

Eric R. Sokol
Senior Quantitative Ecologist
NSF CI Compass virtual workshop: AI Meets CI
Monday, January 12, 2026

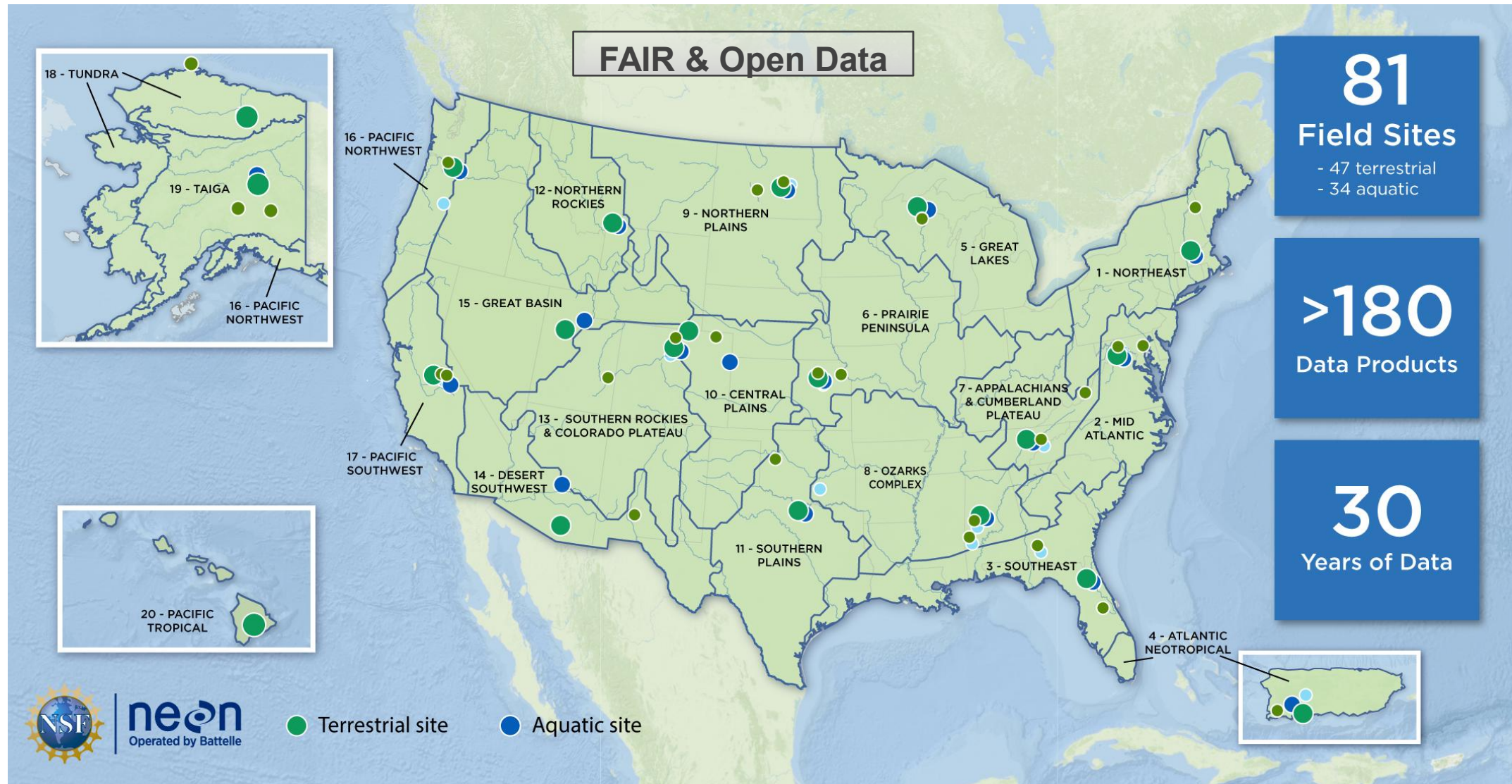


neon
Operated by Battelle

Making the National Ecological Observatory Network (NEON) AI-FAIR

Panel II: Making MF data AI-ready

NEON is a distributed Observatory across the U.S.



