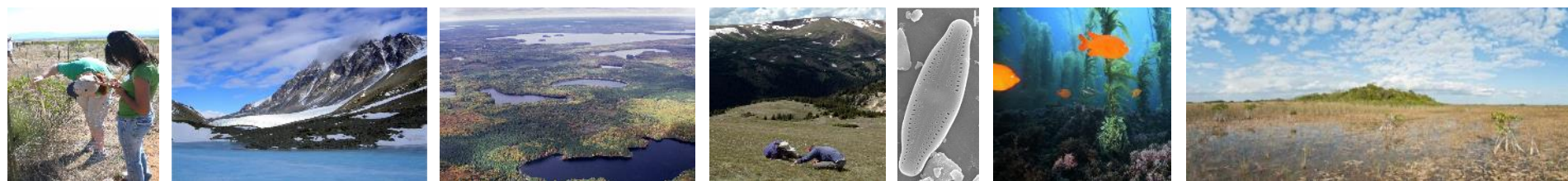




Insights from the synthesis of long-term biodiversity data: resources and tools available to community ecologists



Eric R. Sokol (eric.r.sokol@colorado.edu | esokol@battelleecology.org)

National Ecological Observatory Network (NEON), Battelle Ecology, Inc. (<http://www.neonscience.org>)

Institute of Arctic and Alpine Research (INSTAAR), University of Colorado Boulder

Christopher M. Swan

University of Maryland Baltimore County

Nathan Wisnoski

Indiana University

A growing commitment to FAIR Data Principles is facilitating synthesis in ecology



Data and supplementary materials have sufficiently rich metadata and a unique and persistent identifier.

FINDABLE



Metadata and data are understandable to humans and machines. Data is deposited in a trusted repository.

ACCESSIBLE



Metadata use a formal, accessible, shared, and broadly applicable language for knowledge representation.

INTEROPERABLE



Data and collections have a clear usage licenses and provide accurate information on provenance.

REUSABLE

COPDESS

Coalition for Publishing Data in the Earth and Space Sciences

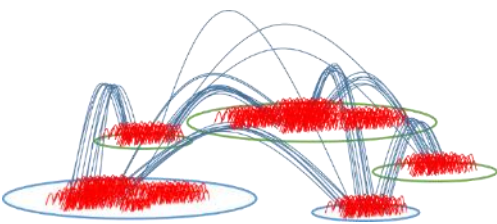
SIGNATORIES

34 Repositories have signed on, including:

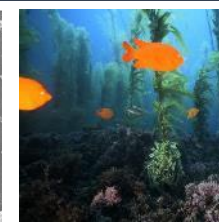
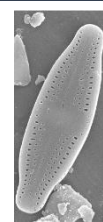
- Environmental Data Initiative (host LTER data)
- National Ecological Observatory Network (NEON)
- Arctic Data Center
- UNAVCO
- Dryad
- KNB Data Repository



Find commitment statement and signatories here: <https://copdess.org/enabling-fair-data-project/commitment-statement-in-the-earth-space-and-environmental-sciences/>



Using long term data to understand links between environmental variability and metacommunity stability



Eric R. Sokol (eric.r.sokol@colorado.edu)

National Ecological Observatory Network (NEON), Battelle Ecology, Inc. (<http://www.neonscience.org>)

Institute of Arctic and Alpine Research (INSTAAR), University of Colorado Boulder

LTER Metacommunities Synthesis Working Group

<https://www.nceas.ucsb.edu/projects/12749>

<https://github.com/sokole/ltermetacommunities>

Contributors

N.I. Wisnoski¹, C.M. Swan², R. Andrade³, A. Compagnoni⁴, M. Castorani⁵,
L. Hallett⁶, T. Lamy⁷, N. Lany⁸, L. Marazzi⁹, S. Record¹⁰, J. Tonkin¹¹, N. Voelker²,
P. Zarnetske⁷, C. Catano¹², A. Smith³, and others

¹ Department of Biology, Indiana University, Bloomington

² Geography and Environmental Studies, University of Maryland

³ School of Geographical Sciences and Urban Planning, Arizona State University

⁴ German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig, Martin Luther University Halle-Wittenberg

⁵ University of Virginia

⁶ University of Oregon

⁷ Marine Science Institute, University of California, Santa Barbara

⁸ Dept. of Forestry, and Ecology, Evolutionary Biology and Behavior Program, Michigan State University

⁹ Southeast Environmental Research Center, Florida International University

¹⁰ Bryn Mawr College, Pennsylvania

¹¹ Department of Integrative Biology, Oregon State University

¹² Department of Plant Biology, Michigan State University



The Long-Term Ecological Research (LTER) Metacommunities working group is supported by the LTER Network Communications Office and funded by NSF grant # DEB-1545288.

