Metadata

Overview:

- What metadata are
- Metadata structure
- Value of metadata
- Tips for creating quality metadata

What do temps represent?

Average Temperature of Observation for Each Species

What units?

Species	Average Temperature	Temperature Standard Deviation	Number of Observations	Minimum Temperature	Maximum Temperature
Northern Red-legged Frog	4.4	---	1	4.4	4.4
Tailed Frog	7.0	3.0	3	4	10
Arizona Toad	10.0	---	1	10	10
Strecker's Chorus Frog	10.5	2.0	11	9	16
Oregon Spotted Frog	11.0	15.5	2	0	22
New Jersey Chorus Frog	11.5	4.5	17	3	22
Wood Frog	12.5	5.5	897	0	28.8
Spring Peeper	13.2	5.6	569	-1	32
Red-legged Frog	13.3	5.9	16	4	27

How?

Courtesy: Viv Hutchison

Where?

What are metadata?

Metadata are: Data about Data

WHO created the data?

WHAT is the content of the data?

WHEN were the data created?

WHERE is it geographically?

WHY were the data developed?

EDI Metadata Template (2017)[1]

Data should be in csv text file. If starting with an Excel spreadsheet, please make sure it does not contain any formulas and comments on cells. If you need comments put them in their own column. If data were used in a database and major table linking is necessary to analyze, please de-normalize into a flat file, not just database table exports.

Dataset Title

(be descriptive, more than 5 words):

Short name or nickname you use to refer to this dataset:

Abstract

(include what, why, where, when, and how)

Investigators

(list in order as for a paper with e-mail addresses, organization and preferably ORCID ID, if you don't have one, get it, it's easy and free: http://orcid.org/) add table rows as needed

First Name	Last Name	Organization	e-mail address	ORCID ID (optional)

Other personnel names and roles

(field crew, data entry etc. with e-mail addresses, organization and ORCID ID)

First Name	Last Name	Organization	e-mail address	ORCID ID (optional)	Role in project

Keywords

(list and separate by comma, please check out these resources http://vocab.lternet.edu,) Please determine one or two keywords that best describe your lab, station, and/or project (e.g., Trout Lake Station, NTL LTER, UW Center for Limnology).

---

[1] This document liberally borrows from similar documents at SBC and GCE

Funding of this work:

Add rows to table if several grants were involved, list only the main PI, start with main grant first:

PI First Name	PI Last Name	PI ORCID ID (optional)	Title of Grant	Funding Agency	Funding Identification Number

Timeframe

- Begin date
- End date
- Data collection ongoing/completed

Geographic location

- Verbal description:
- North bounding coordinates (decimals)
- South bounding coordinates (decimals)
- East bounding coordinates (decimals)
- West bounding coordinates (decimals)

Taxonomic species or groups

Methods

(please be specific, include instrument descriptions, or point to a protocol online, if this is a data compilation please specify datasets used, preferably their DOI or URL plus general citation information)

Data Table

- Column name: exactly as it appears in the dataset. Please avoid special characters, dashes and spaces.
- Description: please be specific, it can be lengthy
- Unit: please avoid special characters and describe units in this pattern: e.g. microSiemenPerCentimeter, microgramsPerLiter, absoptionPerMolePerCentimeter
- Code explanation: if you use codes in your column, please explain in this way: e.g. LR=Little Rock Lake, A=Sample suspect, J=Nonstandard routine followed
- Date format: please tell us exactly how the date and time is formatted: e.g. mm/dd/yyyy hh:mm:ss plus the time zone and whether or not daylight savings was observed.
- If a code for 'no data' is used, please specify: e.g. -99999

To make these metadata useful for data discovery and re-use, computers must be able to extract information from them. The metadata must be standardized.