How we collect data and samples

Standardized, colocated methods across sites



Automated
instrument systems



Observational
sampling

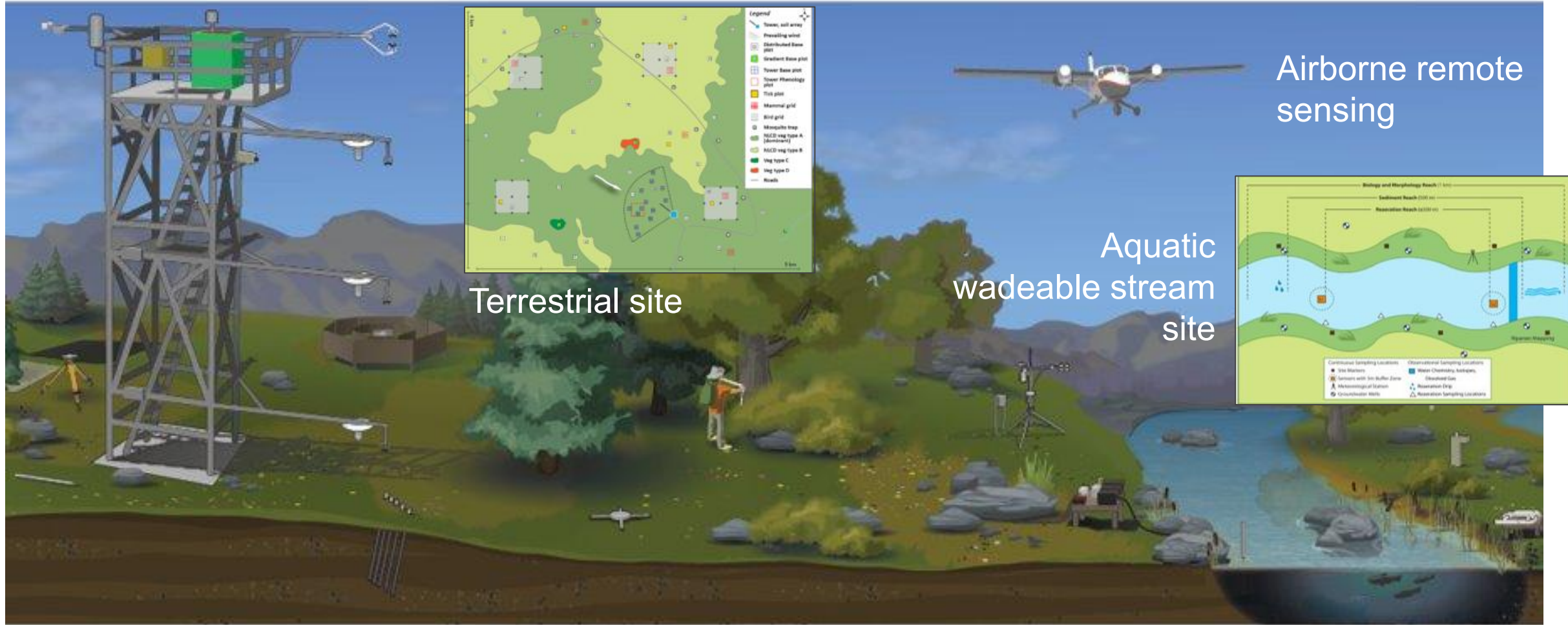


Airborne observation
platform



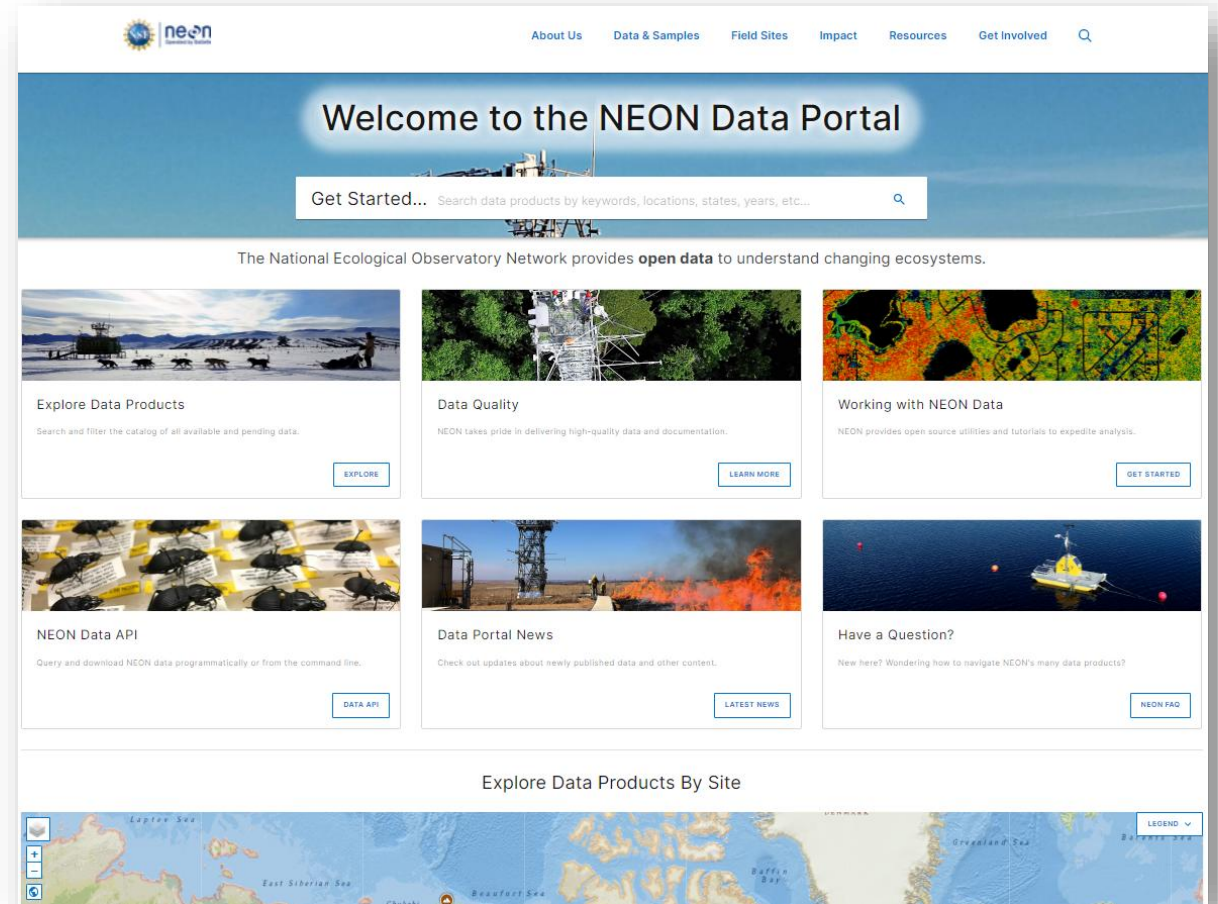
How we collect data and samples

Standardized, colocated methods across sites



Data portal: <https://data.neonscience.org>

- Explore and download FREE data
- Information on programmatic access to NEON data
 - API
 - Code packages
- Metadata:
 - API endpoint with schema.org
 - EML comes with data download
- Access data product user guides, detailed protocols, and other important documents



NEON specimens & samples: NEON Biorepository

>100,000 specimens & samples per year



ASU Arizona State University

BARCODE OF LIFE DATA SYSTEM ^{v4}

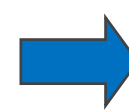
Advancing biodiversity science through DNA-based species identification.



