earthcube.org/decoder

DeCODER Town Hall:

Use Case Applications for a Democratized Cyberinfrastructure for Open Discovery to Enable Research

AGU 2023
Thursday, 14 December 2023
13:00 - 14:00
2010 - West (Level 2, West, Moscone Center)

We gratefully acknowledge the support provided by the US National Science Foundation under awards 1928208 and 2209864.

ILLINOIS NCSA | National Center for Supercomputing Applications
SDSC SAN DIEGO SUPERCOMPUTER CENTER
SCRIPPS INSTITUTION OF OCEANOGRAPHY UC San Diego
VIRGINIA TECH
Syracuse University
Berkeley UNIVERSITY OF CALIFORNIA

Embodying a Welcoming Community

We put people first and do our best to recognize, appreciate, and **respect the diversity** of our contributors. **We welcome** contributions from **everyone** who shares an interest in scientific software and cybertraining, and wants to contribute in a **healthy and constructive** manner within our community.[1]

We would like to acknowledge that the land we occupy today has served as a site of Indigenous peoples, specifically the *Ramaytush Ohlone people*.

1. Adapted from the EarthCube community.

Agenda

- **Welcome:**
  - **Christine Kirkpatrick**, UC San Diego
- **Overview:**
  - **Kenton McHenry**, U of Illinois at Urbana Champaign
- **Tech Update:**
  - **David Valentine**, UC San Diego
- **Use cases**
  - **Deep Ocean Observing: Karen Stocks**, UC San Diego
  - **Low-temperature geochemistry**: **Tao Wen**, Syracuse U, **Shuang Zhang**, Texas A&M
  - **Ecological Forecasting: Freya Olsson** & **Quinn Thomas,** Virginia Tech; **Carl Boettige**r, UC Berkeley
- **Questions**

The NSF EarthCube Program started around 2012 to advance cyberinfrastructure in support of geoscience research

- Funded roughly 95 efforts to date
- Develop tools, data resources, infrastructure, standards, e.g. data sharing, data interface standards, data wrangling, workflows, analytics, reproducibility

Lessons Learned Towards a Successful Community Effort

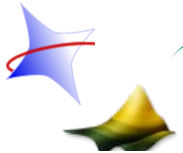
- The criticality of a Code of Conduct!!
- Accepting the sentiment captured in the saying “do not build it expecting us to come”
  - The need to remove oneself from the equation (e.g. pushing ones own software)
  - Leverage! Do not rebuild!
  - Work with the community, align different directions, promote and support directions that have the greatest potential for impact
  - Maintain and promote a deliverable focused mindset
  - Always try to help
- Sustainability means exactly that, sustainability
  - Adopt an approach that could have the potential to be self-driven by a community

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

```
<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 statistics on
    "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.ncdc:C
    "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",
        "name": "Sub dataset 02",
        "description": "Informative description of the second subdataset...",
        "license" : "https://creativecommons.org/publicdomain/zero/1.0/"
      }
    ],
```

Repositories Using Science on Schema

- 26 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, Decade, CReSIS, and DesignSafe

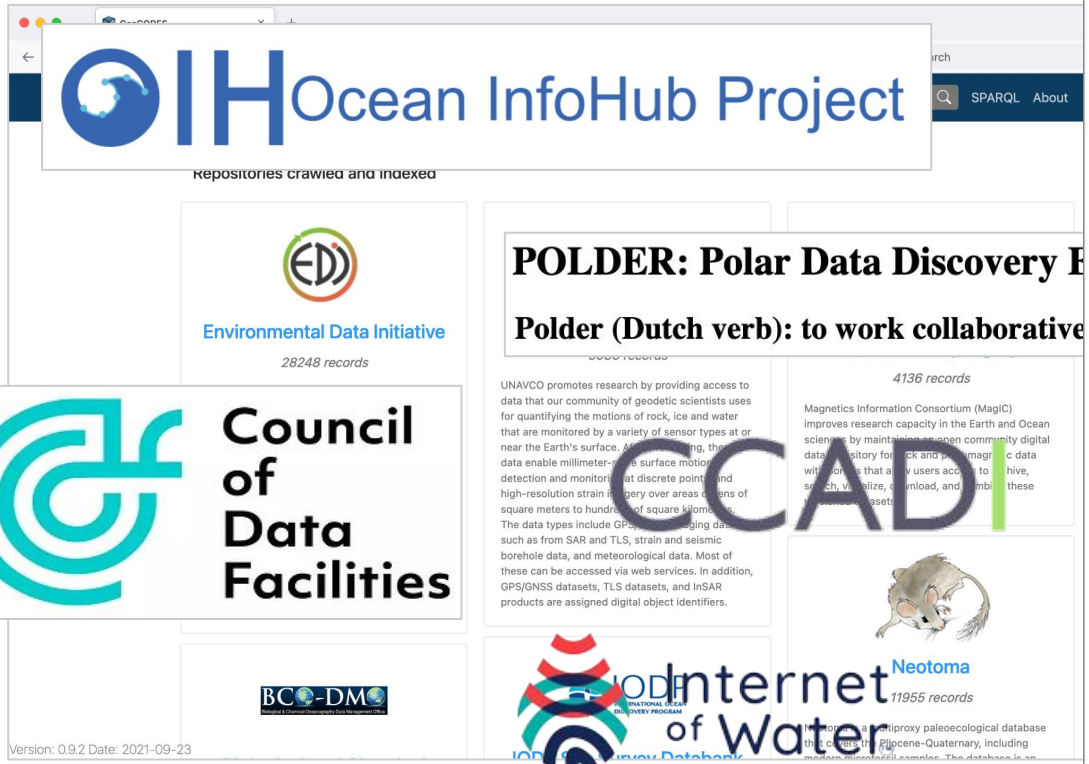


Call for Notebooks - Abstracts Due April 15

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.

- 80 submissions over 3 years with 40 accepted
- Publication in collaboration with AGU as well as leveraging JupyterHub
- Many insights into the future of notebooks replacing PDFs for publication
  - The need for runnability of the final submission!
  - The need to provide style templates
  - The need to support multiple notebook types (e.g. Jupyter, R Markdown)

Potential of GeoCODES



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 [l], 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

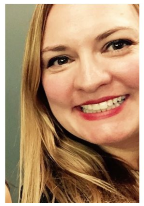
* Corresponding author. Earth and Environmental Systems Institute and Department of Geosciences, The Pennsylvania State University, University Park, PA, USA.
E-mail address: sxb7@psu.edu (S.L. Brantley).

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/© 2021 Elsevier Ltd. All rights reserved.

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

Quinn Thomas
Ecology
VIRGINIA TECH

Carl Boettiger
Ecology
Berkeley
UNIVERSITY OF CALIFORNIA

Lynne Schreiber
Coordinator
UC San Diego

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

↓

Democratized Cyberinfrastructure for Open Discovery to Enable Research (DeCODER)

DeCODER Goals and Activities

- Gleaner community
- Extending Schema.org
- 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

