

Going Beyond the Spreadsheet

Developing Best Practices in 'long-tail' environmental data
curation and publishing



C. Gries, R.F. Brown, G.Gastil-Buhl, S. Elmendorf, H. Garritt,
M. Martin, G. Maurer, A. Nguyen, J.H. Porter, T. Whiteaker



Introduction

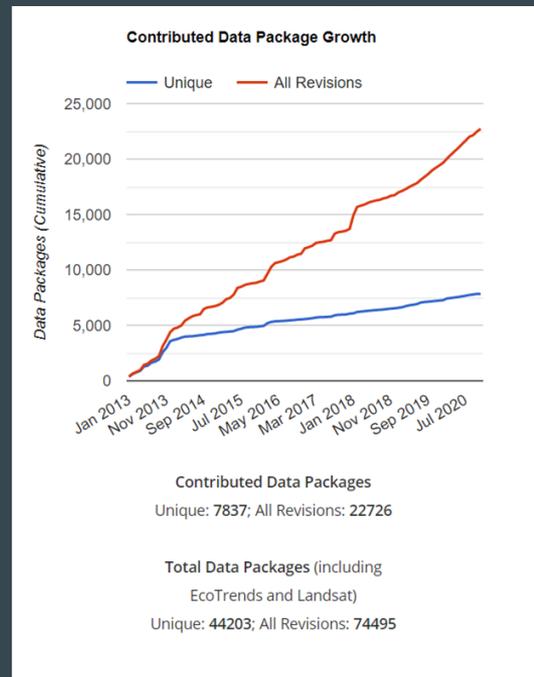
Environmental Data Initiative repository

