HOLOS
BERKELEY ECOINFORMATICS ENGINE

Rescuing and integrating biological and environmental data in the face of global change

The challenge

Predicting biotic responses to global environmental change necessitates a holistic understanding of the complex interactions and feedbacks among organisms, climate, and their physical and biotic environments across space and time.

This level of understanding can only be achieved through the integration and analysis of diverse data types spanning a range of biological, spatial, and temporal scales.

Our mission

- Rescue and integrate a wealth of data including specimens in natural history collections, field surveys and observations, aerial and satellite imagery, measurements from environmental sensor networks, and global change model predictions.

- Provide an open, technical infrastructure for researchers and policymakers to explore, visualize, and analyze this wealth of data on global change.

Data Science

The Berkeley Ecoinformatics Engine aims to be model for informatics that promotes open science and platform agnostic toolkits. We create web-based building blocks to access data from established repositories while promoting best practices to uncover 'dark' datasets from orphaned or remote labs. Our architecture is structured to promote use of the biodiversity and environmental datasets in data-driven, transparent platforms. We welcome developers and researchers of all skills to use the Ecoengine tools and hopefully be inspired to create their own with Ecoengine-based applications.

Berkeley Ecoinformatics Engine Framework

holos.berkeley.edu (Library Interface)
ecoengine.berkeley.edu (API)

Mock-up

Data Access

The Ecoengine has multiple means of access and use.

- Users access data and its metadata via the Online Portal, named Holos, our browsable Library Interface. Holos will come with built-in searching, charting, visualization and analysis tools.

- The Data API's open framework will allow third-party developers to write their own tools, including visualization and charting apps and expert analysis tools, like ROpenSci has done (see side bar).

- The Ecoengine is a self-describing API built on the Django REST framework.

- Output from GET requests is provided as GeoJSON, XML, a browsable API format, and a map and photo view if applicable.

- Resources can be searched and filtered either directly, using a number of query parameters, or indirectly, using a search engine to generate faceted results.

Apps
(Internally and Externally Developed)

Expert Analysis Tools
(See R package!)