The need for long-term data

Long-Term Ecological Research (LTER) Metacommunities *Synthesis working group*

1. Compare metacommunities from different LTER sites
2. Do species' dispersal characteristics predict biodiversity stability?
3. Does environmental variability predict biodiversity stability?



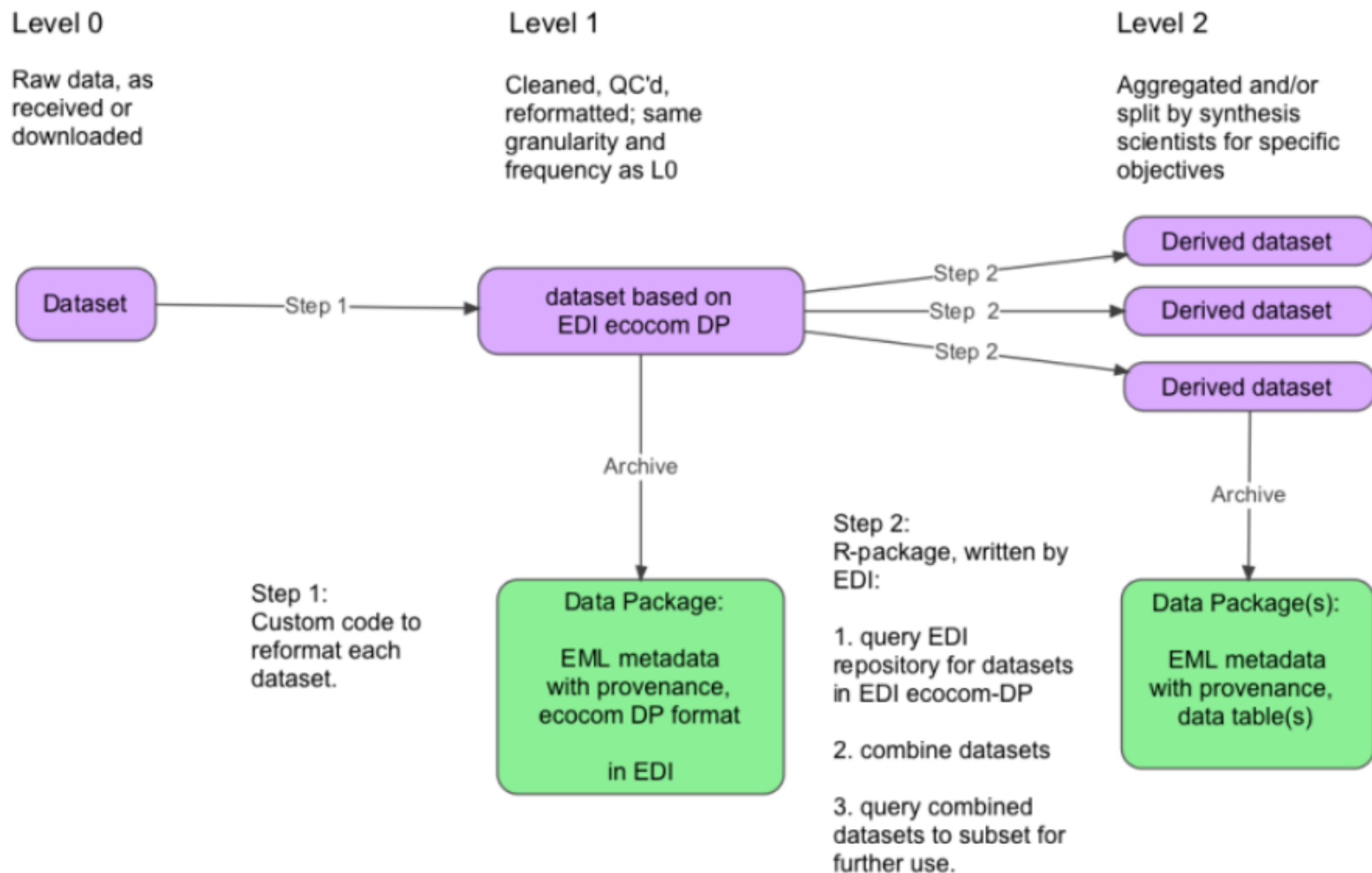


Figure: Abstract view of dataset levels. A flexible intermediate (L1, middle) lies between datasets of primary observations (L0, left) and the aggregated views used by synthesis projects. If datasets are in a recognized format, EDI can create tools for some basic functions