Tech update

Let’s get Geeky

David 7-8 min

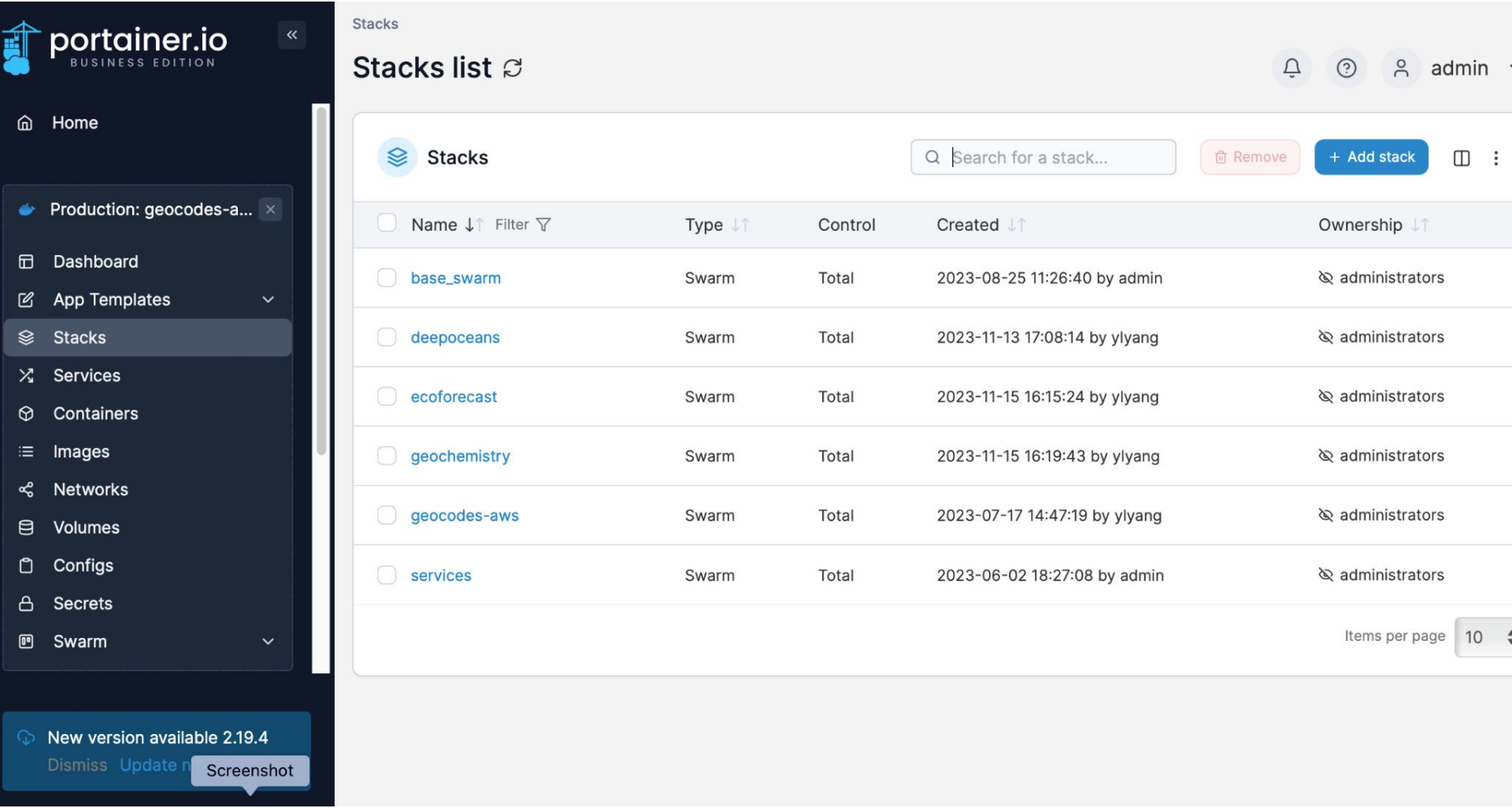
Focus

Decoder/Geocodes

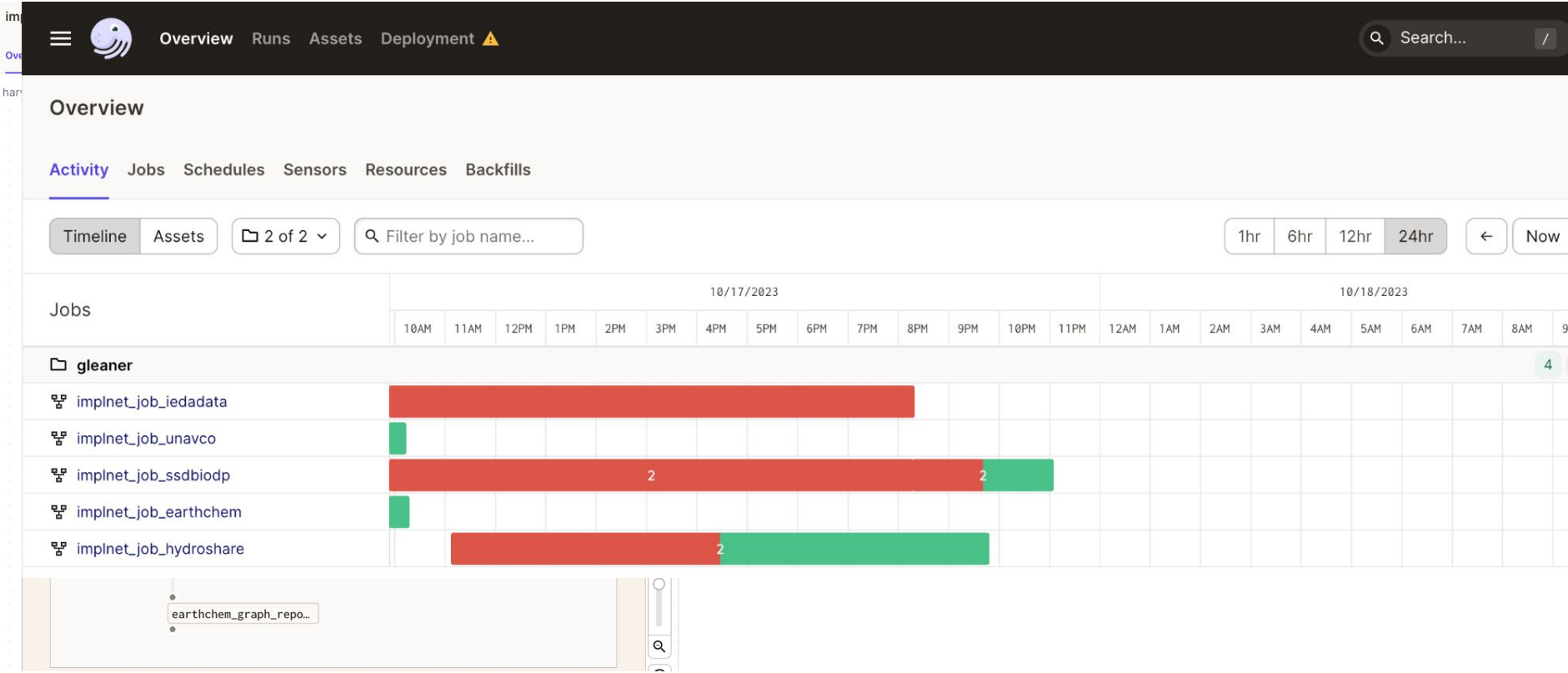
- Improve the geocodes stack
  - Docker Stacks managed by a container management system (Portainer)
- Implement repeatable workflows for data ingest
  - Using Dagster to create repeatable workflow

Community

- Gather Requirements
- Create tools to import/ingest a communities datasets
- Improve Reporting


portainer.io
BUSINESS EDITION
Home
Production: geocodes-a...
Dashboard
App Templates
Stacks
Services
Containers
Images
Networks
Volumes
Configs
Secrets
Swarm
New version available 2.19.4
Dismiss
Update n
Screenshot
Stacks
Stacks list
admin
Stacks
Search for a stack...
Remove
+ Add stack
Name
Filter
Type
Control
Created
Ownership
base_swarm
Swarm
Total
2023-08-25 11:26:40 by admin
administrators
deepoceans
Swarm
Total
2023-11-13 17:08:14 by ylyang
administrators
ecoforecast
Swarm
Total
2023-11-15 16:15:24 by ylyang
administrators
geochemistry
Swarm
Total
2023-11-15 16:19:43 by ylyang
administrators
geocodes-aws
Swarm
Total
2023-07-17 14:47:19 by ylyang
administrators
services
Swarm
Total
2023-06-02 18:27:08 by admin
administrators
Items per page
10

Dagster UI


Overview
Runs
Assets
Deployment
Search...
Overview
Activity
Jobs
Schedules
Sensors
Resources
Backfills
Timeline
Assets
2 of 2
Filter by job name...
1hr
6hr
12hr
24hr
Now
Jobs
10/17/2023
10/18/2023
10AM
11AM
12PM
1PM
2PM
3PM
4PM
5PM
6PM
7PM
8PM
9PM
10PM
11PM
12AM
1AM
2AM
3AM
4AM
5AM
6AM
7AM
8AM
gleaner
4
implnet_job_iedadata
implnet_job_unavco
implnet_job_ssdbiodp
2
2
implnet_job_earthchem
implnet_job_hydroshare
2
earthchem_graph_repo...

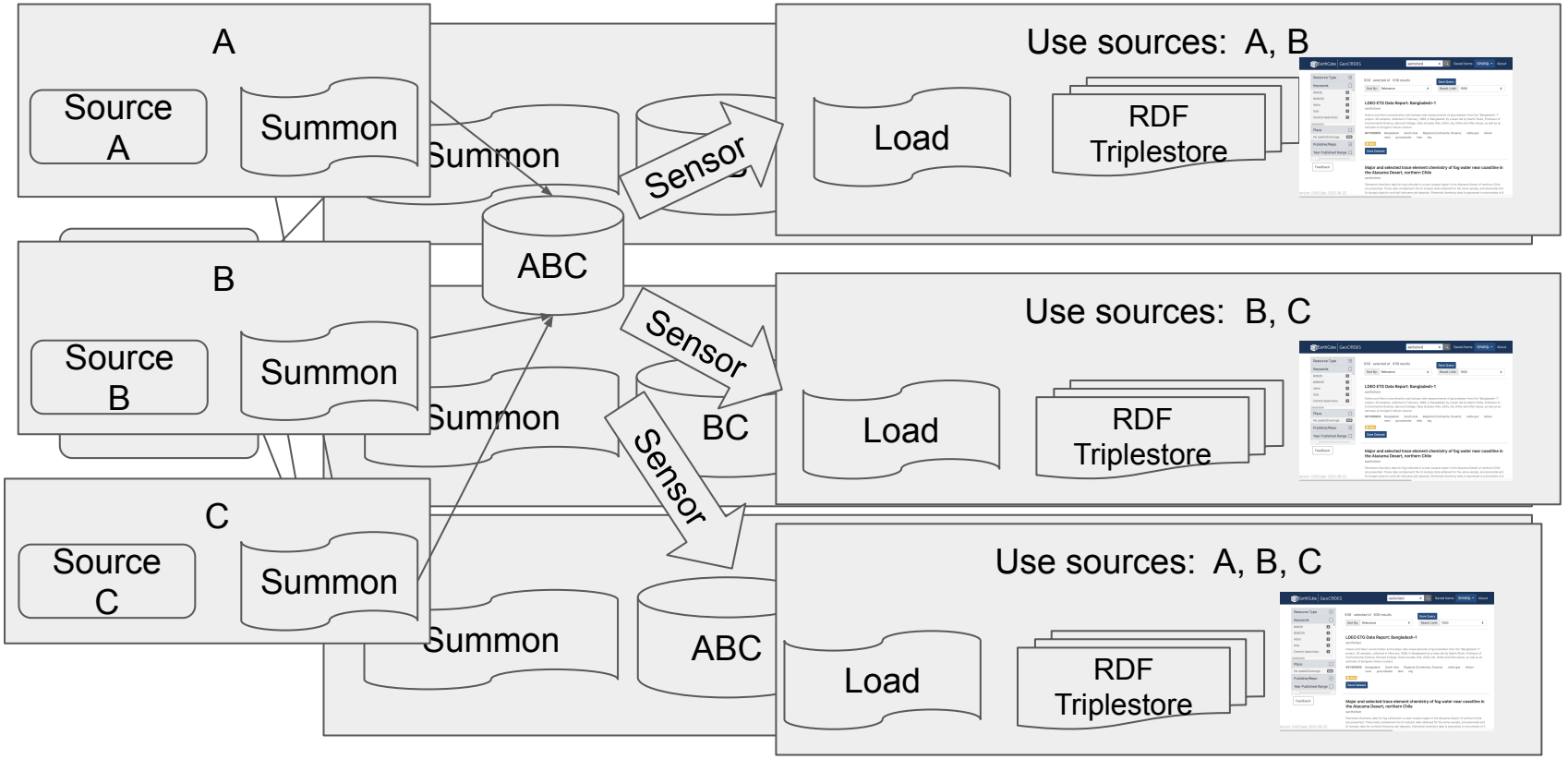
Creating workflows as code using Dagster enables DeCoder/Geocodes to

- Automate formerly manual SOS extraction from web pages listed in sitemaps, and transformation to a knowledge graph
- Expand information checks on knowledge graphs
- Provide reporting for community sources

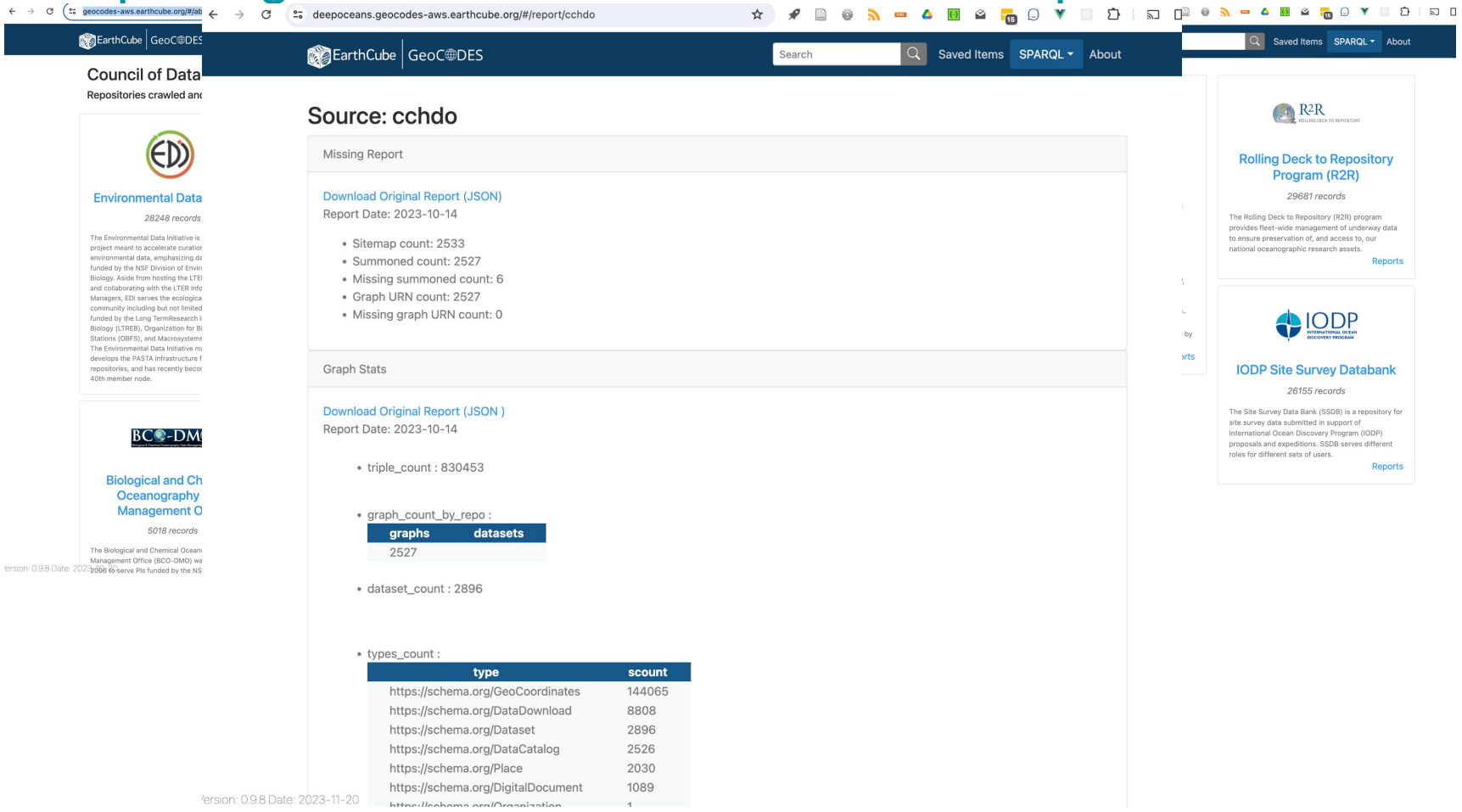
Workflows

Presently, each source is workflow. All utilizing Dagster 'sensors' we can trigger when a dependent entity source(s) is updated