Data RESTful API
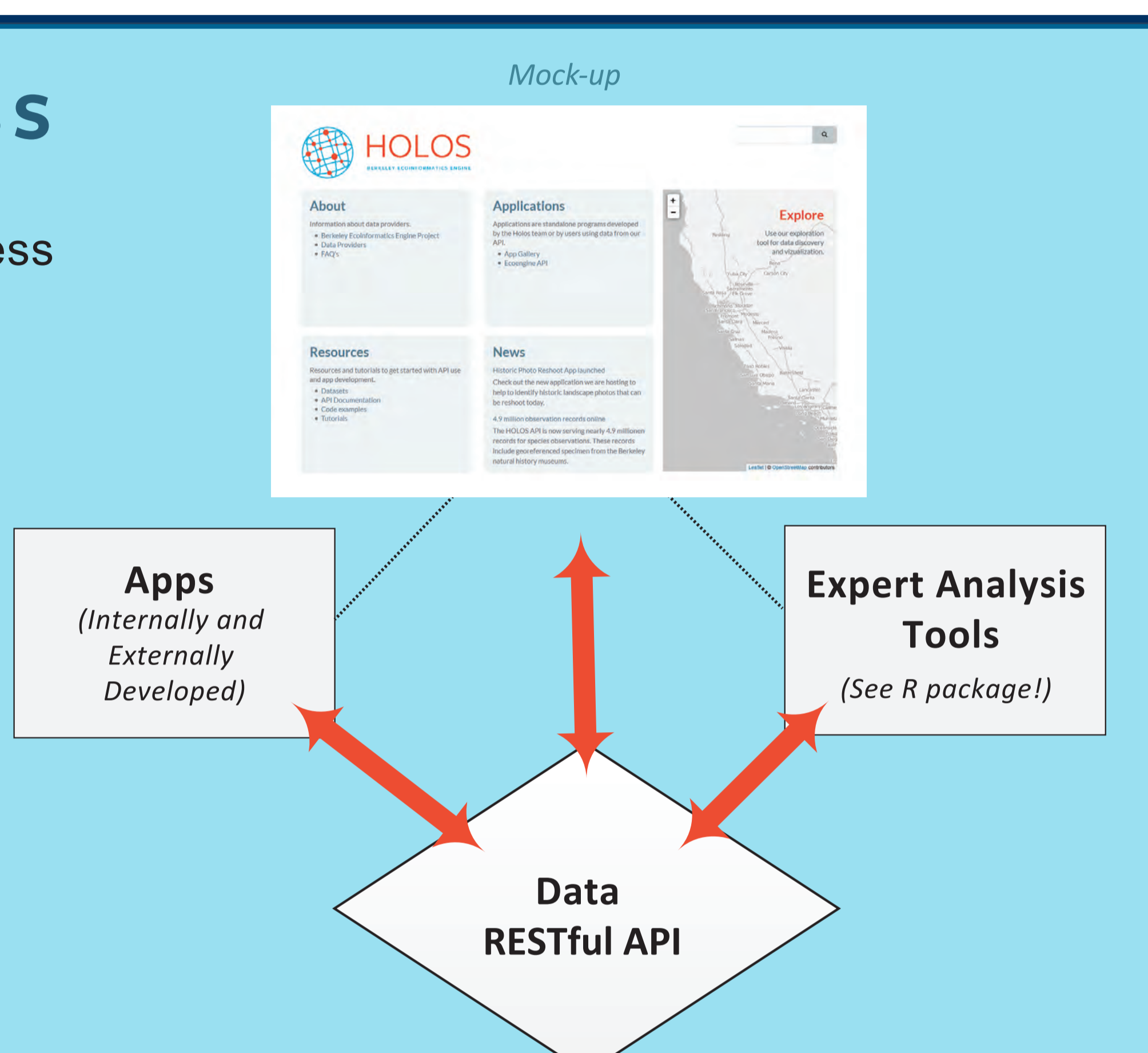
Data Resources

The Ecoengine integrates two basic data types, *Events* and *Baselayers,* from a diversity of sources:

Events → Specimens, Photos, Sensors

Baselayers → Thematic, Continuous, Vector

Events - specimens and observations

- DarwinCore archives of biocollection datasets including Berkeley Natural History collections:
  - Museum of Vertebrate Zoology
  - UC & Jepson Herbaria
  - Essig Museum of Entomology
  - Museum of Paleontology
  - Consortium of California Herbaria
- Datasets from field stations
  - Species checklists
  - Environmental monitoring network data (e.g., 4.1 million geophysical sensor measurements, Keck HydroWatch Center, Angelo Coast Range Reserve FS)
- Historic and current vegetation maps, habitat and species photos, and data from / California's Vegetation Type Mapping project (1940's)
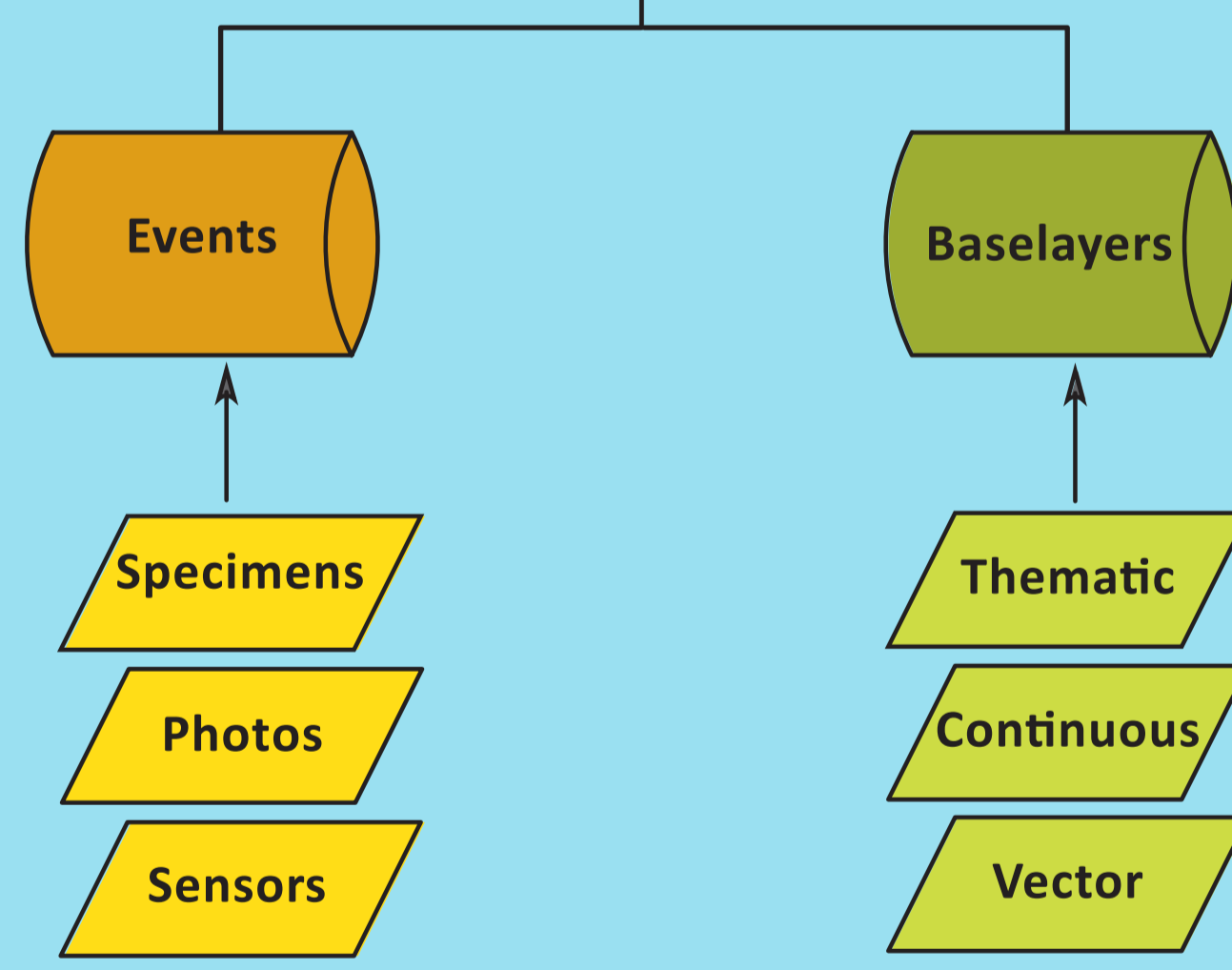
Baselayers- GIS datasets

- Models of past, current and future climates: including newly downscaled paleoclimate models of the Last Glacial Maxima and the Mid-Holocene

- Land Cover and remote sensing:
  - e.g., MODIS, National Land Cover Database, Cropland