Column name	Description	Unit or code explanation or date format	Empty value code
Species	Common name of frog		
Average Temperature	Average temperature of frog's skin	celsius	-9999.99
Temperature Standard Deviation	Standard deviation of temperature of the frog's skin	celsius	-9999.99

Two Metadata Files: Standardizing Content and Structure

Gaiser, Evelyn
Water Quality Data from Shark River Slough
Data were collected monthly

Water Quality Data from Shark River Slough, Florida Everglades National Park from June 1, 2000 to March 30, 2017
Grab samples were collected once a month
Collected by Evelyn Gaiser

Different content and different structure

Standardizing Content:

Originator: Evelyn Gaiser
Water Quality Data from Shark River Slough in the Everglades National Park
BeginDate: 2000-06-01
EndDate: 2017-03-30
Methods: Data were collected monthly using grab samples

Water Quality Data from Shark River Slough, Everglades National Park from June 1, 2000 to March 30, 2017
Methods: Grab samples were collected once a month
Originator: Evelyn Gaiser

Content is standardized, but structure is not

Standardizing Structure:

```
<title>Water Quality Data from Shark River Slough, Everglades National Park</title>
<originator>
    <firstName>Evelyn</lastName>
    <lastName>Gaiser</lastName>
</originator>
<method>Grab samples of water were collected monthly </method>
<date>
    <begin>2000-06-01</begin>
    <end>2017-03-30</begin>
</date>
```

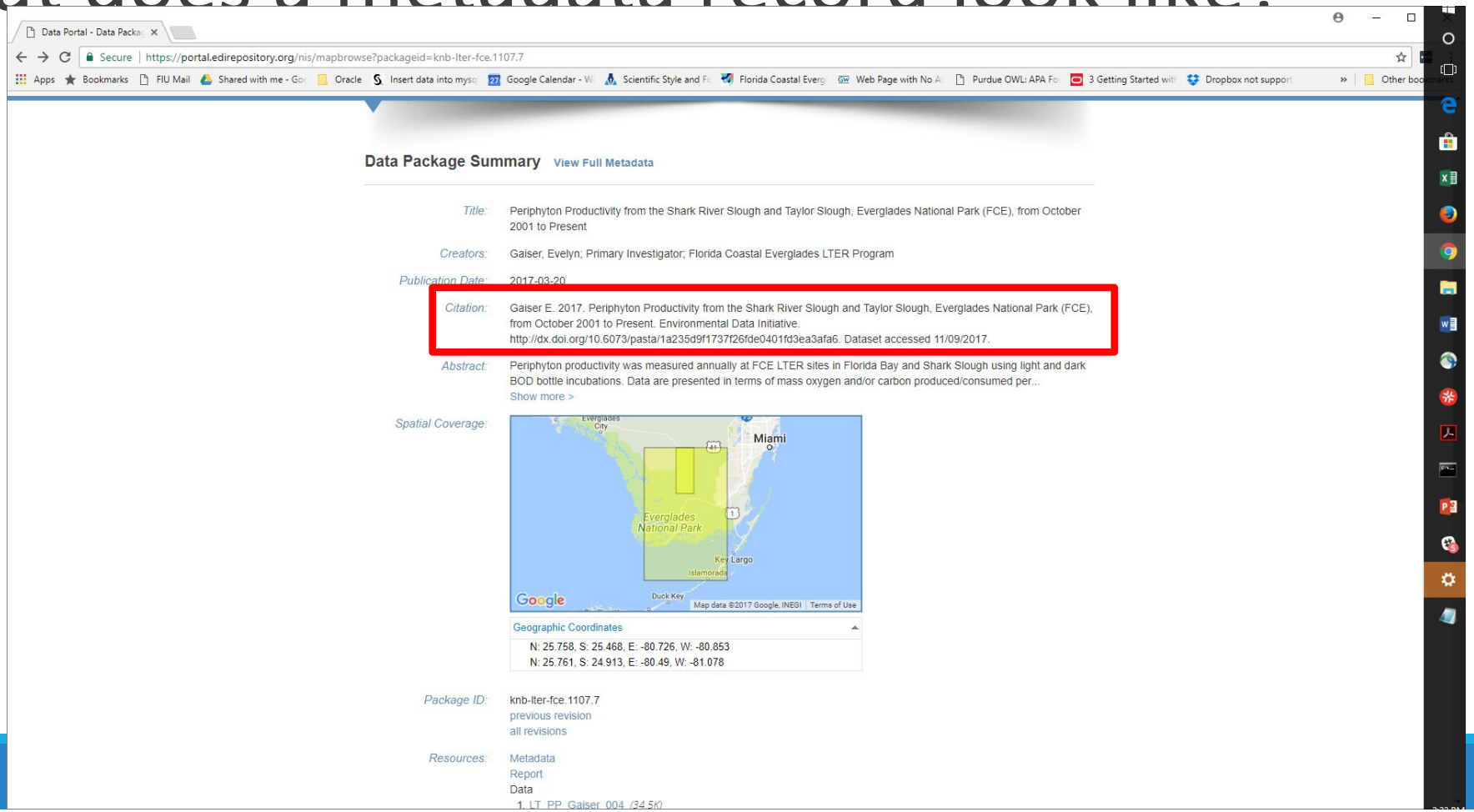
```
<title>Water Quality Data from Shark River Slough, Everglades National Park</title>
<originator>
    <firstName>Evelyn</lastName>
    <lastName>Gaiser</lastName>
</originator>
<method>Grab samples of water were collected monthly </method>
<date>
    <begin>2000-06-01</begin>
    <end>2017-03-30</begin>
</date>
```