By breaking into multiple workflows, we can trigger when a dependent entity source(s) is updated



Reporting Based on a Communities Graph



Plan

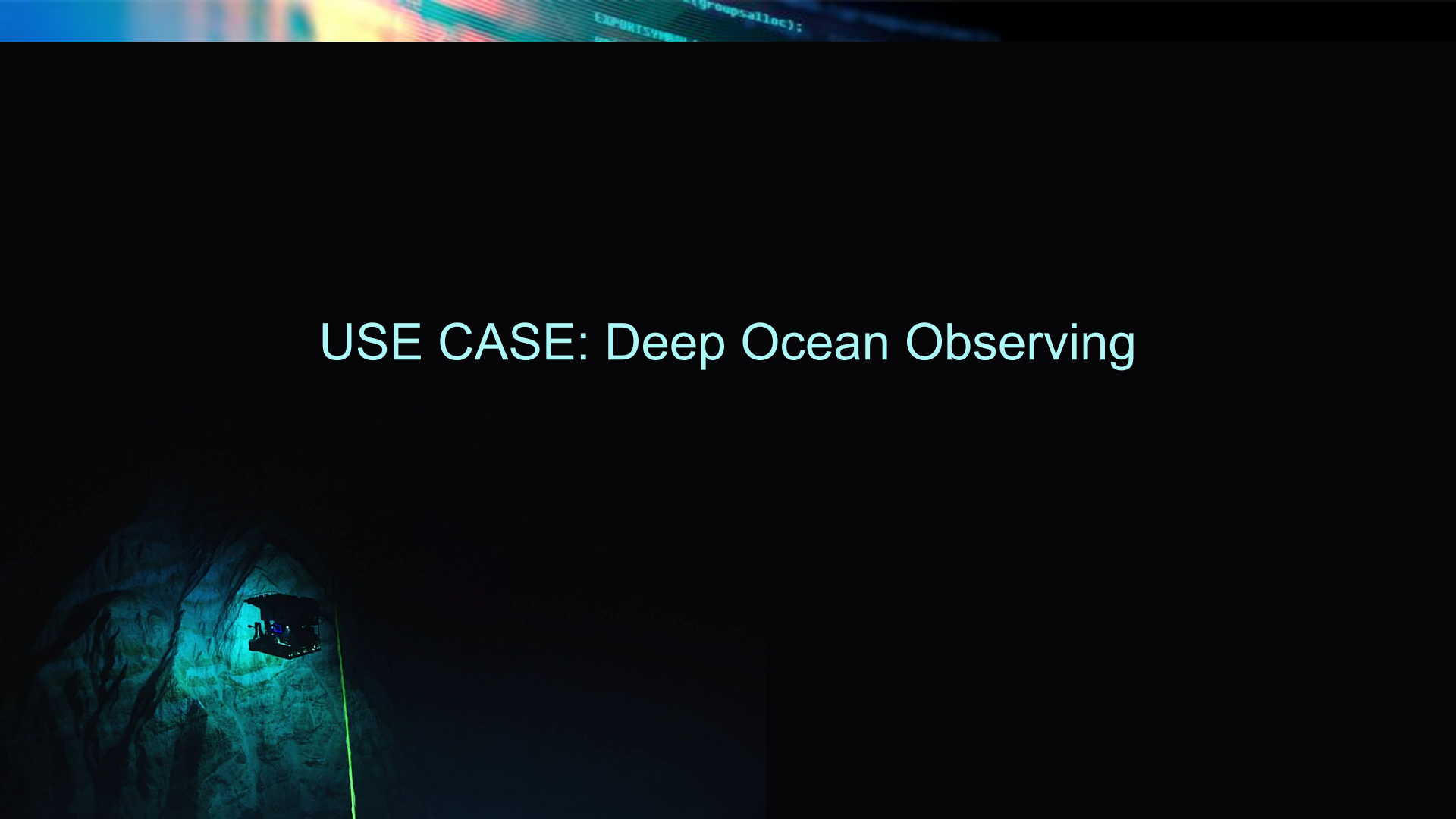
Decoder/Geocodes

- Complete data ingest refactoring
- Validation of SOS using community requirements
- Modify Geocodes UI to handle community requirements
- Create workflows that can be used to manage community data
  - Allow for communities to create communities
- Complete Tools UI

Community

- Implement community requirements
  - Depth
  - Spatial
- Provide reporting on community requirements
- Implement integration
- Provide tools for communities to manage their community portal
  - Add datasets
  - Add tools

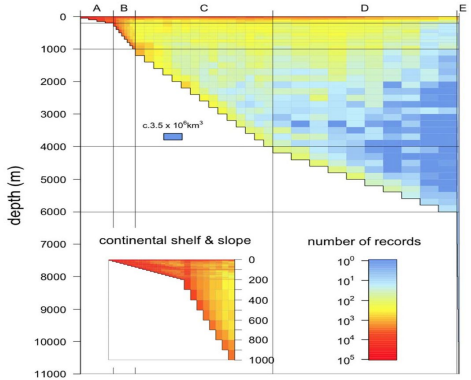
USE CASE: Deep Ocean Observing



Science Drivers

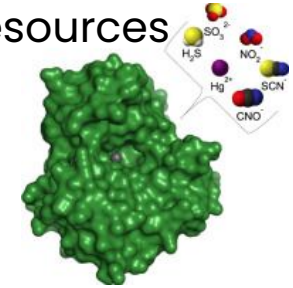
Climate Change

Biodiversity Baselines

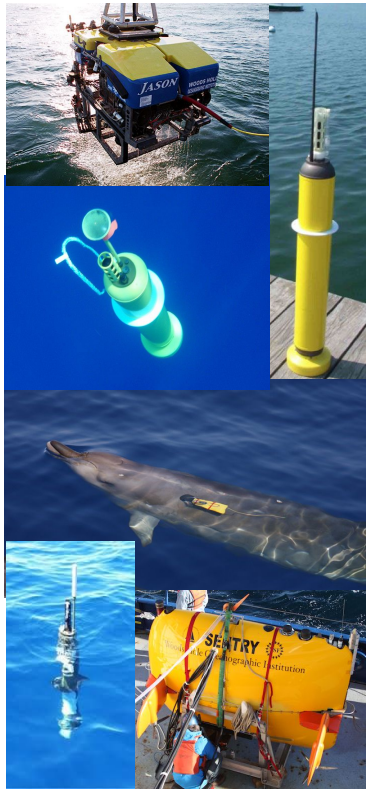


Hazards

Genetic Resources

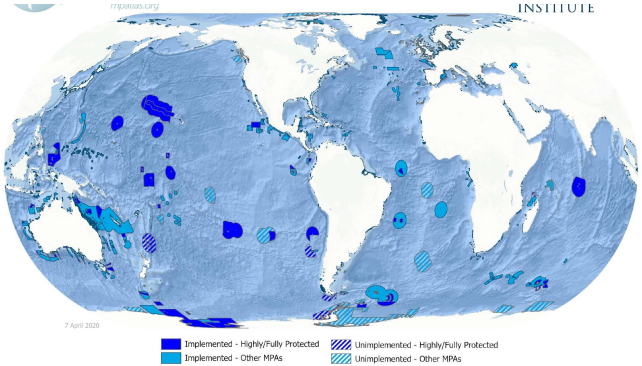


Observatory Planning

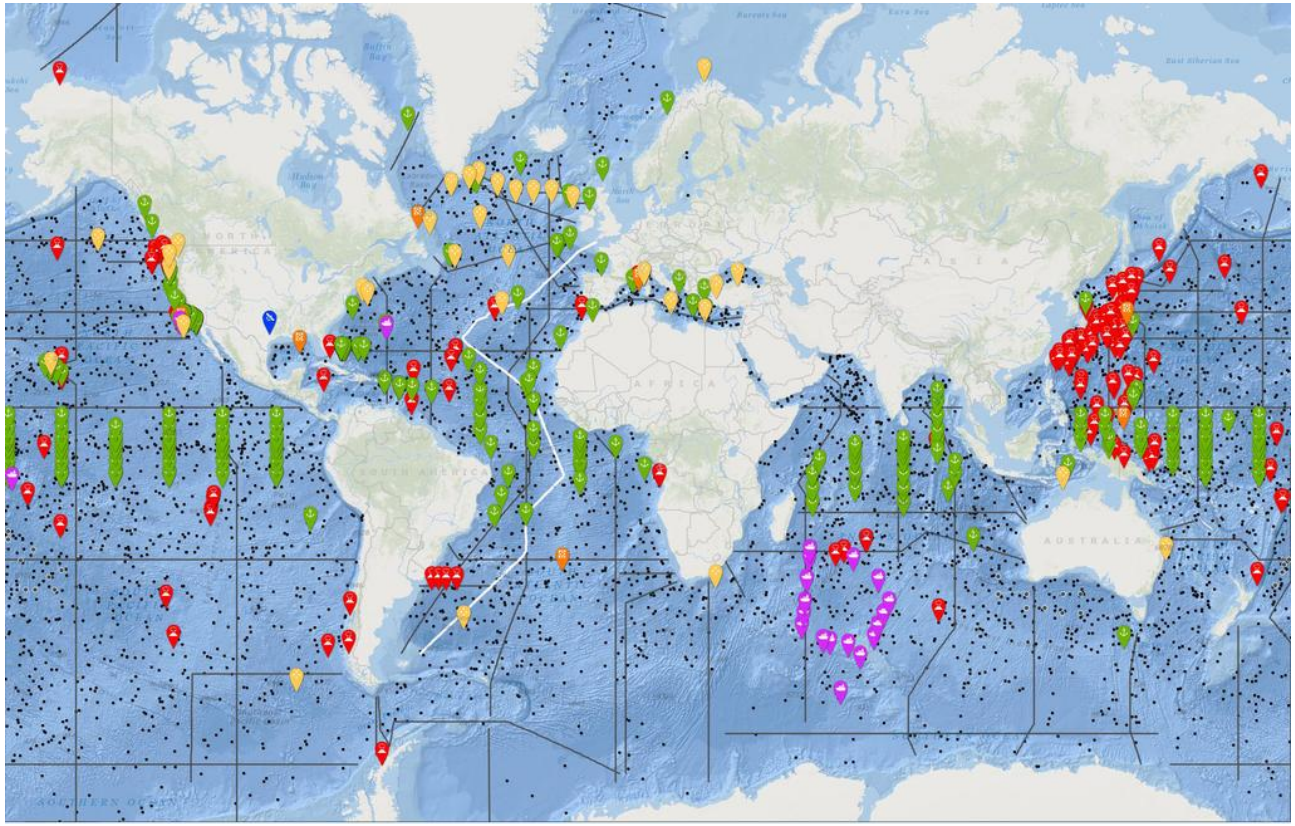


Management Drivers

Marine Protected Areas



Globally Distributed Data



DEEP OCEAN OBSERVING STRATEGY

Partner: Deep Ocean Observing Strategy (DOOS)

A globally integrated network of networks to observe the deep ocean effectively in support of science, policy, and planning for sustainable oceans.