GBIF

Global Biodiversity
Information Facility

>1.3M
occurrences



iDigBio

Integrated Digitized Biocollections

65 sample types

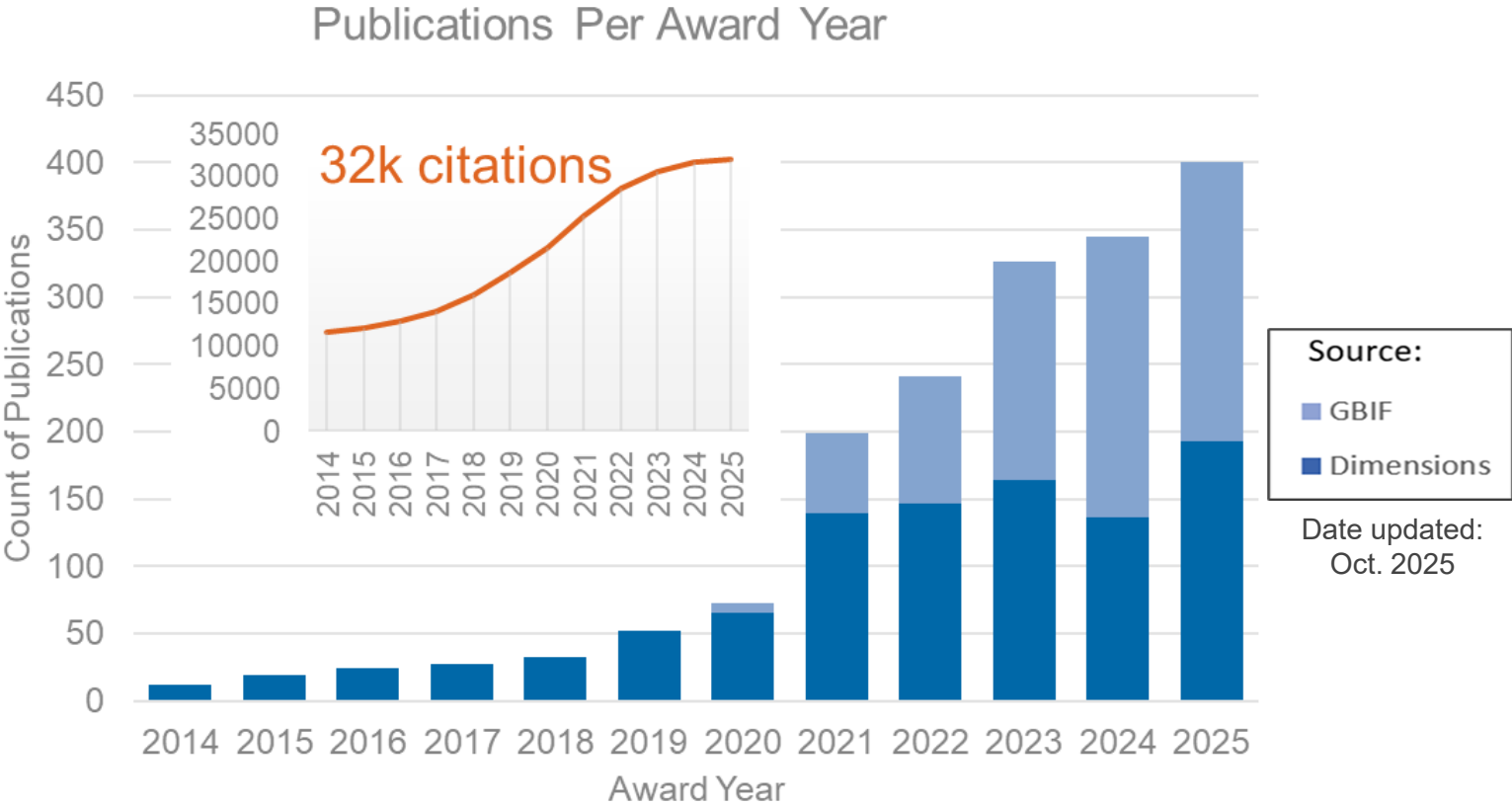
> 650,000 samples

> 5,200 taxa

- Small mammals
- Fishes
- Ground beetles
- Mosquitos
- Ticks

- Soil
- Dust
- ...many more

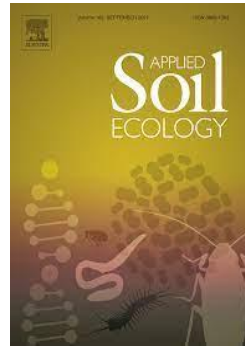
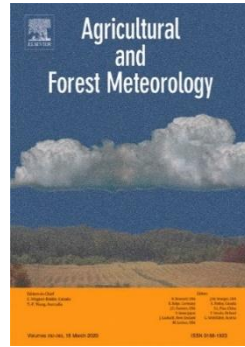
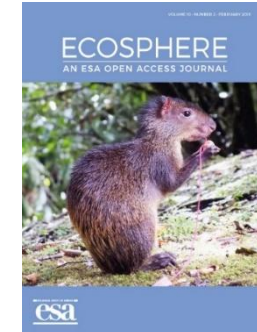
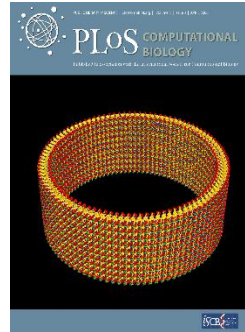
Research enabled by NEON data & specimens