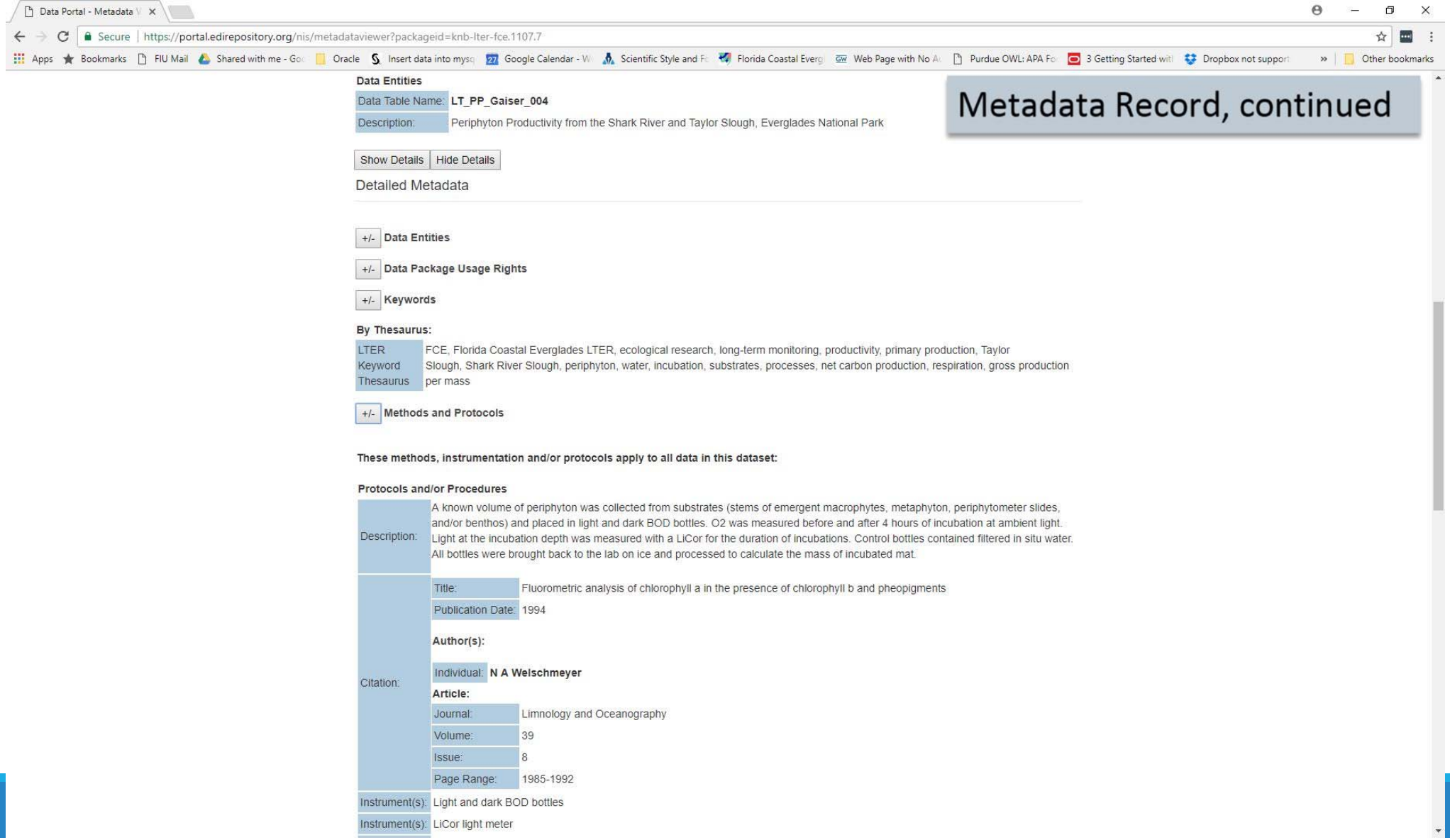
Content and structure are standardized, making it easy for computers to automatically extract information from the metadata