2021 2030 United Nations Decade of Ocean Science for Sustainable Development

www.deepoceanobserving.org

NSF

The DOOS Network

Implementing a Deep Ocean Observing Strategy (iDOOS)

within the Global Ocean Observing System (GOOS)

Observing & Exploration Networks

Argo (Core, BGC, Deep)
Challenger 150 (DOSI/SCOR)
COBRA
EMSO
GEO BON
GO-SHIP
iAtlantic
MBON
NDSF/UNOLS
NOAA Ocean Exploration
OceanSITES
OECI
ONC
OOI
REV Ocean
Schmidt Ocean Institute
JTF SMART Cables
SOOS/SOCCOM
TPOS 2020
US-IOOS

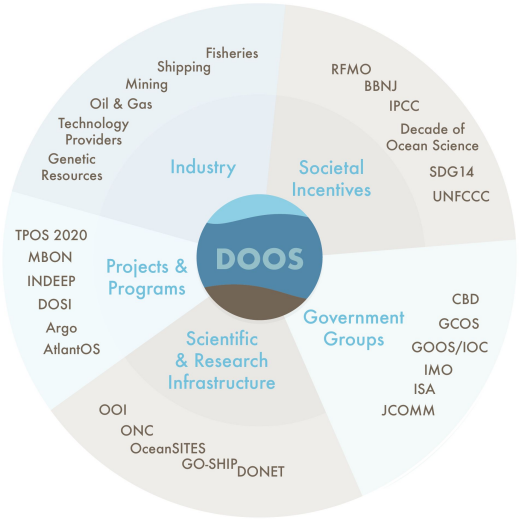
Data & Modelling CI Networks

CCHDO
CLIVAR GSOP
CLIVAR/OMDP & CMIP
ECCO
EMODnet
ESIP/MDC
Esri
FathomNet
IODE/ODIS
IRIS
ISA DeepData
Mercator Ocean
OBIS
OBPS
OceanPredict
Seabed 2030

Management & Policy Users

AtlantOS
DOSI
GEO-BluePlanet
GOOS
Internat. CLIVAR
InterRidge
ISA
IUCN
POGO
UN Decade
UN Global Compact
U.S. CLIVAR
U.S. OCB
U.S. Sanctuaries & Monuments

Operating across defined working groups



The Ocean Info Hub

- a UN Decade of Ocean Science Programme
- building a sustainable, interoperable, and inclusive digital ecosystem for ocean data
- by connecting existing local, national, and regional digital systems and infrastructures…with a schema.org based architecture

unesco
Intergovernmental Oceanographic Commission

2021 2030 United Nations Decade of Ocean Science for Sustainable Development

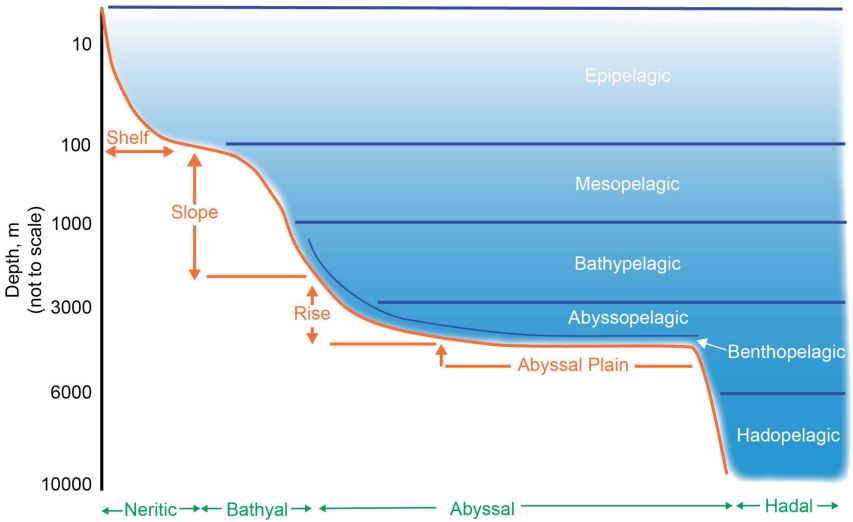
IODE

Flanders State of the Art

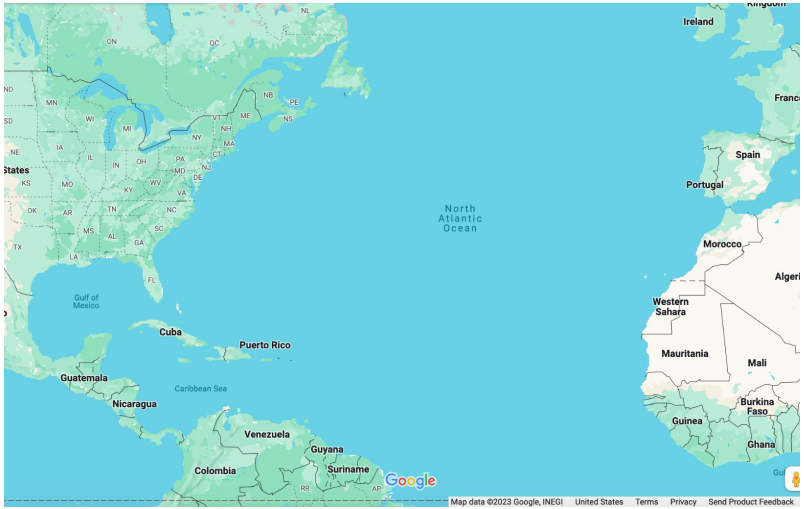
OIH
Ocean InfoHub

First priority: adding Depth to science on schema.org

This

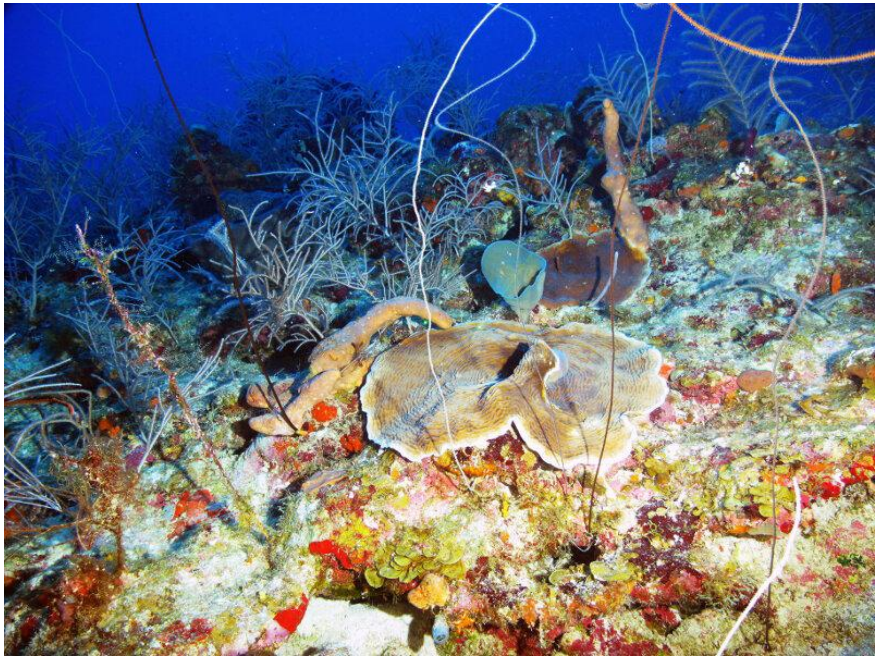


Not (just) this



Expanding schema to cover critical deep ocean discovery concepts

- **DEPTH**
- habitat types,
- sampling platforms & devices,
- parameters



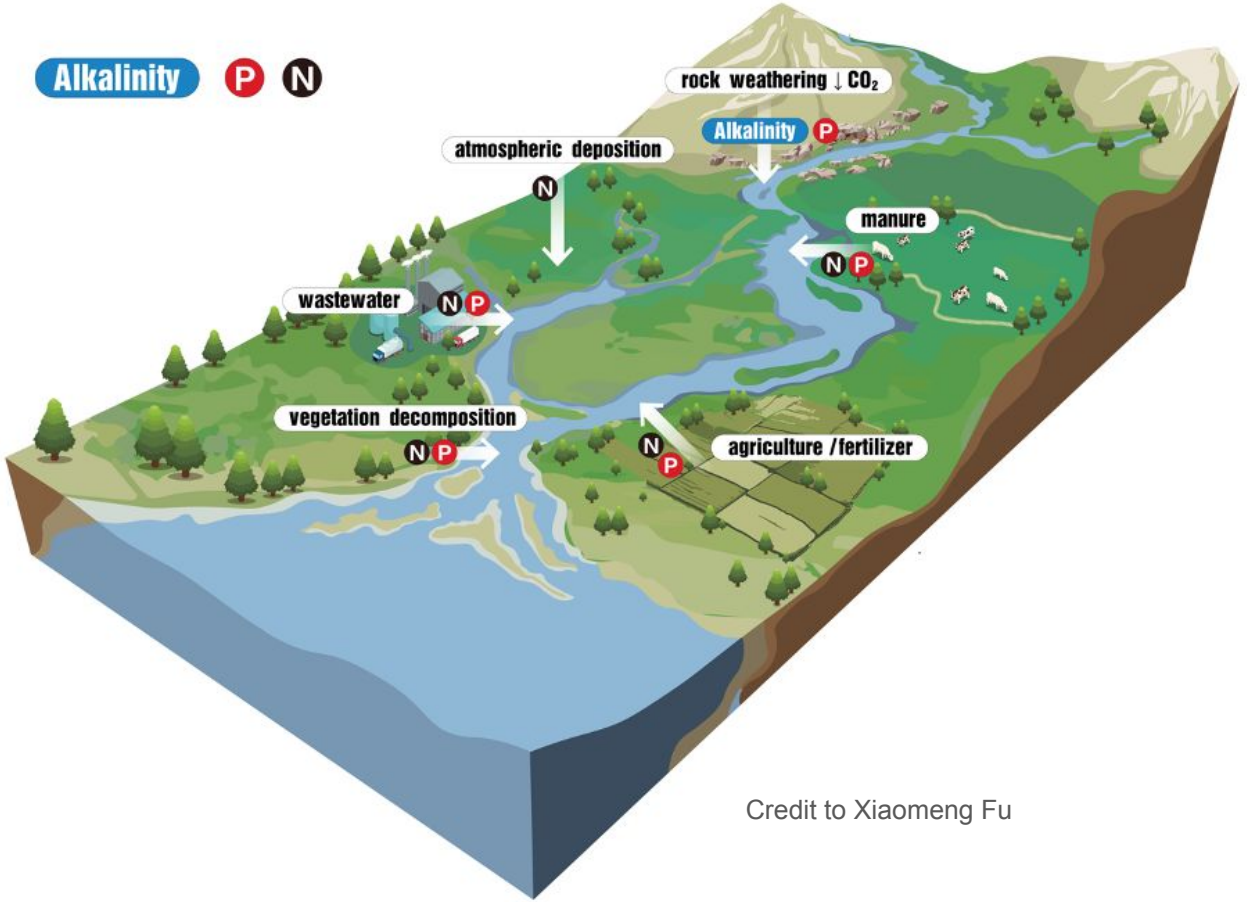
Supporting uptake by data repositories & supporting DOOS working groups

USE CASE: Low-temperature Geochemistry

Tao Wen
Syracuse University

Shuang Zhang
Texas A&M University

Low-temperature geochemistry results from complex processes occurring within its corresponding watershed, regulated by a wide array of variables.