- Topography, Geology, Soils, Disturbance Regimes

- Ecoregion, political and nature reserve boundaries
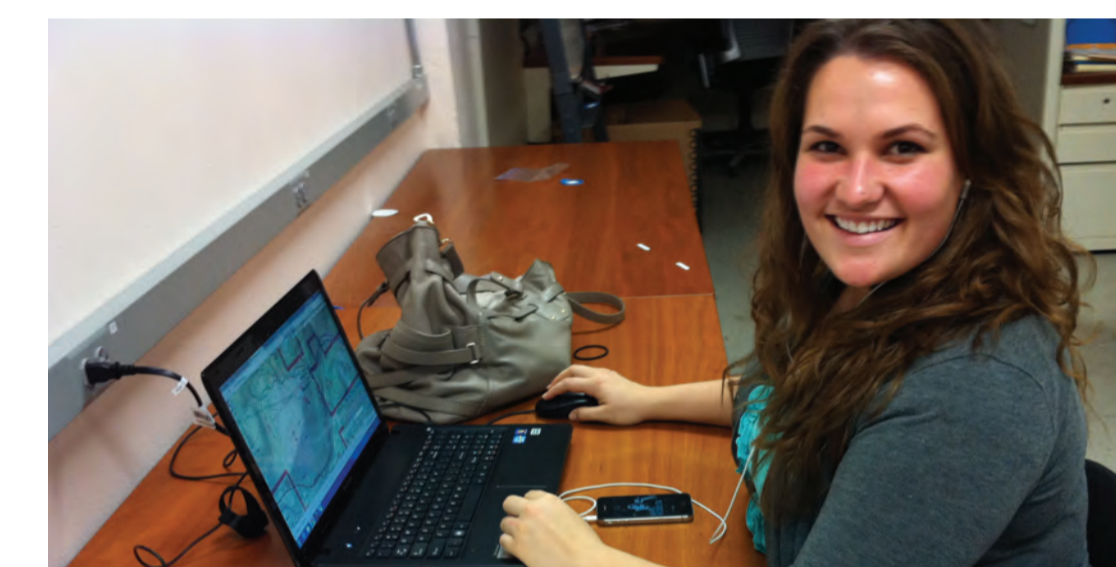
Accessing the Ecoengine with R

Collaboration partners ROpenSci has used the Ecoengine API to create a programmatic way to query, download, view and analyse data in the Ecoengine. Their R package, *ecoengine*, lets users access the Ecoengine in the R Statistical programming environment, already a familiar part of the workflow of many scientists.