Ecological Metadata Language (EML)

- Developed for documenting ecological and environmental datasets
- Based on previous work done by the Ecological Society of America
- Implemented in XML
  - Tags surrounding content that describe its meaning
  - A markup language that defines a set of rules for encoding documents in a format that is machine-readable

What does a metadata record look like?




Data Portal - Metadata
Secure | https://portal.edirepository.org/nis/metadataviewer?packageid=knb-lter-fce.1107.7
Apps Bookmarks FIU Mail Shared with me - Go Oracle Insert data into mysq Google Calendar - W Scientific Style and Fo Florida Coastal Everg Web Page with No A Purdue OWL: APA Fo 3 Getting Started wit Dropbox not support Other bookmarks
Metadata Record, continued
Data Entities
Data Table Name: LT_PP_Gaiser_004
Description: Periphyton Productivity from the Shark River and Taylor Slough, Everglades National Park
Show Details Hide Details
Detailed Metadata
+/- Data Entities
+/- Data Package Usage Rights
+/- Keywords
By Thesaurus:
LTER Keyword Thesaurus: FCE, Florida Coastal Everglades LTER, ecological research, long-term monitoring, productivity, primary production, Taylor Slough, Shark River Slough, periphyton, water, incubation, substrates, processes, net carbon production, respiration, gross production per mass
+/- Methods and Protocols
These methods, instrumentation and/or protocols apply to all data in this dataset:
Protocols and/or Procedures
Description: A known volume of periphyton was collected from substrates (stems of emergent macrophytes, metaphyton, periphytometer slides, and/or benthos) and placed in light and dark BOD bottles. O2 was measured before and after 4 hours of incubation at ambient light. Light at the incubation depth was measured with a LiCor for the duration of incubations. Control bottles contained filtered in situ water. All bottles were brought back to the lab on ice and processed to calculate the mass of incubated mat.
Citation:
Title: Fluorometric analysis of chlorophyll a in the presence of chlorophyll b and pheopigments
Publication Date: 1994
Author(s):
Individual: N A Welschmeyer
Article:
Journal: Limnology and Oceanography
Volume: 39
Issue: 8
Page Range: 1985-1992
Instrument(s): Light and dark BOD bottles
Instrument(s): LiCor light meter

Metadata: Data set attributes

Table Column Descriptions

	Project	Group	Transect	Site	substrate	ID Number	date	rate	rate	rate
Column Name:	PROJECT	Group	Transect	Site	Substrate	Replicate	Date	Net_Carbon_Production	Net_Respiration	GrossProductionPerMass
Definition:	Name of Organization	Name of sampling group	Location along freshwater-saline gradient	Name of LTER site	Substrate-Periphytometers	Replicate ID Number	Collection date	Periphyton Net carbon production	Periphyton Net carbon respiration	Periphyton Gross Production per Mass
Storage Type:	text	text	text	ordinal	text	ordinal	datetime	data	data	data
Measurement Type:	nominal	nominal	nominal	ordinal	nominal	ordinal	dateTime	ratio	ratio	ratio
Measurement Values Domain:	Definition: Name of Organization	Definition: Name of sampling group	+/- Allowed Values and Definitions Enumerated Domain Code: SRS Definition: Shark River Slough; TS Source:	Definition: Name of LTER site	+/- Allowed Values and Definitions Enumerated Domain Code: P Definition: periphytometer; M Source:	Definition: Replicate ID Number	Format: YYYY-MM-DD Precision: 1	Unit: milligramsPerGramPerHour Precision: 0.000001 Type: real	Unit: milligramsPerGramPerHour Precision: 0.000001 Type: real	Unit: milligramsPerGramPerHour Precision: 0.000001 Type: real
Missing Value Code:								Code: -9999.000000 Expl: Value will never be recorded	Code: -9999.000000 Expl: Value will never be recorded	Code: -9999.000000 Expl: Value will never be recorded
Accuracy Report:										
Accuracy Assessment:										
Coverage:										
Methods:										