FAIR data repository

Ecological Metadata Language

Data package: data table(s), metadata

Metadata generation tools

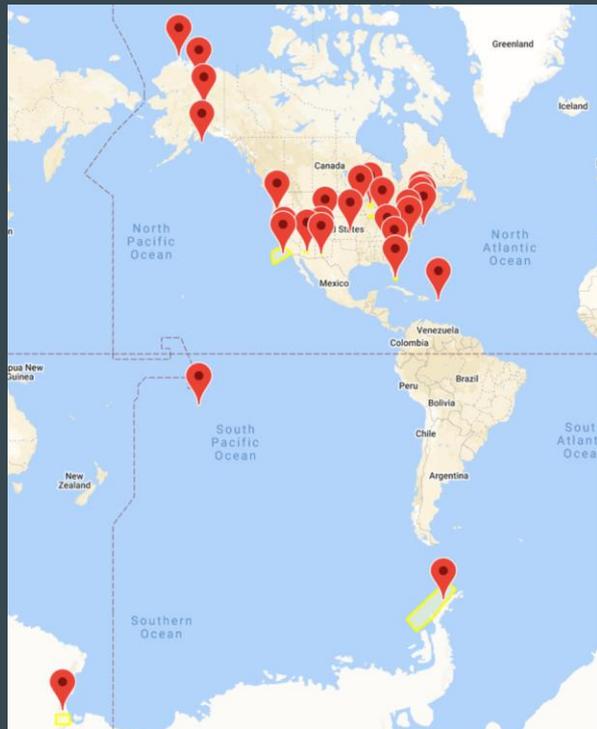


Introduction

Long-Term Ecological Research network

40 years of data management

Professional data managers at each site



Archiving other data types

Modeled Data

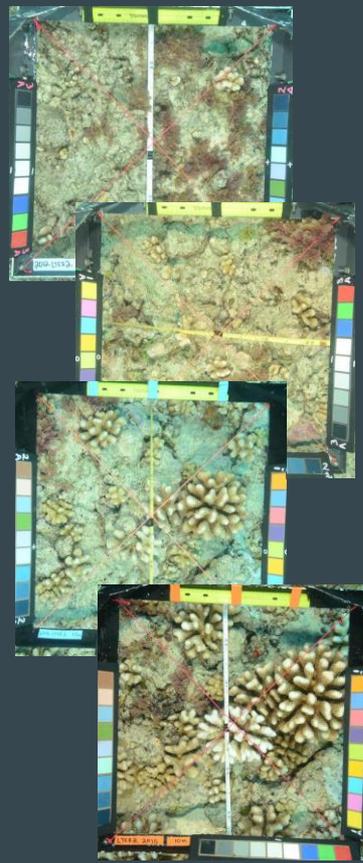
Code and Scripts

Data in More than one Repository

Images and Documents as Data

Spatial Data

Data from Small, Moving Platforms



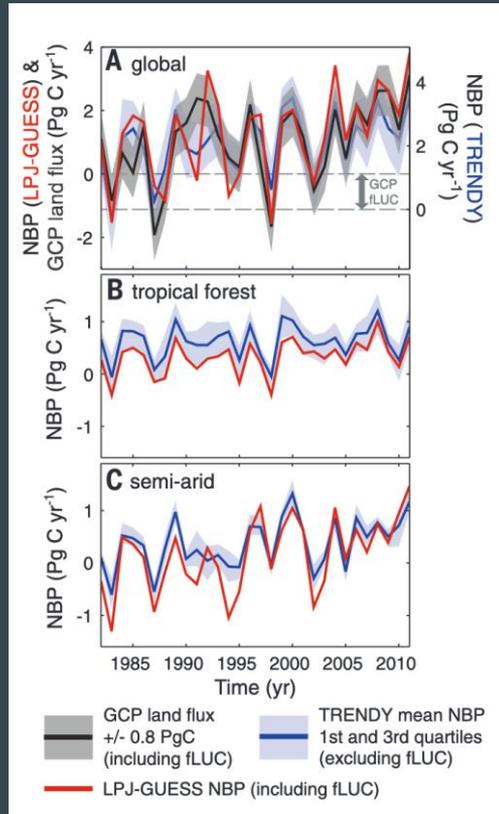
Modeled Data

Transparency / Reproducibility

Parameter settings, input data, output data

Model code

Size



From Ahstrom et al 2015



Code and Scripts

```
6  
7 # Fix any interval or ratio columns mistakenly read in as nominal and nominal columns read as numer  
8  
9 if (class(dt1$SEQ)=="factor") dt1$SEQ <-as.numeric(levels(dt1$SEQ))[as.integer(dt1$SEQ) ]  
0 if (class(dt1$YEAR)=="factor") dt1$YEAR <-as.numeric(levels(dt1$YEAR))[as.integer(dt1$YEAR) ]  
1 if (class(dt1$MONTH)=="factor") dt1$MONTH <-as.numeric(levels(dt1$MONTH))[as.integer(dt1$MONTH) ]  
2 if (class(dt1$LONG)=="factor") dt1$LONG <-as.numeric(levels(dt1$LONG))[as.integer(dt1$LONG) ]  
3 if (class(dt1$LAT55)=="factor") dt1$LAT55 <-as.numeric(levels(dt1$LAT55))[as.integer(dt1$LAT55) ]  
4 if (class(dt1$LAT525)=="factor") dt1$LAT525 <-as.numeric(levels(dt1$LAT525))[as.integer(dt1$LAT525) ]  
5 if (class(dt1$LAT50)=="factor") dt1$LAT50 <-as.numeric(levels(dt1$LAT50))[as.integer(dt1$LAT50) ]  
6
```

Model code, data manipulation scripts (not binary software packages)

In EDI / code repository

Inside data package / referenced

CodeMeta file



Data in More than one Repository

Related datasets managed in different repositories

E.g., GenBank, FluxNet, IDigBio

Data inventory table (accession #)



```
ACGGTAGCTAATACCGCATAACGTCGCAAG/  
TTACTAGCGATTCCAAC TTCATAAGGTCGAG  
CGTATTCACCGCGGCATGCTGATCCGCGATT/  
GGGGAAAGATTTATCGCCAAAAGATTGGCC
```



Images and Documents as Data

Images, digital field sheets, lab notebooks, protocols

Reanalysis, transparency, documentation

Design data packages for usability

Data inventory table

Sequence # 6301 A

ATER BIOLOGICAL FIELD DATA

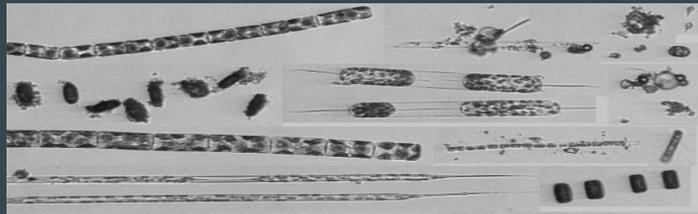
Lake SP Station 1 Date 7 SEP 2020 Page 1 of 2