Credit to Xiaomeng Fu

Science of The Total Environment
Volume 889, 1 September 2023, 164138

Human and natural impacts on the U.S. freshwater salinization and alkalinization: A machine learning approach

Beibei E [a], Shuang Zhang [b], Charles T. Driscoll [c], Tao Wen [a]

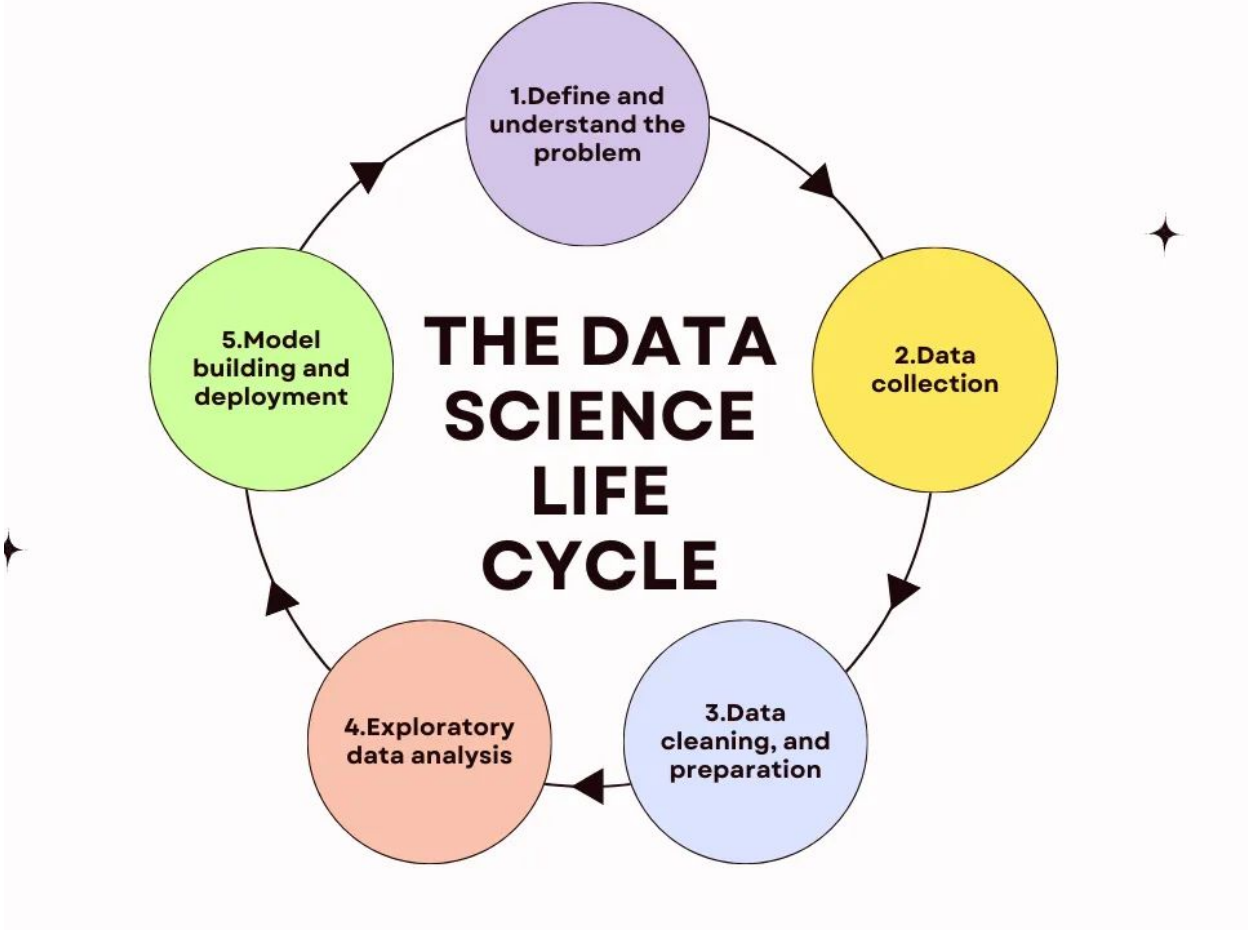
High-dimensional data from multi-disciplines

E et al. STOTEN (2023)

Table 1
List of all 32 watershed attributes along with their corresponding unit, feature category, and source. Features selected for the model development are marked by *.

Feature name	Unit	Category	Source
Runoff*	mm per month	Hydrology	Ghiggi et al. (2021)
Soil moisture*	m^3 per m^3	Hydrology	Wang et al. (2021)
Temperature*	Celsius	Climate	Karger et al. (2017)
Precipitation*	mm per month	Climate	
Watershed area	km^2	Geomorphology	This study
Erosion rate*	mm per year	Geomorphology	Amatulli et al. (2020); Larsen et al. (2014)
Elevation*	Meter	Geomorphology	Amatulli et al. (2020)
Slope	Degree	Geomorphology	
Aspect	Degree	Geomorphology	
Soil organic carbon*	g per kg	Soil chemistry	Poggio et al. (2021)
Soil pH*	pH unit	Soil chemistry	
Evaporite	Percentage	Geology	Hartmann and Moosdorf (2012)
Carbonate sediment*	Percentage	Geology	
Siliciclastic sediment*	Percentage	Geology	
Pyroclastic sediment	Percentage	Geology	
Mixed sediment	Percentage	Geology	
Unconsolidated sediment*	Percentage	Geology	
Igneous basic*	Percentage	Geology	
Igneous intermediate	Percentage	Geology	
Igneous acid	Percentage	Geology	
Metamorphic*	Percentage	Geology	
Impervious surface area*	Percentage	Land use	Brown de Colstoun et al., 2017
Population density*	# per km^2	Land use	Center For International Earth Science Information Network (2016)
Cultivated vegetation*	Percentage	Land use	Tuanmu and Jetz (2014)
Urban	Percentage	Land use	
Trees*	Percentage	Land cover	
Shrubs	Percentage	Land cover	
Herbaceous vegetation	Percentage	Land cover	
Flooded vegetation*	Percentage	Land cover	
Snow ice	Percentage	Land cover	
Barren	Percentage	Land cover	
Water	Percentage	Land cover	

Data Life Cycle



Credit: Madison Hunter | The data science project lifecycle.

DeCODER

EarthCube
GeoCODES
Saved Items
SPARQL
About
GeoCODES
Search
All
Tool
Data
a schema.org/Dataset search

DeCODER

scientific **data**

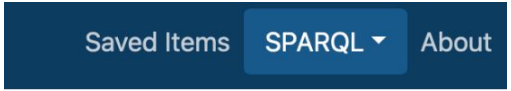
Explore content · About the journal · Publish with us

nature › scientific data › data descriptors › article

Data Descriptor | Open access | Published: 28 May 2020

Geomorpho90m, empirical evaluation and accuracy assessment of global high-resolution geomorphometric layers

Giuseppe Amatulli, Daniel McInerney, Tushar Sethi, Peter Strobl & Sami Domisch



Geomorpho90m

Geomorpho90m - Global High-Resolution Geomorphometric Layers

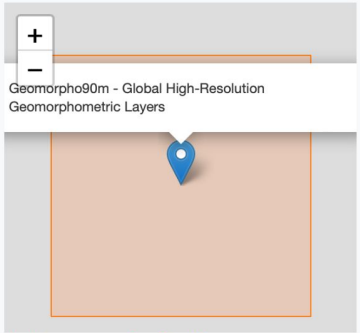
TYPE Data

ABSTRACT Topographical relief comprises the vertical and horizontal variations of the Earth's terrain and drives processes in geomorphology, biogeography, climatology, hydrology and ecology. Its characterisation and assessment, through geomorphometry and feature extraction, is fundamental to numerous environmental modelling and simulation analyses. We, therefore, developed the Geomorpho90m global dataset comprising of different geomorphometric features derived from the MERIT-Digital Elevation Model (DEM) - the best global, high-resolution DEM available. The fully-standardised 26 geomorphometric variables consist of layers that describe the (i) rate of change across the elevation gradient, using first and second derivatives, (ii) ruggedness, and (iii) geomorphological forms. The Geomorpho90m variables are available at 3 (~90 m) and 7.5 arc-second (~250 m) resolutions under the WGS84 geodetic datum, and 100 m spatial resolution under the Equi7 projection. They are useful for modelling applications in fields such as geomorphology, geology, hydrology, ecology and biogeography.

LINKS Object URL
https://doi.org/10.5069/G91R6NPX

Cite

Location



Downloads

Geomorpho90m

Geomorpho90m - Global High-Resolution Geomorphometric Layers

OT Dataspace

Dataset Information

Topographical relief comprises the vertical and horizontal variations of the Earth's terrain and drives processes in geomorphology, biogeography, climatology, hydrology and ecology. Its characterisation and assessment, through geomorphometry and feature extraction, is fundamental to numerous environmental modelling and simulation analyses. We, therefore, developed the Geomorpho90m global dataset comprising of different geomorphometric features derived from the MERIT-Digital Elevation Model (DEM) - the best global, high-resolution DEM available. The fully-standardised 26 geomorphometric variables consist of layers that describe the (i) rate of change across the elevation gradient, using first and second derivatives, (ii) ruggedness, and (iii) geomorphological forms. The Geomorpho90m variables are available at 3 (~90m) and 7.5 arc-second (~250m) resolutions under the WGS84 geodetic datum, and 100m spatial resolution under the Equi7 projection. They are useful for modelling applications in fields such as geomorphology, geology, hydrology, ecology and biogeography.

opentopoID: OTDS.012020.4326.1

DOI: https://doi.org/10.5069/G91R6NPX

Platform: Satellite Data

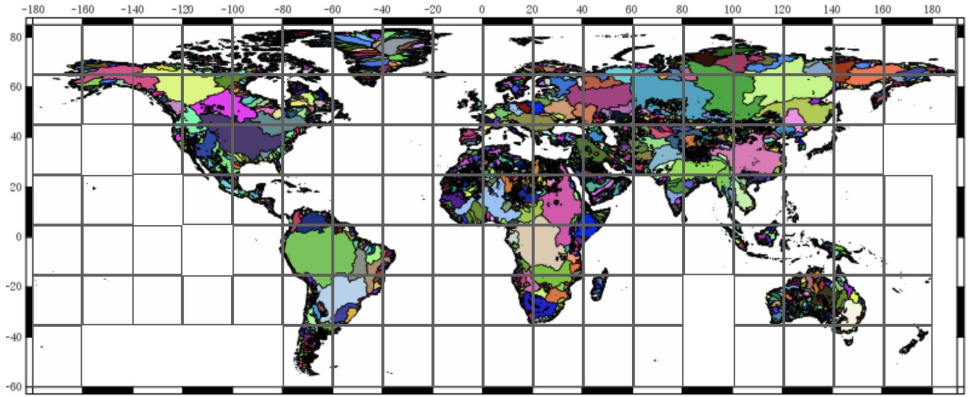
Download Products:

Bulk Download **opentopoID**: OTDS.012020.4326.1

Download KMZ of tiling scheme

Ongoing Work

- Indexing critical data resources on individual websites



Amatulli et al. ESSD (2022)

Output map description	Map reference	Unit	Command	Output tif file name
Flow accumulation (raster)	Fig. 6d	km^2 accumulation=acc	r.watershed -b	accumulation_*.tif
Flow direction (raster)	Fig. 6f	NE–N–NW–W–SW–S–SE–E correspond to 1–2–3–4–5–6–7–8	r.stream.extract direction=dir threshold=0.05	direction_*.tif
Drainage basin (raster)	Fig. 6g	IDs from 1 to 1 676 628	r.stream.basins -l basins=basin	basin_*.tif
Drainage basin (vector)	Fig. 5b	IDs from 1 to 1 676 628	gdal_polygonize.py	basin_*.gpkg
Outlets (raster)	Fig. 6h	ID=1 stream_vector=stream threshold=0.05; v.to.rast input=stream	r.stream.extract	outlet_*.tif
Outlets (vector)	Fig. 6h	ID=1	gdal_polygonize.py	outlet_*.gpkg
Depression (raster)	–	ID = 1	pksetmask	depression_*.tif
Stream segment (raster)	Fig. 6h	IDs from 1 to 726 723 221	r.stream.extract stream_raster=stream threshold=0.05	segment_*.tif
Sub-catchment (raster)	Fig. 6i 1 to 726 723 221	IDs from basins=sub_catchment	r.stream.basins	sub_catchment_*.tif
Sub-catchment (vector)	Fig. 5d	IDs from 1 to 726 723 221	gdal_polygonize.py	sub_catchment_*.gpkg
Regional unit (raster)	Fig. 7	IDs from 1 to 116 IDs from 150 to 200	pkreclass pksetmask	regional_unit_*.tif

Ongoing Work

- Indexing PANGAEA data resources that are essential to low-T geochemistry and Earth surface processes research

Variable Type	Database Name	Variable Description	Source	Download Link
Vegetation Cover	Global GPP	Gross primary production	Zhang et al. (2017)	https://doi.pangaea.de/10.1594/PANGAEA.879560
Lithology	GLiM	Surface lithology	Hartmann and Moosdorf (2012)	https://doi.pangaea.de/10.1594/PANGAEA.788537
Agriculture	Nitrogen Fertilizer	$NH4^+$ and $NO3^-$ application in synthetic nitrogen fertilizer	Nishina et al. (2017)	https://doi.pangaea.de/10.1594/PANGAEA.861203
Agriculture	Manure Nitrogen	Manure nitrogen production and application in cropland	Zhang et al. (2017)	https://doi.pangaea.de/10.1594/PANGAEA.871980
River Hydrochemistry	GLORICH	River chemistry and watershed property	Hartmann et al. (2019)	https://doi.pangaea.de/10.1594/PANGAEA.902360
River Hydrochemistry	GSIM	River discharge	Do et al. (2018)	https://doi.pangaea.de/10.1594/PANGAEA.887477

USE CASE: Ecological Forecasting

Ecological Forecasting Initiative

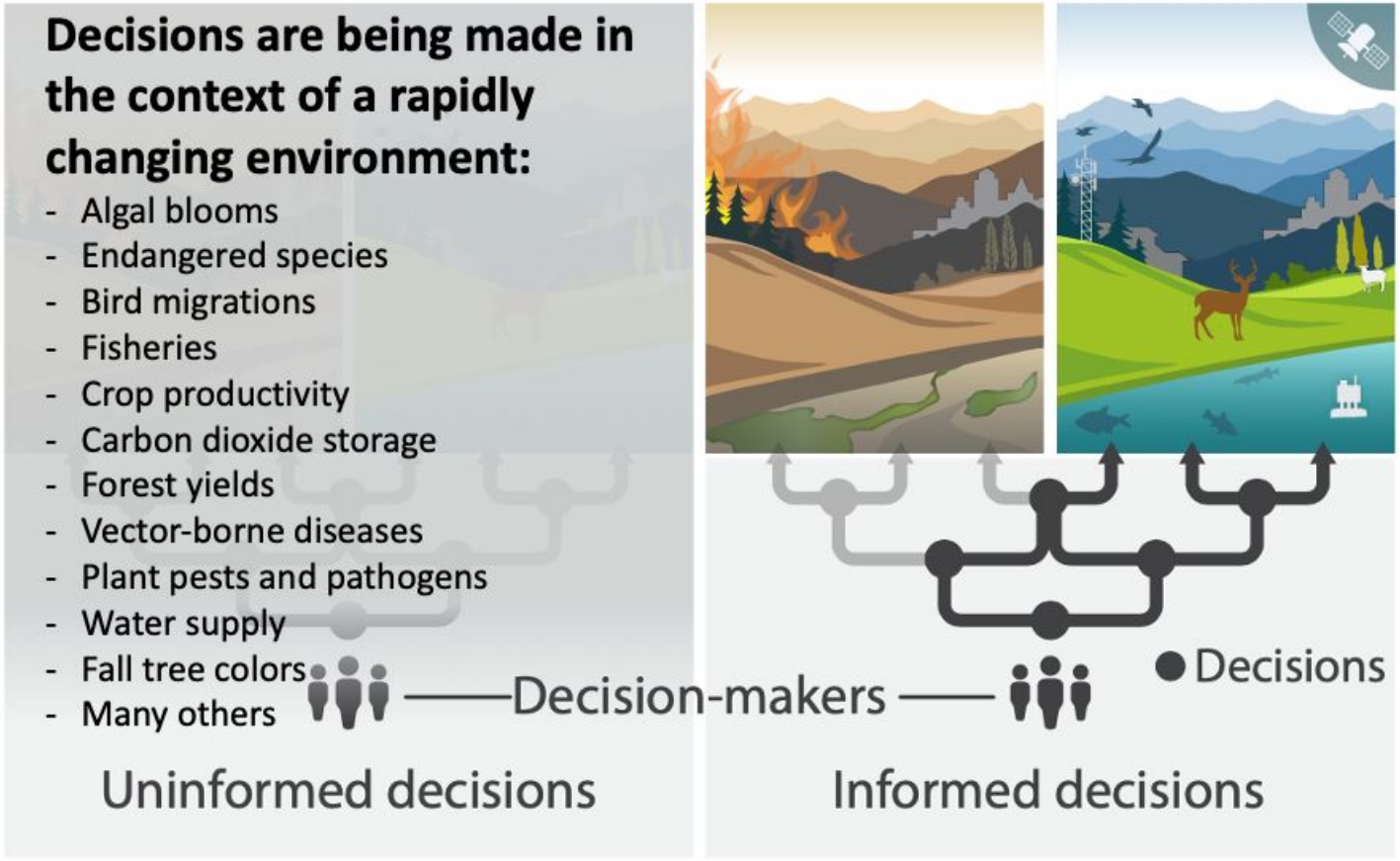
UNDERSTAND · MANAGE · CONSERVE

ecoforecast.org

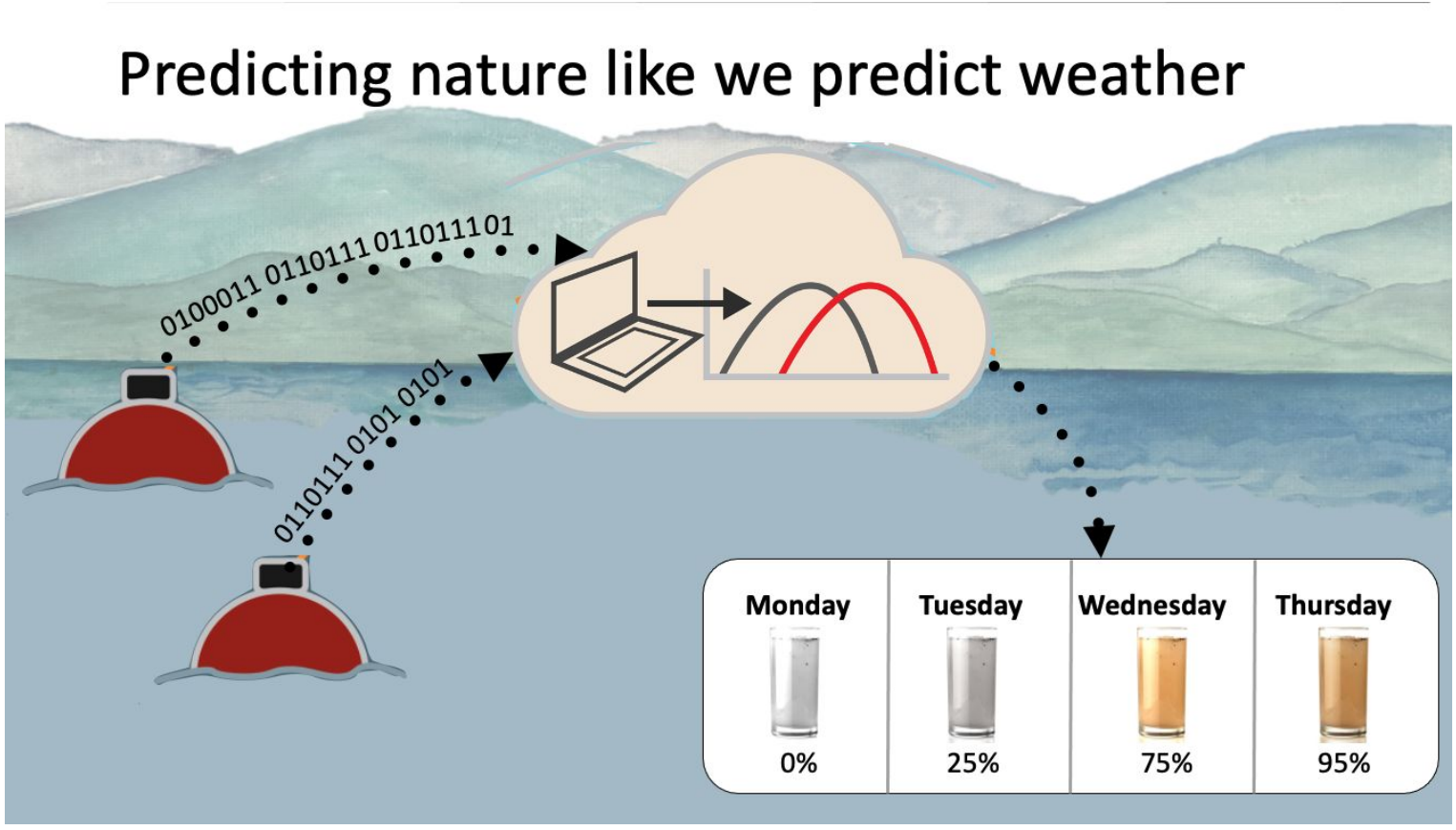
Quinn Thomas
Virginia Tech

Carl Boettiger
UC Berkeley

USE CASE: Ecological forecasting



USE CASE: Ecological forecasting



USE CASE: Ecological forecasting

What makes ecological forecasting an exciting use-case?