Abstract: Periphyton productivity was measured annually at FCE LTER sites in Florida Bay and Shark Slough using light and dark BOD bottle incubations. Data are presented in terms of mass oxygen and/or carbon produced/consumed per...
Show more >

Spatial Coverage:

Geographic Coordinates

N: 25.758, S: 25.468, E: -80.726, W: -80.853
N: 25.761, S: 24.913, E: -80.49, W: -81.078

Package ID: knb-lter-fce.1107.7
previous revision
all revisions

Resources: Metadata
Report
Data
1. LT_PP_Gaiser_004 *(34.5K)*

Download Zip Archive

Intellectual Rights: These data are classified as 'Type II' whereby original FCE LTER experimental data collected by individual FCE researchers to be released to restricted audiences according to terms specified by the owners of the data. Ty...
Show more >

Digital Object Identifier: doi:10.6073/pasta/1a235d9f1737f26fde0401fd3ea3afa6

PASTA Identifier: https://pasta.lternet.edu/package/eml/knb-lter-fce/1107/7

Provenance: Generate provenance metadata for use within your derived data package

Code Generation: Analyze this data package using Matlab, R, SAS, SPSS

R Code

Package ID: knb-lter-fce.1107.7

File Download: knb-lter-fce.1107.7.r

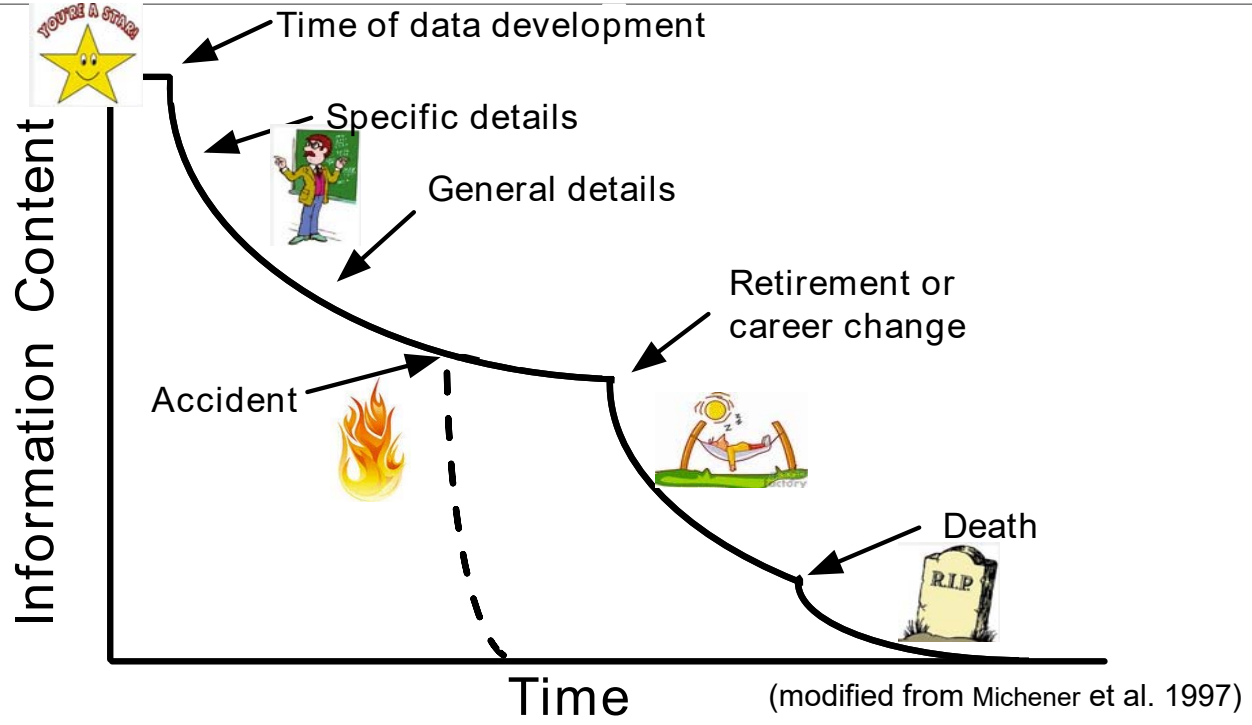
Instructions: Download the R program and open it in R to run. Alternatively, you can copy and paste the program code into the R console.