NEON Publications in Dimensions (data, assets)



NEON Publications in GBIF (specimens)



Challenges present opportunities for understanding what AI-readiness can mean for NEON

Challenges

- **Operations:** coordinating fieldwork, sample collection, sensor maintenance, sample processing, etc., at scale
- **Evolving standards and technology:** maintaining long-term (30 year) data products as domain standards and technology evolve
- **Data management:** QAQC, provenance tracking, delivery of 180+ products
- **Meeting FAIR principles** in age of AI-enabled science, with a diverse user base

Opportunities

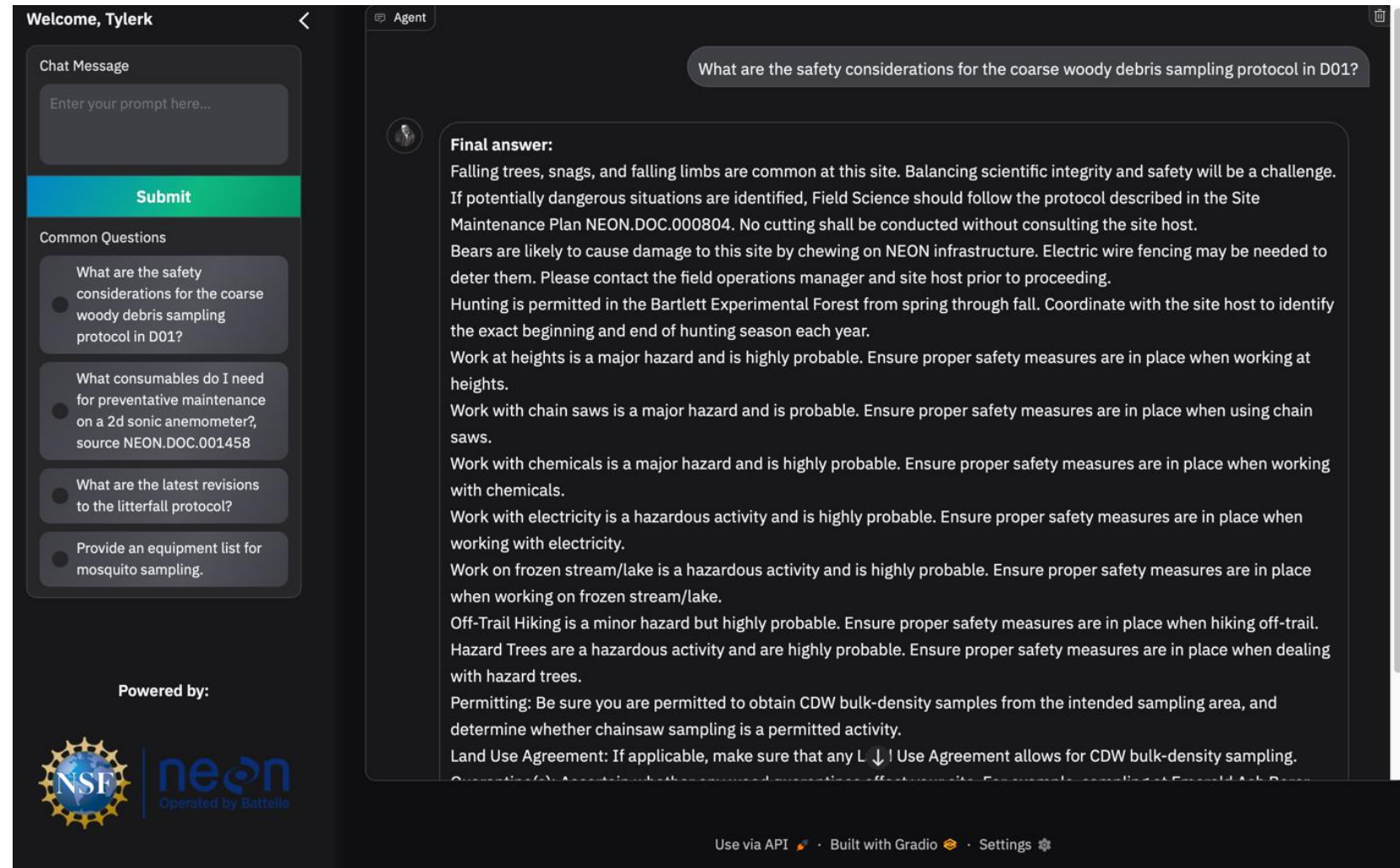
- **AI-FAIR documentation:** tools to gain efficiencies in operations (NEON VINE)
- **Open infrastructure:** NEON as a testbed/benchmark for new data acquisition technologies (specimen imaging, Sage AI)
- **Community of practice:** define FAIR4AI vs. AI-ready and develop standards in ecology
- **Partnerships:** support user projects developing AI-enabled science to understand requirements

AI-FAIR documentation

Tools to gain efficiencies in operations (NEON VINE)

NEON VINE: Virtual Interface for NEON Expertise

- Safety-forward AI assistant for NEON field staff (site/protocol-aware).
- Rapid help with safety protocols, info retrieval, field packing/prep, and protocol questions.
- Ingests and understands a wide range of document formats.
- Supports contextual, visual, and document-grounded conversations.
- Easy to update/extend with minimal engineering effort.
- **RAG-based:** grounds answers in up-to-date, context-specific documents to cut hallucinations and boost accuracy/relevance.