Findable Accessible Interoperable Reproducible

Challenges to finding data

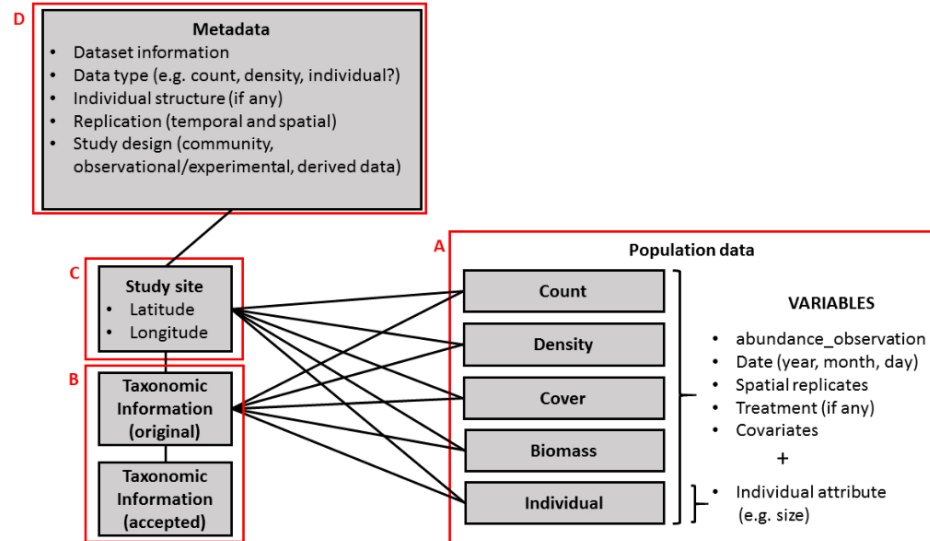
- Where do I start?
- Are relevant data sets actually discoverable?
- How do keywords map to actual data? Do keywords map consistently?
- How do I know if a data actually meet my criteria?
(LTER metacommunities looking for spatial and temporal replication)

Findable Accessible Interoperable Reproducible

Tools: `popler` R package and database (Compagnoni et al.)
<https://github.com/ropensci/popler>