For datasets that require authenticated access to data tables, you may need to download the data separately and alter the `infile <-` lines to reflect where the data is stored on your computer.

Code:

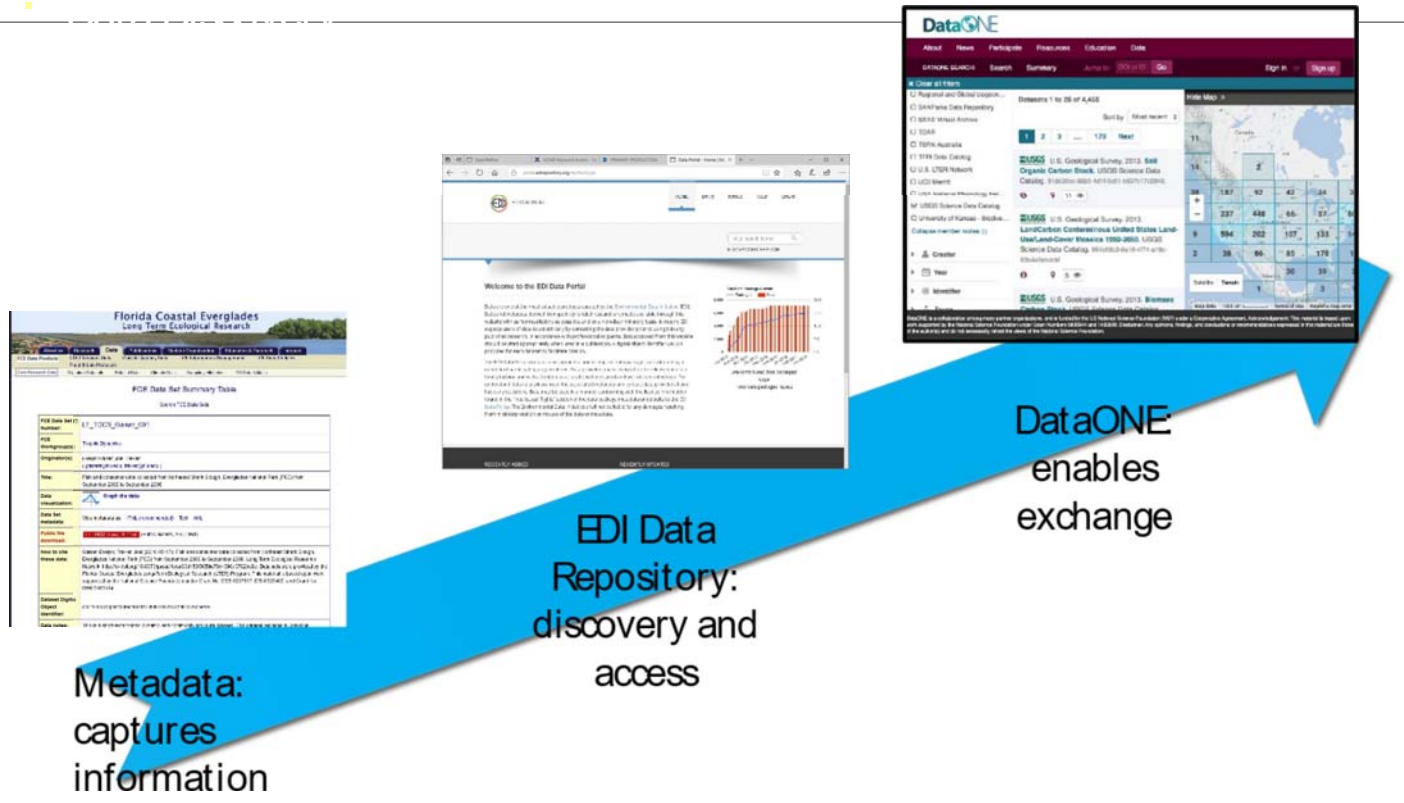
```
# Contact:    - Information Manager Florida Coastal Everglades LTER Program  - fcelter@fiu.edu
# Stylesheet for metadata conversion into program: John H. Porter, Univ. Virginia, jporter@virginia.edu

infile1  <- "https://pasta.lternet.edu/package/data/eml/knb-lter-fce/1107/7/ac394fca14259073329ffe19ecf096f6"
infile1 <- sub("^https","http",infile1)
 dt1 <-read.csv(infile1,header=F
          ,skip=1
            ,sep=","
        , col.names=c(
                    "PROJECT",
                    "Group",
                    "Transect",
                    "Site",
                    "Substrate",
                    "Replicate",
                    "Date",
                    "Net_Carbon_Production",
                    "Net_Respiration",
                    "GrossProductionPerMass"    ), check.names=TRUE)


# Fix any interval or ratio columns mistakenly read in as nominal and nominal columns read as numeric or dates read as strings

if (class(dt1$PROJECT)!="factor") dt1$PROJECT<- as.factor(dt1$PROJECT)
if (class(dt1$Group)!="factor") dt1$Group<- as.factor(dt1$Group)
if (class(dt1$Transect)!="factor") dt1$Transect<- as.factor(dt1$Transect)
```