NEON VINE: Virtual Interface for NEON Expertise

Current Knowledge Base & Roadmap

- **NEON Site-Specific Safety Docs**
Detects sites/domains in queries • Pulls relevant guidance • Reinforces safe-practice reminders
- **NEON Observational Protocols (OS & AOS)**
Full protocol awareness • Answers from text or images • Provides field packing lists
- **NEON Sensor Maintenance Docs (TIS & AIS)**
Full maintenance procedures • Answers from text or images • Provides field packing lists

Potential additions

- Real-time weather and safety conditions
- Resolved ServiceNow ticket knowledge base
- Protocol training library (public)
- Data product information (public)
- Site history information (public)

— SCIENTIFIC LEARNING HUB

Explore NEON Data Tutorials

Access a curated collection of data science tutorials. Learn R and Python workflows to analyze ecological data from across the continent.

🔍 Search by topic, keyword, or language...

🔍 AI Search

All

.csv

.las

.laz

.tif

2013 Colorado Floods

AIS

AIS data

AOP

AOP Reflectance

AOP data

AOS data

API

API Ac

R

New Resources for Working with NEON Data

R

Quantifying The Drivers and Impacts of Natural Disturbance

PYTHON

Tutorial Introduction to Bidirectional Hyperspectral

Open infrastructure

NEON as a testbed/benchmark for new data acquisition technologies

✓ Node Status

□ All Nodes

(••) Sensors

17 Nodes | 5:53:32 PM



NEON



Project ▾

Focus ▾

City ▾

State ▾

Sensors ▾

☐ Show all



NEON Biorepository

- Small mammals
- Fishes
- **Ground beetles**
- Mosquitos
- Ticks
- Zooplankton
- Vascular plants & algae
- Microbes
- Soil
- Dust
- Precipitation
- ...and more!

Datasets:  imageomics/**sentinel-beetles**

Tasks:  Image Feature Extraction Modalities:  Image  Text

Libraries:  Datasets  Dask  Croissant + 1 License:  cc-by-4

 Dataset card  Data Studio  Files and versions  xet

Dataset Viewer

Split (2)
train · 22.4k rows

Search this dataset

file_path
image · width (px)



colorpicker_full_path
image · width (px)



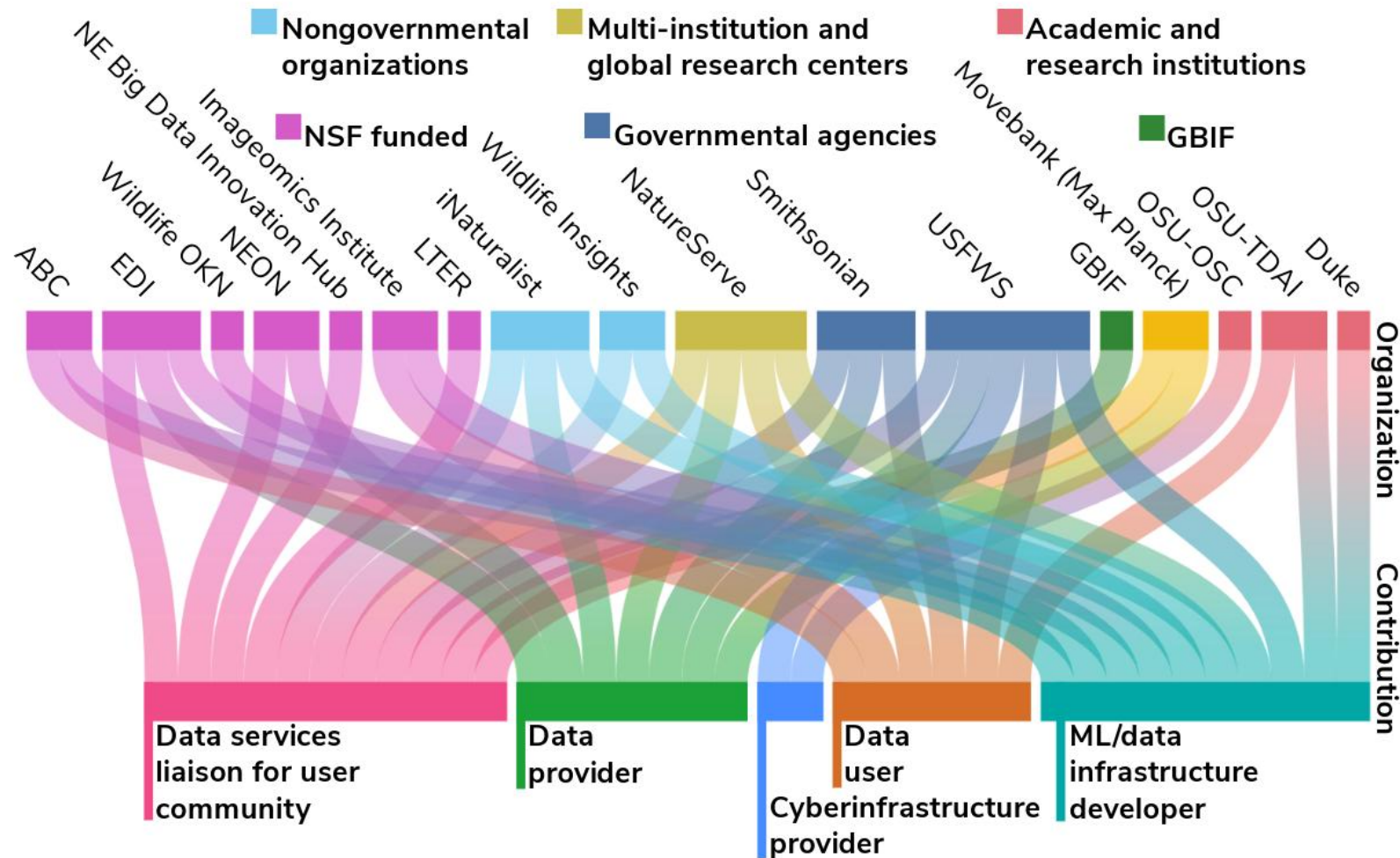
scalebar_full_path
image · width (px)