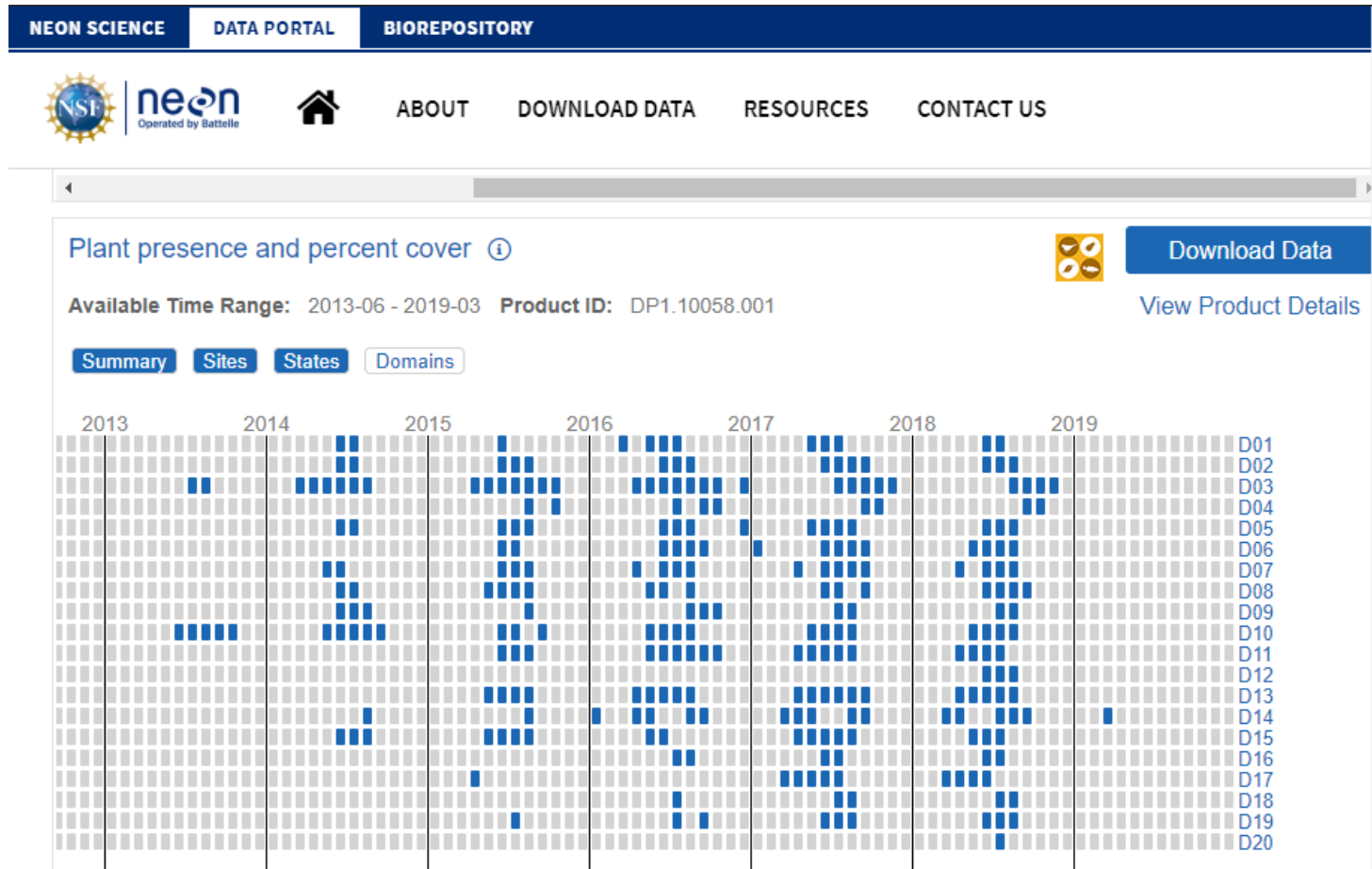
An R package to browse and query population-level datasets from the US Long Term Ecological Research (LTER) network



Findable Accessible Interoperable Reproducible

Tools: NEON data browser

<https://data.neonscience.org/static/browse.html>



Findable Accessible Interoperable Reproducible

Tools: `ecocomDP` R package and data model
<https://github.com/EDlorg/ecocomDP>



Environmental Data Initiative

Use `ecocomDP` data

- Find:
 - Use the `view_all_ecocomDP()` function to list all `ecocomDP` datasets. This function is apart of the `ecocomDP` R package.
 - [Environmental Data Initiative \(EDI\)](#) Enter "ecocomDP" in the 'simple search' box in the EDI data repository.
 - [National Ecological Observatory Network \(NEON\)](#) Use the `view_all_ecocomDP()` function to list all NEON data available in the `ecocomDP` format. This function is apart of the `ecocomDP` R package.

Findable Accessible Interoperable Reproducible

Challenges to accessing data

- How easy is it to import data into my R workspace?