Metadata: Why are they important?



(modified from Michener et al. 1997)

Structured Metadata: What are they good for?



What is the Value of Metadata to Data Creators and the Organizations they work for ?

Metadata allows data developers to:

- Preserve investment in research project
- Re-use data after initial intended purpose
- Publicize efforts – promote the work of a scientist and his/her contributions to a field of study through citations

For organizations, metadata transcend people and time:

- Offers data permanence
- Creates institutional memory
- Advertises an organization's research
- Creates possible new partnerships and collaborations through data sharing

Tips for Writing Quality Metadata

Titles, Titles, Titles…

Titles are critical in helping readers find your data

- While individuals are searching for the most appropriate datasets, they are most likely going to use the title as the first criteria to determine if a dataset meets their needs.

A complete title includes: What, Where, and When (and Who, if relevant)

Tips for Writing Quality Metadata

A Clear Choice: Which title is better?

Periphyton

OR

Periphyton Abundance data collected by FCE LTER from Northeast Shark River Slough, Everglades National Park, Florida from September 2006 to September 2008

Periphyton abundance data (what) collected by FCE LTER (who) from Northeast Shark River Slough, Everglades National Park, Florida (where) from September 2006 to September 2008 (when)

USGS.gov | Science for a cha | Florida Coastal Everglades L | DataONE Data Catalog

https://search.dataone.org/#data

Search

Most Visited | Getting Started | FIUmail | Google Calendar - We... | Florida Coastal Evergl... | krvander | Environme... | EDIorg | University Libraries | T... | Shared with me - Goo... | Boards | Trello | Thalassa login to EDI | EDI Customer Log | Environment Ontology

DataONE

About | News | Participate | Resources | Education | Data

DATAONE SEARCH: Search | Summary | Jump to: DOI or ID | Go | Sign in or Sign up

Clear all filters

Search

Search phrase

My Search

plant phenology

Filter by:

- Data attribute
- Data files
- Member Node
- Creator
- Year
- Identifier
- Taxon
- Location

1

Heidi G Rodenhizer, Edward A.G. Schuur, Christopher J. Greyson-Gaito, Marguerite Mauritz, and Bonanza Creek LTER. 2017. **Eight Mile Lake Research Watershed, Carbon in Permafrost Experimental Heating Research (CiPEHR): Phenology of Dominant Plant Species III - Flowering Date 2013-2016..** LTER Network Member Node. https://pasta.lternet.edu/package/metadata/eml/knb-lter-bnz/581/14.