Community of practice

Define FAIR4AI vs. AI-ready and develop standards in ecology

Understanding the requirements for AI-ready ecology and biodiversity data infrastructure for open science



Partners:

- Imageomics Institute, Ohio State University, USA
Tanya Berger-Wolf, Arnab Nandi
- NEON, Battelle, USA
Eric Sokol, Christine Laney
- NatureServe, USA
Max Tarjan
- Duke University, USA
Hilmar Lapp

Continuum of data “readiness” for AI applications



FAIR4AI - Data that are FAIR enough for AI to be used in making data fully “AI-Ready”

AI-Ready - Data that are ready to be included in AI infrastructure at scale

From Alex White, Elizabeth Campolongo, Nick Lyon, and Hilmar Lapp, “AI Readiness” Defined. A presentation at the August 2025 Biodiversity AI Workshop

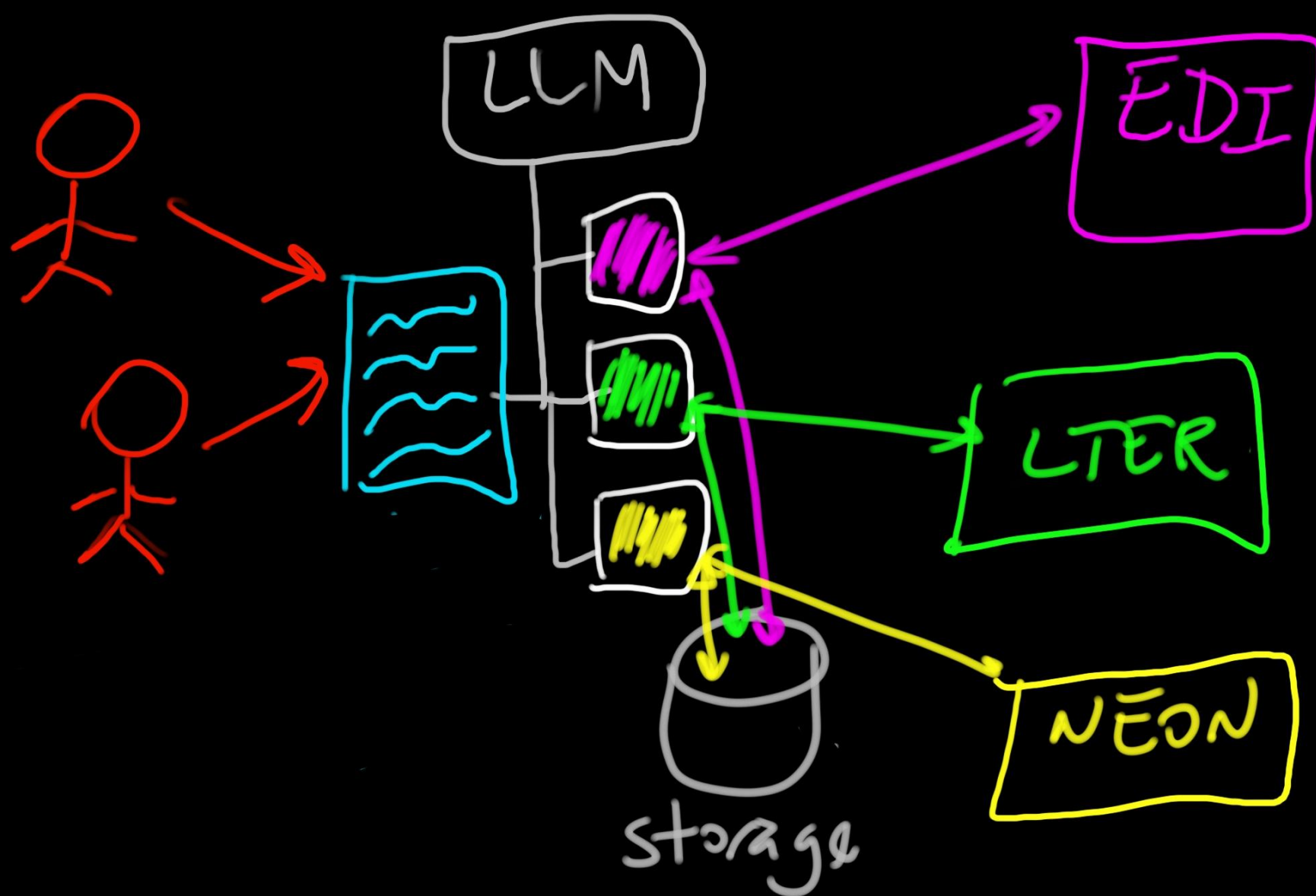
Partnerships

Support user projects developing AI-enabled science to understand requirements