Findable Accessible Interoperable Reproducible

Tools: popler R package and database (Compagnoni et al.)
<https://github.com/ropensci/popler>



It only takes 2 lines of code to find and import LTER data into your R working environment!

```
# create a browse object and use it to get data

penguins <- pplr_browse(lterid == 'PAL')

# unpack covariates as well

penguin_raw_data <- pplr_get_data(penguins, cov_unpack = TRUE)
```

Findable Accessible Interoperable Reproducible

Tools: `neonUtilities` R package

CRAN: <https://cran.r-project.org/web/packages/neonUtilities/index.html>

GitHub: <https://github.com/NEONScience/NEON-utilities>



Load data into your R environment with 1 line of code!

```
# To download plant foliar properties data from all sites, expanded data package:  
cfc <- loadByProduct(dpID="DP1.10026.001", site="all", package="expanded")
```

Find tutorials here: <https://www.neonscience.org/resources/data-tutorials>

Findable Accessible Interoperable Reproducible

Tools: `ecocomDP` R package and data model
<https://github.com/EDlorg/ecocomDP>



Environmental Data Initiative

Browse, query, and download LTER, Macrosystems, LTREB, NEON data sets and more!

Works through API calls to data repositories, so data sets are current

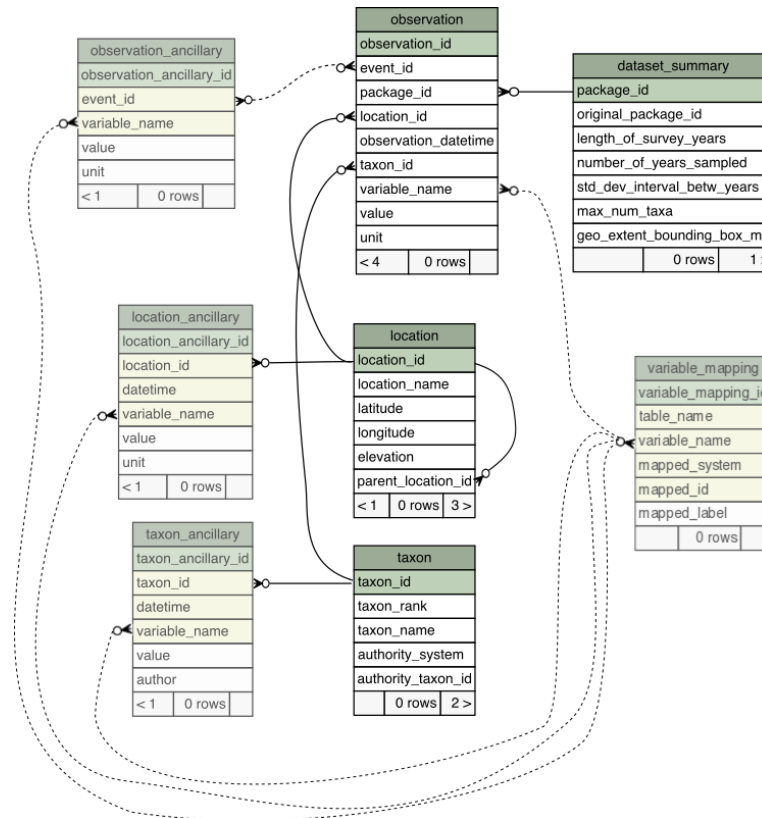
Findable Accessible Interoperable Reproducible

Challenges to comparing across datasets

- Data formatting
- Observational scale and grain size
- Taxonomic concepts (biodiversity data)
- Comparable metadata, covariates (environmental and spatial data)

Findable Accessible Interoperable Reproducible

Tools: ecocomDP Standard data pattern



Use `aggregate_ecocomDP()` to combine datasets.

Findable Accessible Interoperable Reproducible

Challenges to maintaining reproducible workflows

- Will data sources always be findable, accessible?
- Can data munging for a particular project be re-created?
- How easy is it for others to make the same calculations?

Findable Accessible Interoperable Reproducible



Tools: Version controlled scripted workflow and R packages

- Data munging:
LTER metacommunities github page
<https://github.com/sokole/ltermetacommunities>
- Metacommunity variability metrics
ltmc package for R
<https://github.com/sokole/ltermetacommunities/tree/master/ltmc>

Acknowledgements

NSF LTER DEB-1545288



NATIONAL SCIENCE FOUNDATION

LTER NETWORK
LONG TERM ECOLOGICAL RESEARCH



NCEAS

National Center for Ecological Analysis and Synthesis

To download these slides:

https://github.com/sokole/ltermetacommunities/tree/master/ESA_2019/Sokol_INSPIRE