- Critical need to be able to discover past forecasts to evaluate with new data
- The dataset is continuously updated as new forecasts are generated
- Focus on cloud storage to support the evolving dataset
- Multiple concepts of time (time being forecasted and time that forecast was generated) and emphasis on uncertainty quantification
- Emerging community of practice so opportunities to shape the standards for describing and archiving forecasts.

Ecological Forecasting Initiative

UNDERSTAND · MANAGE · CONSERVE

Grassroots consortium aimed at building an interdisciplinary community of practice in ecological forecasting

Standardized **terrestrial** and **freshwater** data with ongoing collection that is freely available at 81 sites across the U.S.

NSF-sponsored

Ecological Forecasting Initiative

Research Coordination Network

5-year project

Create a community of practice that builds capacity for ecological forecasting by leveraging NEON data products.

https://ecoforecast.org/rcn/

Funded by the National Science Foundation (DEB-1926388)

The NEON Ecological Forecasting Challenge

R Quinn Thomas[1*], Carl Boettiger[2], Cayelan C Carey[1], Michael C Dietze[3], Leah R Johnson[1], Melissa A Kenney[4], Jason S McLachlan[5], Jody A Peters[5], Eric R Sokol[6], Jake F Weltzin[7], Alyssa Willson[5], Whitney M Woelmer[1], and Challenge contributors[8]

[1]*Virginia Tech* [*](rqthomas@vt.edu); [2]*University of California–Berkeley;* [3]*Boston University;* [4]*University of Minnesota;* [5]*University of Notre Dame;* [6]*National Ecological Observatory Network;* [7]*US Geological Survey;* [8]*See WebPanel 1 for additional authors*

Front Ecol Environ 2023; 21(3): 112–113, doi:10.1002/fee.2616

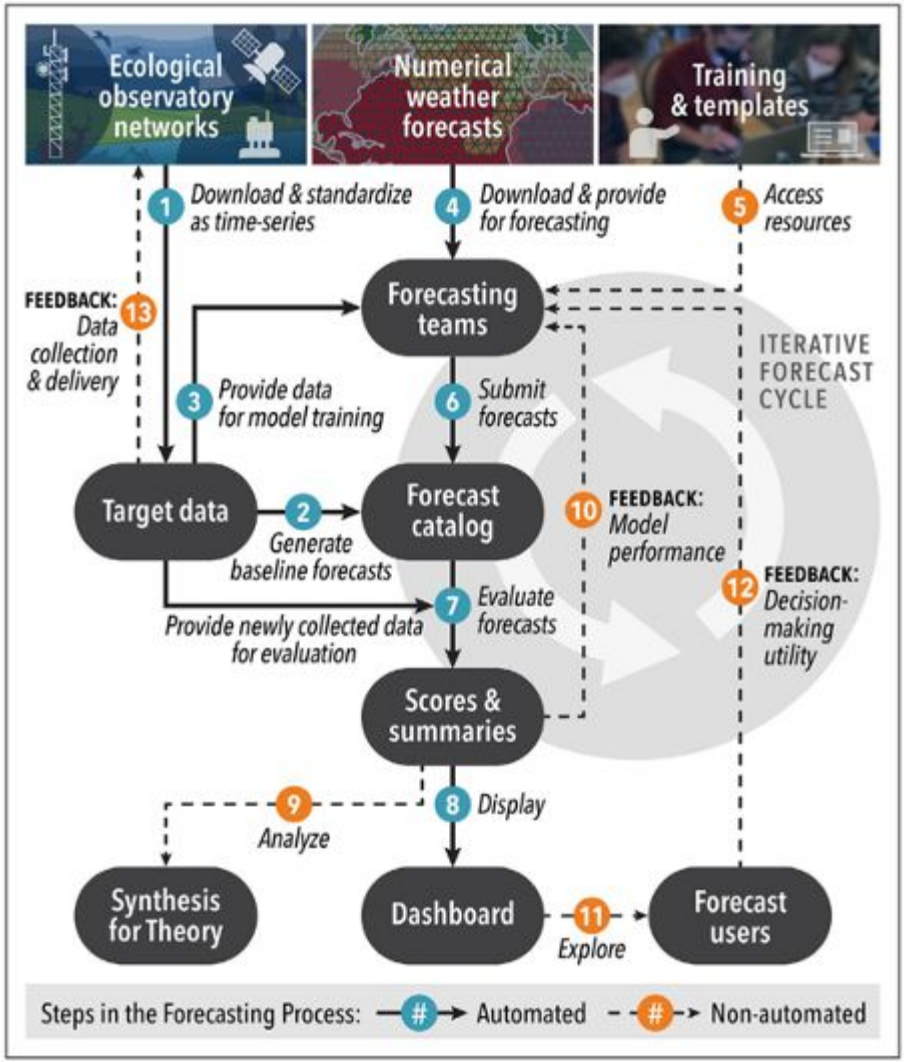
A platform that challenges and empowers the community to submit **iterative near-term forecasts** of *yet-to-be-collected* NEON data

Ecological Forecasting Initiative
UNDERSTAND · MANAGE · CONSERVE

NSF | neon Operated by Battelle

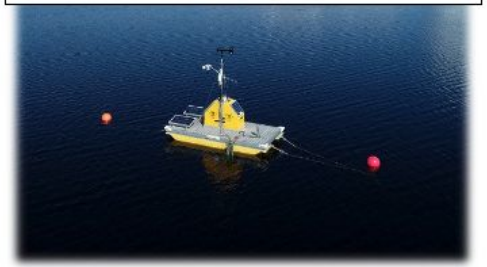
Funded by the National Science Foundation (DEB-1926388)

www.neon4cast.org


Ecological observatory networks
Numerical weather forecasts
Training & templates
1 Download & standardize as time-series
4 Download & provide for forecasting
5 Access resources
FEEDBACK: 13 Data collection & delivery
Forecasting teams
ITERATIVE FORECAST CYCLE
3 Provide data for model training
6 Submit forecasts
Target data
2 Generate baseline forecasts
Forecast catalog
10 FEEDBACK: Model performance
12 FEEDBACK: Decision-making utility
Provide newly collected data for evaluation
7 Evaluate forecasts
Scores & summaries
9 Analyze
8 Display
Synthesis for Theory
Dashboard
11 Explore
Forecast users
Steps in the Forecasting Process: # Automated # Non-automated

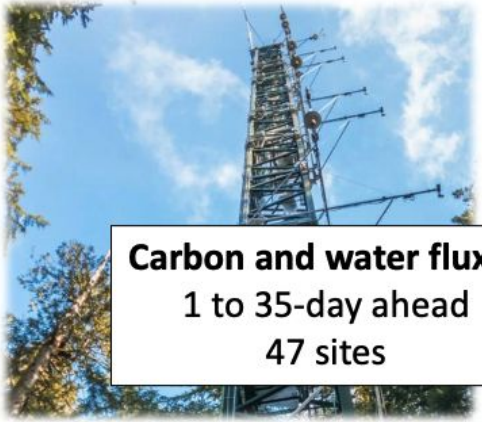
Challenge themes

Temperature, oxygen, chl-a
1 to 35-day ahead
34 sites

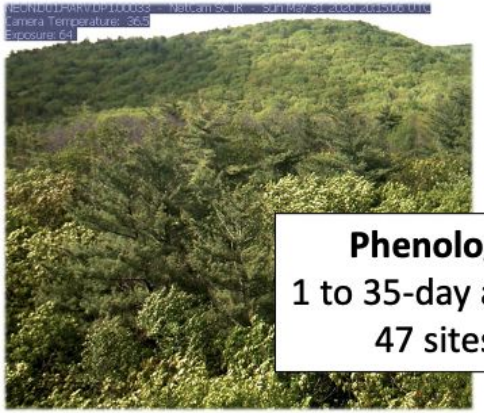


Beetle community richness
1 week to 1 year ahead
47 sites

Tick larvae abundance
1 week to 1 year ahead
47 sites

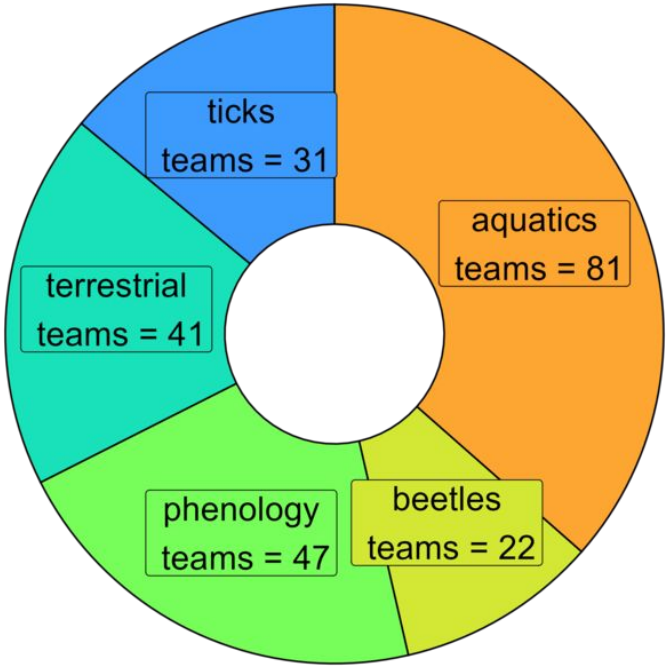


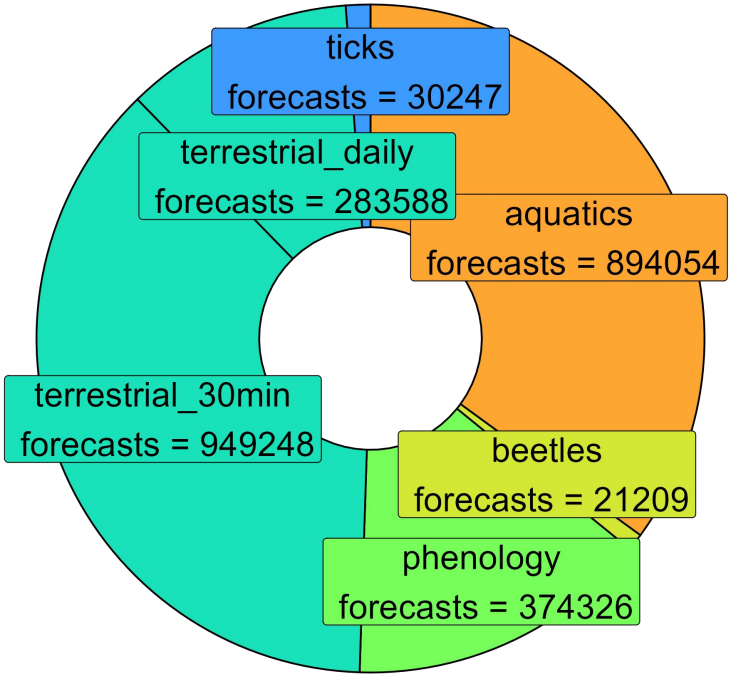
Carbon and water fluxes
1 to 35-day ahead
47 sites



Phenology
1 to 35-day ahead
47 sites

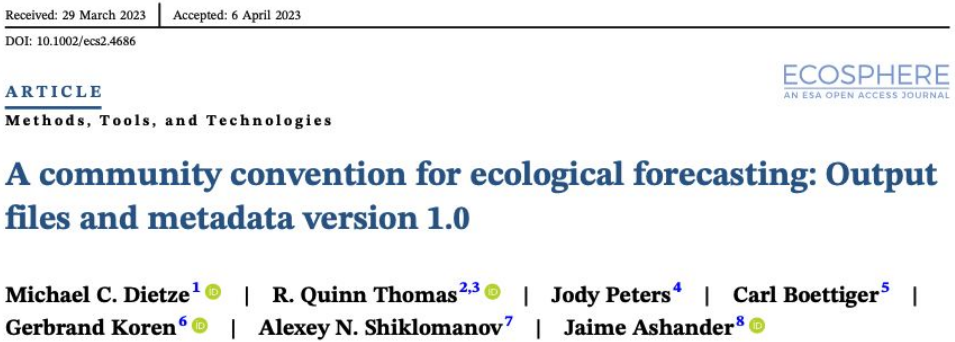
222 teams have produced 2,252,672 forecasts





USE CASE: Ecological forecasting

How do we describe, archive, and discover these forecasts from different ecological systems and scales?