Prototype NEON Agent for Environmental Data Exploration and Analysis

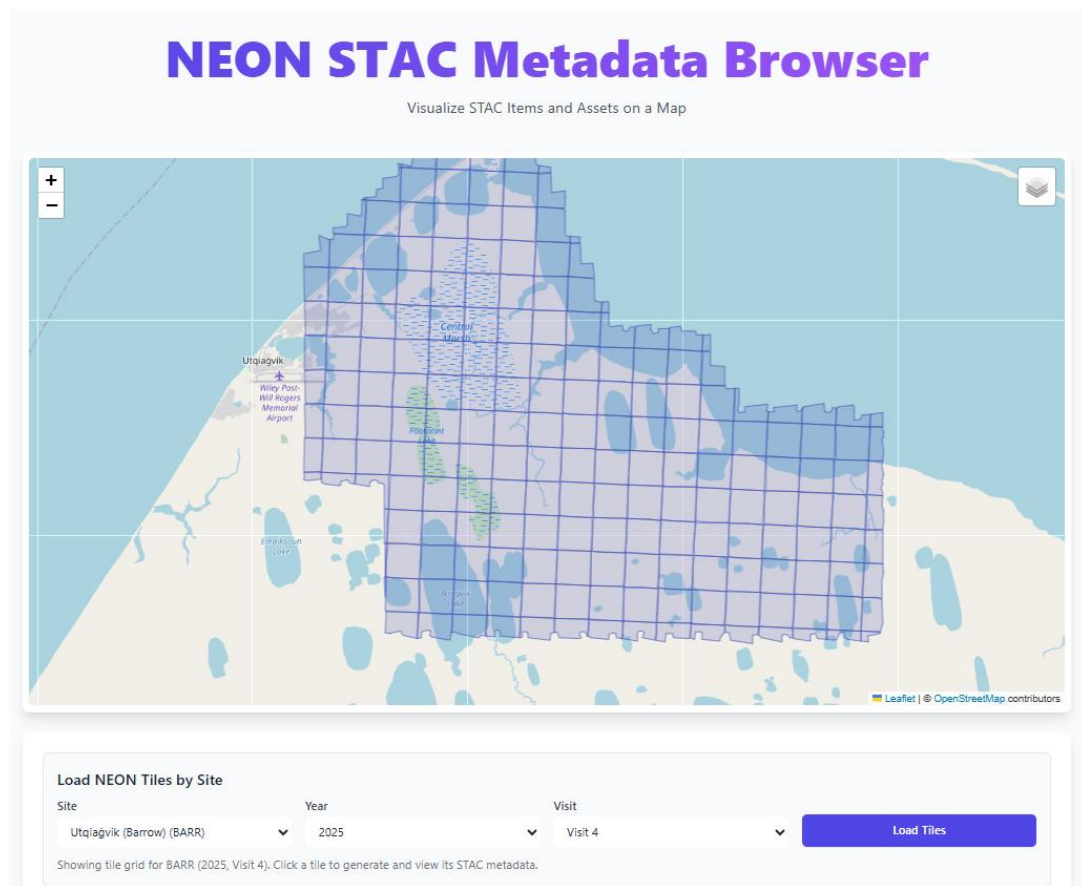
Renato Figueiredo, Laura Kuo, Zehua Zhu (EECS)
Julia Jones (Geography)
Oregon State University

In collaboration/co-development with:
Michael Elliott, Manny Luciano, Jose A. B. Fortes
ACIS Lab/ECE, University of Florida

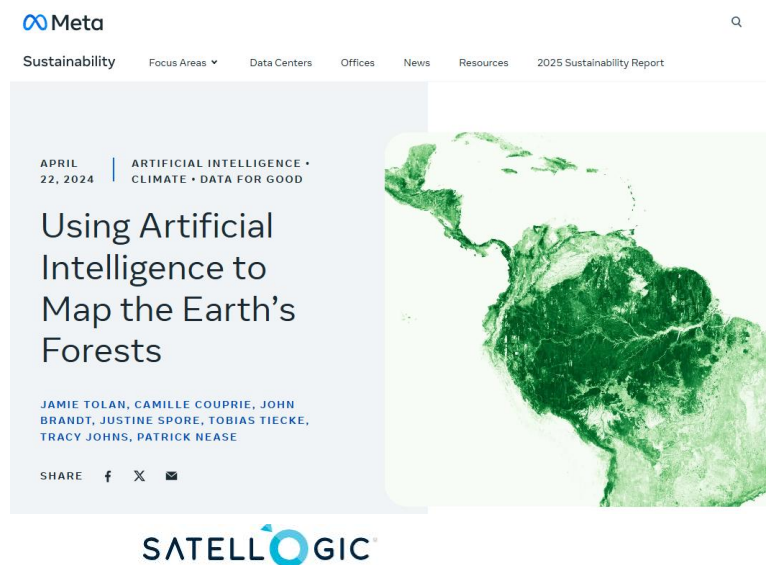


Vision: we are researching natural language agents that can access multiple repositories and extract data...

Airborne Remote Sensing



Prototype of STAC Metadata Browser for NEON Utqiagvik (BARR) site. AOP is working towards improved standardization of data, shapefiles, and metadata to enable this enhanced representation.



Meta / World Resources Institute AI application using NEON CHM data along with satellite data to map global forests.



Satellogic incorporated NEON hyperspectral and lidar data in an AI training dataset on HuggingFace, called EarthView.