ropensci.org/packages
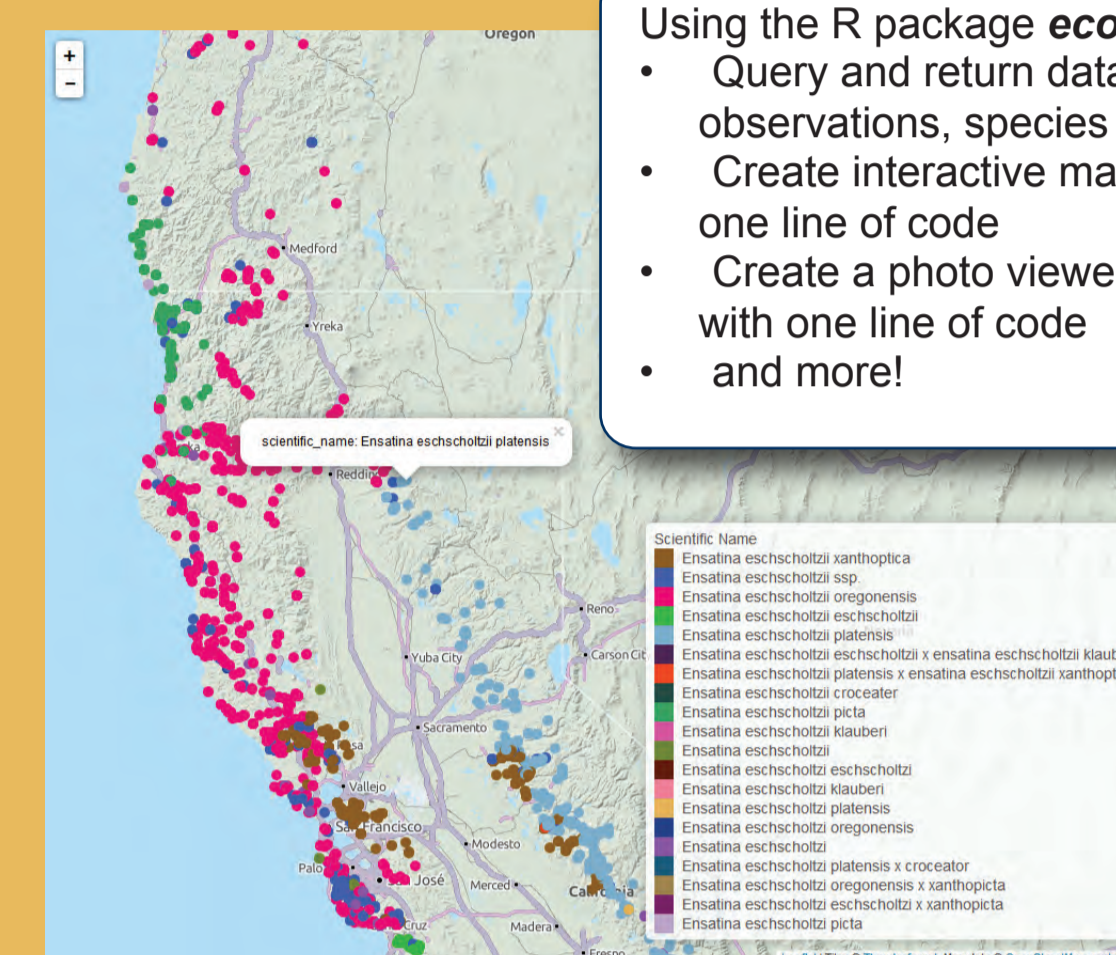
Using the R package *ecoengine*:
- Query and return data from photos, observations, species checklists, sensors
- Create interactive maps of results with one line of code
- Create a photo viewer website with one line of code
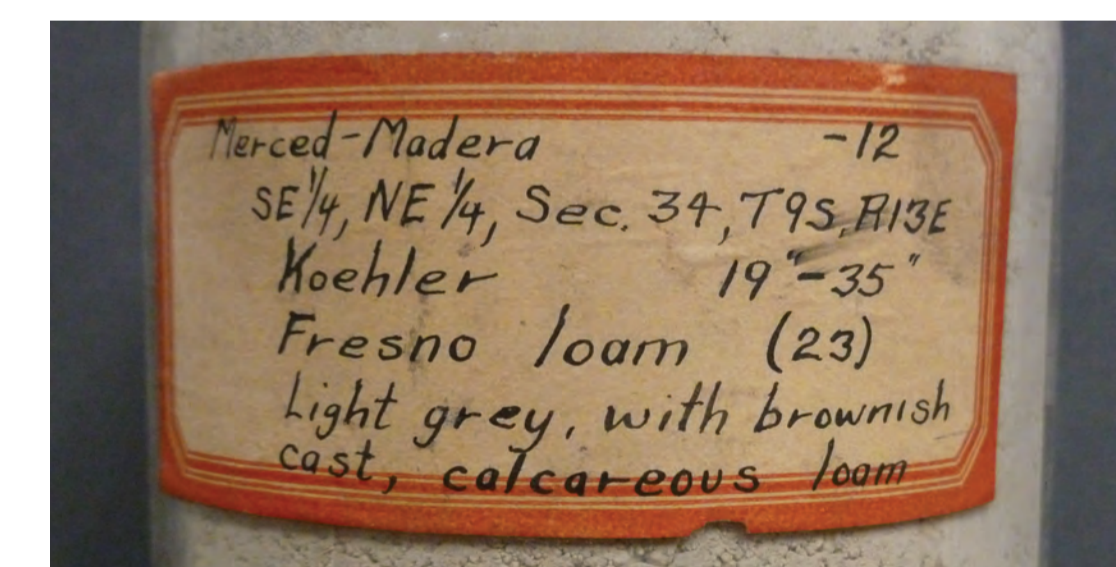- and more!

Ecoengine Photo Viewer

Rescuing Data

FIELD STATION DATASETS & SPECIES CHECKLISTS
We are actively cataloging and digitizing all past and present species checklists and observations, long-term ecological studies, experimental studies, and climate and weather measurements that are associated with the four UC Berkeley Field Stations.