Received: 29 March 2023 | Accepted: 6 April 2023
DOI: 10.1002/ecs2.4686

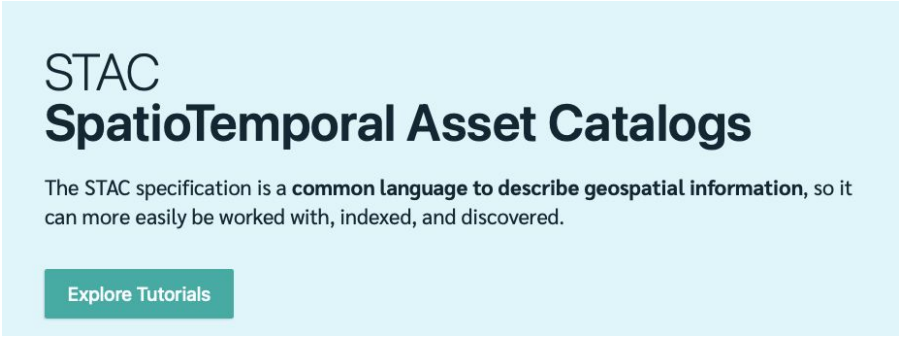
ECOSPHERE
AN ESA OPEN ACCESS JOURNAL

ARTICLE
Methods, Tools, and Technologies

A community convention for ecological forecasting: Output files and metadata version 1.0

Michael C. Dietze[1] | R. Quinn Thomas[2,3] | Jody Peters[4] | Carl Boettiger[5] | Gerbrand Koren[6] | Alexey N. Shiklomanov[7] | Jaime Ashander[8]

https://doi.org/10.1002/ecs2.4686



Set of linked json files

https://stacspec.org/en

Up | Browse

Description

A STAC (Spatiotemporal Asset Catalog) describing forecasts and forecast scores for the neon4cast Forecasting Challenge

Catalogs 7

Tiles | List | Ascending | Descending

Filter catalogs by title

Forecast Summaries

Parquet Summaries are the forecasts statistics of the raw forecasts (i.e., mean, median, confidence intervals). You can

1/1/2023, 12:00:00 AM UTC - 11/25/2024, 12:00:00 AM UTC

Inventory

Parquet The catalog contains forecasts for the NEON Ecological Forecasting Challenge. The forecasts are the raw

1/1/2023, 12:00:00 AM UTC - 11/25/2024, 12:00:00 AM UTC

Scores

Parquet The catalog contains scores for the NEON Ecological Forecasting Challenge. The scores are summaries of the

1/1/2023, 12:00:00 AM UTC - 11/25/2024, 12:00:00 AM UTC

Forecasts

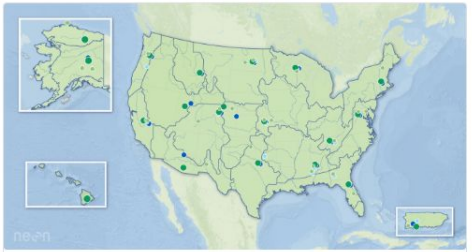
Parquet Forecasts are the raw forecasts that includes all ensemble members or distribution parameters. Due to the size

1/1/2023, 12:00:00 AM UTC - 11/25/2024, 12:00:00 AM UTC

NOAA-Forecasts

Parquet The catalog contains NOAA forecasts used for the NEON Ecological Forecasting Challenge. The forecasts are the

9/25/2020, 12:00:00 AM UTC - 1/11/2024, 12:00:00 AM UTC



Site Metadata

Parquet The catalog contains site metadata for the NEON Ecological Forecasting Challenge

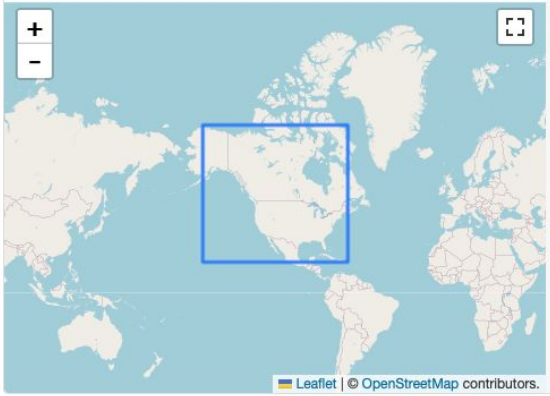
Invalid Date

Forecasts

Description

Forecasts are the raw forecasts that includes all ensemble members or distribution parameters. Due to the size of the raw forecasts, we recommend accessing the scores (summaries of the forecasts) to analyze forecasts (unless you need the individual ensemble members). You can access the forecasts at the top level of the dataset where all models, variables, and dates that forecasts were produced (reference_datetime) are available. The code to access the entire dataset is provided as an asset. Given the size of the forecast catalog, it can be time-consuming to access the data at the full dataset level. For quicker access to the forecasts for a particular model (model_id), we also provide the code to access the data at the model_id level as an asset for each model.

License	CC0-1.0
Temporal Extent	1/1/2023, 12:00:00 AM UTC - 11/25/2024, 12:00:00 AM UTC



Map

Thumbnails

Assets

Database Access DATA PARQUET

Catalogs 6

Tiles | List | Ascending | Descending

Filter catalogs by title

Aquatics

Parquet This page includes variables for the Aquatics group.

1/1/2023, 12:00:00 AM UTC - 11/25/2024, 12:00:00 AM UTC

Beetles

Parquet This page includes variables for the...

1/1/2023, 12:00:00 AM UTC - 11/25/2024, 12:00:00 AM UTC

Models

Parquet Forecasts are the raw forecasts that includes all ensemble members...

1/1/2023, 12:00:00 AM UTC - 11/25/2024, 12:00:00 AM UTC

Phenology

Parquet This page includes variables for the...

1/1/2023, 12:00:00 AM UTC - 11/25/2024, 12:00:00 AM UTC

Terrestrial

Parquet This page includes variables for the...

1/1/2023, 12:00:00 AM UTC - 11/25/2024, 12:00:00 AM UTC

EarthCube | GeoCODES

Saved Items SPARQL About

GeoCODES

water temperature forecasts NEON lakes

All Tool Data

an interdisciplinary geoscience data and tool search engine

Ecological Forecasting Initiative Forecasting Challenges Catalog

Source | Share | Language: English

Browse

Description

A STAC (Spatiotemporal Asset Catalog) describing forecasts and forecast scores for ecological forecasting challenges

Catalogs 2

Tiles | List | Ascending | Descending

Filter catalogs by title

NEON Ecological Forecasting Challenge Catalog

A STAC (Spatiotemporal Asset Catalog) describing forecasts and forecast scores for the neon4cast Forecasting Challenge

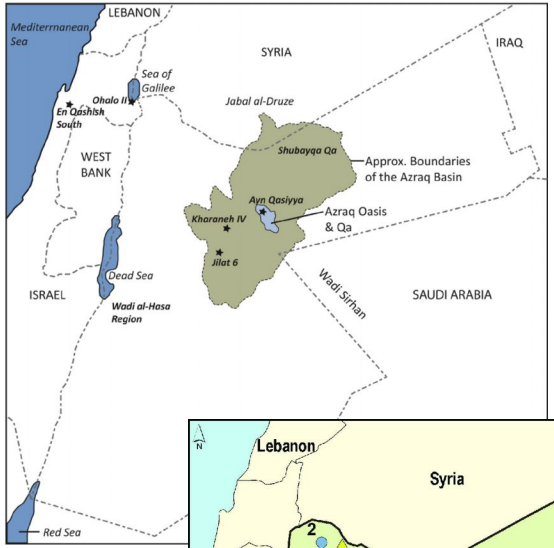
Virginia Ecoforecast Reservoir Analysis Catalog

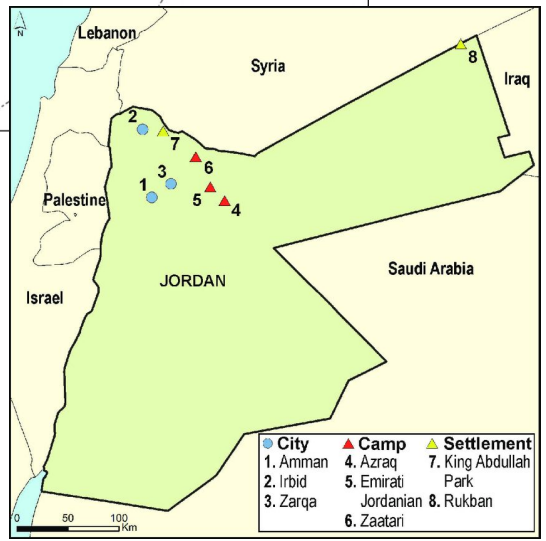
A STAC (Spatiotemporal Asset Catalog) describing forecasts and forecast scores for the vera4cast Forecasting Challenge

https://radiantearth.github.io/stac-browser/#/external/raw.githubusercontent.com/eco4cast/challenge-catalogs/main/catalog.json?.language=en

DeCODER for Climate and Health

- Global Center on Climate Change, Water, Energy, Food, and Health Systems (GC3WEFH) - funded by NIH/NIEHS; PI: Wael Al-Delaimy, UC San Diego
- Partners: UCSD, UCSF, Texas A&M, University of Jordan, Hashemite University, The Royal Science Society in Jordan, Climate Action Network
- DeCODER to present data related to climate, health, water, energy, food especially in the Azraq basin of Jordan
  - Supporting modeling
  - Data collection for engagement and research teams
- DeCODER & GC3WEFH overlap - Kirkpatrick & Zaslavsky co-leading Data Core, David Valentine, data integration





This work funded by an NIH, NIEHS grant #1P20TW012709-01. Image sources: Occupying wide open spaces? Late Pleistocene hunter–gatherer activities in the Eastern Levant - Scientific Figure on ResearchGate. Available from: https://www.researchgate.net/figure/Map-of-the-Azraq-Basin-showing-the-extent-of-this-hydrological-system-in-eastern-Jordan_fig1_283040843; MIGRANTS, ASYLUM SEEKERS AND REFUGEES IN JORDAN, 2017 - Scientific Figure on ResearchGate. Available from: https://www.researchgate.net/figure/Refugee-camps-and-settlements-in-Jordan-Source-UNHCR-2018c_fig3_325594771

Questions