Scientific Modeling out of distribution (Scientific-Mood) ML Challenge

- **Neural Forecasting:** We forecast the activations of a cluster of neurons given previous signals from the same cluster. This targets the critical problem of brain-artificial neuron interfaces, and these models can be used in brain-chip interfaces for artificial limb control, amongst many others.
- **Climate prediction using ecological data:** We predict drought conditions at field sites over short, medium, and long timescales using images of an important group of ecological indicator organisms (ground beetles) collected there. The resulting models can be used to understand how the impacts of future climate conditions may be reflected in observable features of sentinel taxa.
- **Coastal flooding prediction over time:** We model the sea levels at various sites over decades, with the aim of predicting future coastal floods. The resulting model will be essential for understanding the real-world impacts of climate change.



neon
Operated by Battelle

720.746.4844 | neonscience@battelleecology.org | neonscience.org

>180 Data Products

Atmospheric



H₂O, CO₂
Heat
Micromet
Isotopes
Turbulence
Storage
Fluxes

Organismal



Abundance
Composition
Biomass
Productivity
Pathogens
Phenology
Marker Genes
Metagenomics

Ecohydrology



Water quality
Precipitation
Discharge
Radiation
Geomorphology
Riparian Structure

Biogeochemistry



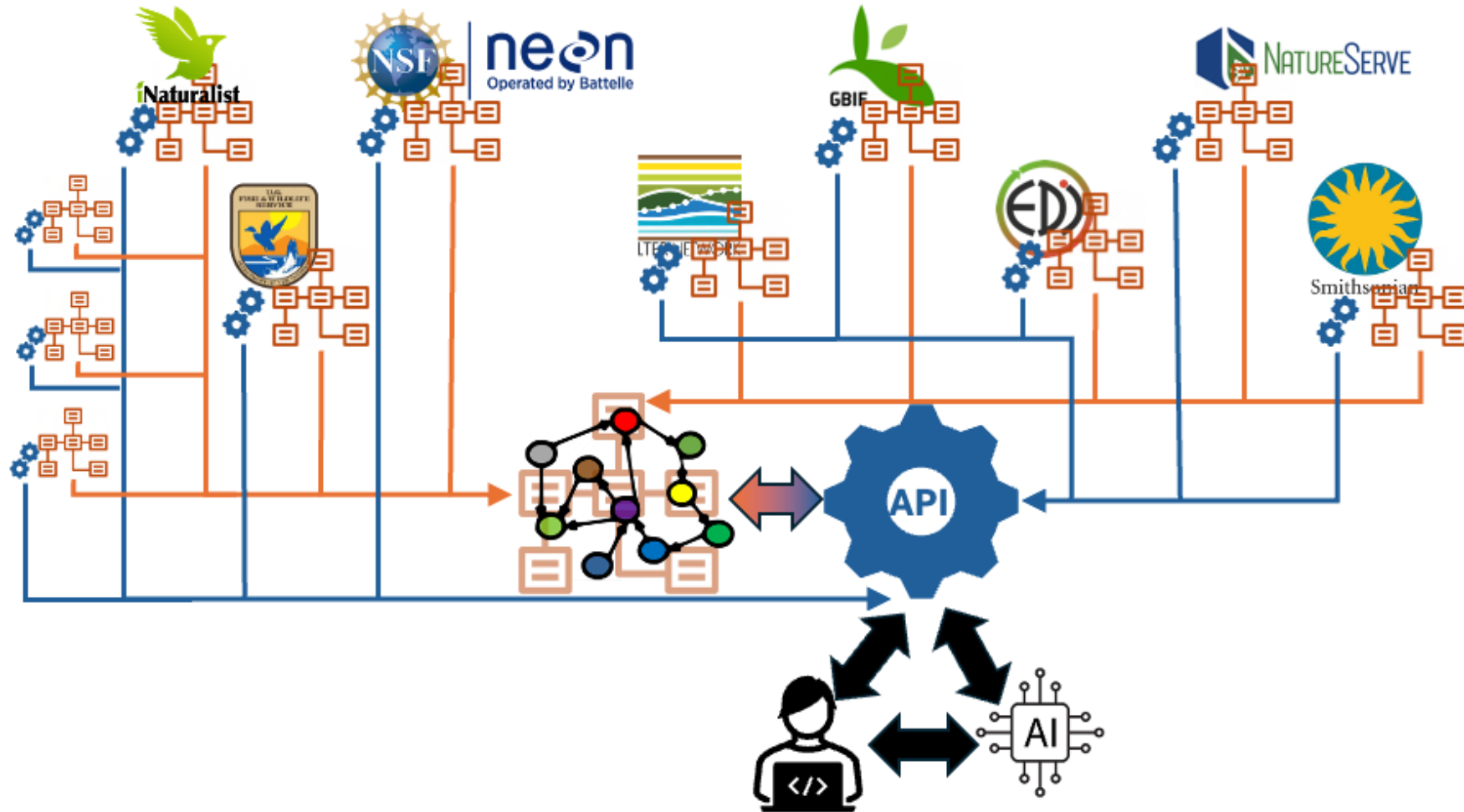
Soil conditions
Chemistry
Particulates
Foliar characteristics

Remote Sensing



Spectrometry
Hi-Res Imagery
LiDAR

A vision for AI-ready biodiversity data cyberinfrastructure



Objective: Predict drought status from characteristics of ground beetles collected during a sampling event.

Target variable: *Standardized Precipitation Evapotranspiration Index (SPEI)*

- Derived from remote sensing data for a given location (e.g., NEON Site)
- Calculated over a given time window (e.g., 30 days, 1 year, 2 years)
- Range is [-3.0, 3.0]
- < -1 indicates dry conditions
- > 1 indicates wet conditions