HISTORIC SOILS
Over 12,000 historic soil samples, invaluable for habitat reconstructions, and dating from the 1920-1970's, are being imaged and cataloged.

POLLEN CORES
50 sediment cores collected in California are now in the UCMP collections database and public. Pollen data associated with these cores are being added and will significantly improve our ability to reconstruct vegetation changes over the last 15,000 years.

Integrating Data
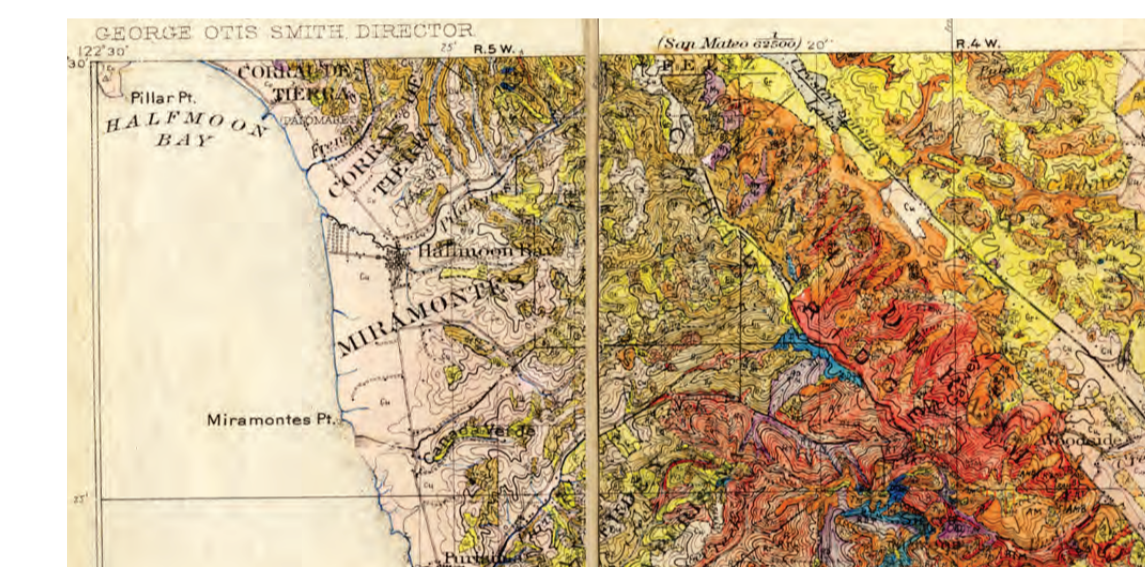
VEGETATION TYPE MAPPING
The Wieslander Vegetation Type Map (VTM) Project is a collection of plots maps, vegetation maps, photos, and specimens surveyed in the 1920s and 30s throughout California. The data provide a snapshot of the state's vegetation in the early 20th century, making the collection an invaluable resource for examining changes in land cover and use, and habitat restoration.

The goal of the current VTM Project, a collaboration between UC Berkeley and UC Davis, is to digitize and integrate the datasets to inform studies of global change. We have initiated two such studies: 1) identifying 20th century changes in vegetation structure in coastal and southern California by comparing USFS data to VTM data, and 2) identifying factors associated with vegetation diversity and identity.

Project Coordinator
Sarah Hinman

Lead PI's
Rosemary Gillespie
Charles Marshall

Co-PI's
David Ackerly
Maggi Kelly
Michelle Koo
Kevin Koy
Craig Moritz
Mary Power

Development Team
Collin Bode
Brian Galey
Joyce Gross
Ginger Ogle
Falk Schuetzenmeister
Jessica Rothery

Postdocs
Giovanni Rapacciuolo

Partners
Karthik Ram, UC Berkeley
John Chiang, UC Berkeley
Jim Thorne, UC Davis

BiGCB Coordinator
Verna Bowie