Perm. Anchor? N
Observers P. S. P. K. M.
On Station Time 923
Off Station Time 1105

Equip: 2m Sch-P 53umesh PERISTALTIC PUMP
Start: 940
End: 1038

| DEPTH (m) | ZOOPL. SAMPLE TAKEN | CHLOROPHYLL | | | PHYTOPLANKTON | |
|-----------|---------------------|---------------|---------|-----------------|---------------|-----------------------|
| | | FILTER LETTER | MAX PSI | VOLUME FILTERED | SPAN (m) | STRATUM VOLUME PUMPED |
| Surf. | | D | 15 | 2700 | | |
| 1 | X | | | | | |
| 2 | | | | | | |
| 3 | X | R | 15 | 2790 | | |
| 4 | | | | | | |
| 5 | X | J | 15 | 2860 | | |
| 6 | | | | | | |
| 7 | X | | | | | |
| 8 | | R | 15 | 1670 | | |

Air Temp (C) 8
Wind Sp. (mph) 4
Wind Direction NW
Wave Size (cm) 5
% Cloud Cover 70
Bottom at (m) 28.1
with (gear) Sinker
Secchi Disk Depth (m) 7
Water Temp (C) 11.9



Spatial Data

Widely readable format (shapefile, geoTIFF)

Include metadata files in other formats

EML specific recommendations

Metadata Geography Table

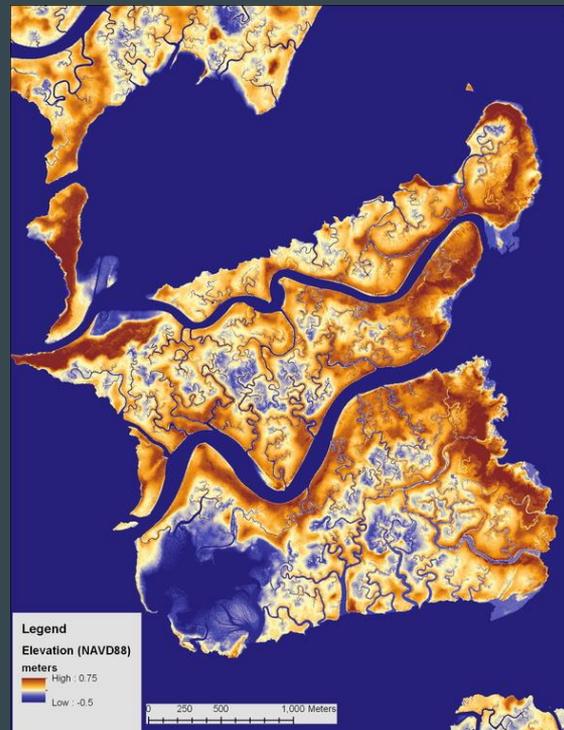
Road GIS Layer for the Eastern Shore of Virginia 2013

Type Shapefile

Tags transportation, Roads, Street Centerline

Summary

The TIGER/Line shapefiles and related database files (.dbf) are an extract of selected geographic and cartographic information from the U.S. Census Bureau's Master Address File / Topologically Integrated Geographic Encoding and Referencing (MAF/TIGER) Database (MTDB). The MTDB represents a seamless national file with no overlaps or gaps between parts, however, each TIGER/Line shapefile is designed to stand alone as an independent data set, or they can be combined to cover the entire nation. The All Roads Shapefile includes all features within the MTDB Super Class "Road/Path Features" distinguished where the MAF/TIGER Feature Classification Code (MTFCC) for the feature in MTDB that begins with "S". This includes all primary, secondary, local neighborhood, and rural roads, city streets, vehicular trails (4wd), ramps,



Data from Small, Moving Platforms

Sensors mounted on uncrewed aerial or underwater systems

Data Inventory Table

Publish raw data, processing code, derived data



What's new

Data inventory table

with additional metadata and accession numbers/GUIDs

Publish raw data and processing code

File with metadata in other formats

as part of the data package