1

Heidi G Rodenhizer, Edward A.G. Schuur, Christopher J. Greyson-Gaito, Marguerite Mauritz, and Bonanza Creek LTER. 2017. **Eight Mile Lake Research Watershed, Carbon in Permafrost Experimental Heating Research (CiPEHR): Phenology of Dominant Plant Species II - Berry Production 2013-2016..** LTER Network Member Node. https://pasta.lternet.edu/package/metadata/eml/knb-lter-bnz/580/13.

1

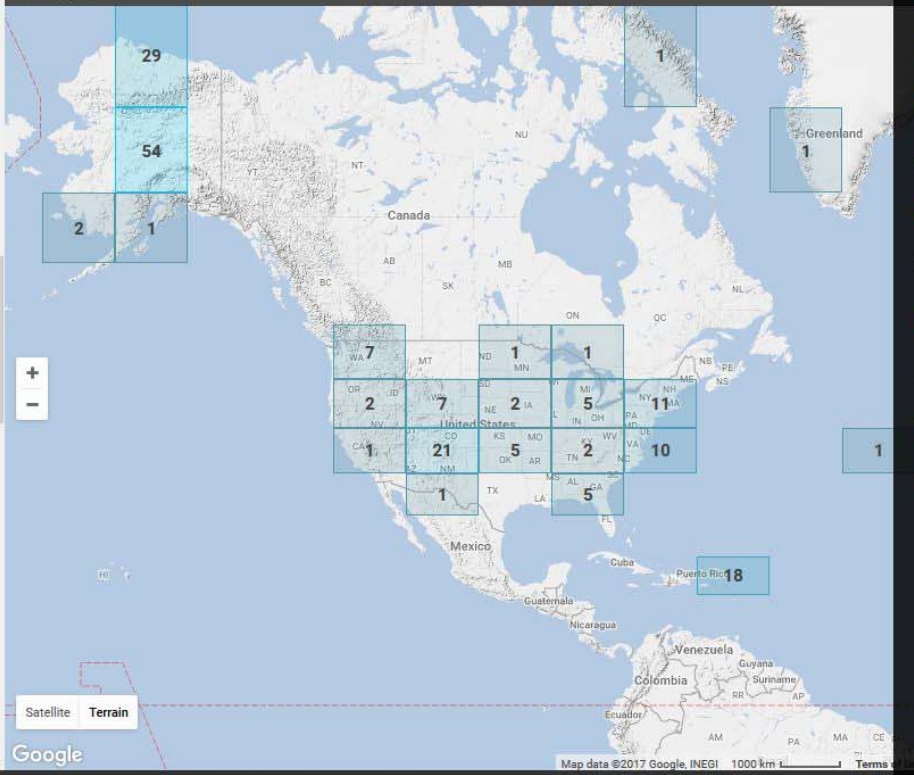
Heidi G R…izer, Edward A.G. Schuur, Christopher J. Greyson-Gaito, Marguerite Mauritz, and Bonanza Creek LTER. 20… **…ight Mile Lake Research Watershed, Carbon in Permafrost Experimental Heating Re…ch (CiPEHR…enology of Dominant Plant Species I - Bud burst and Senescence 2013-2016.** LTER N… Member Node. http…sta.lternet.edu/package/metadata/eml/knb-lter-bnz/570/11.

1

Krab, Eveline J., Rönnefarth, Jonas, Becher, Marina, Blume-Werry, Gesche, Keuper, Frida, et al. 2017. **Plant_responses.** Dryad Digital Repository. http://dx.doi.org/10.5061/dryad.90d2g/2?ver=2017-09-25T16:04:02.047-04:00.

Krab, Eveline J., Rönnefarth, Jonas, Becher, Marina, Blume-Werry, Gesche, Keuper, Frida, et al. 2017. **Abiotics.** Dryad Digital Repository. http://dx.doi.org/10.5061/dryad.90d2g/1?ver=2017-09-25T16:03:59.460-04:00.

Hide Map



DataONE is a collaboration among many partner organizations, and is funded by the US National Science Foundation (NSF) under a Cooperative Agreement. Acknowledgement: This material is based upon work supported by the National Science Foundation under Grant Numbers 0830944 and 1430508. Disclaimer: Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation. (MetacatUI v1.14.11)

3:56
11/9/

Tips for Writing Quality Metadata

Select keywords wisely

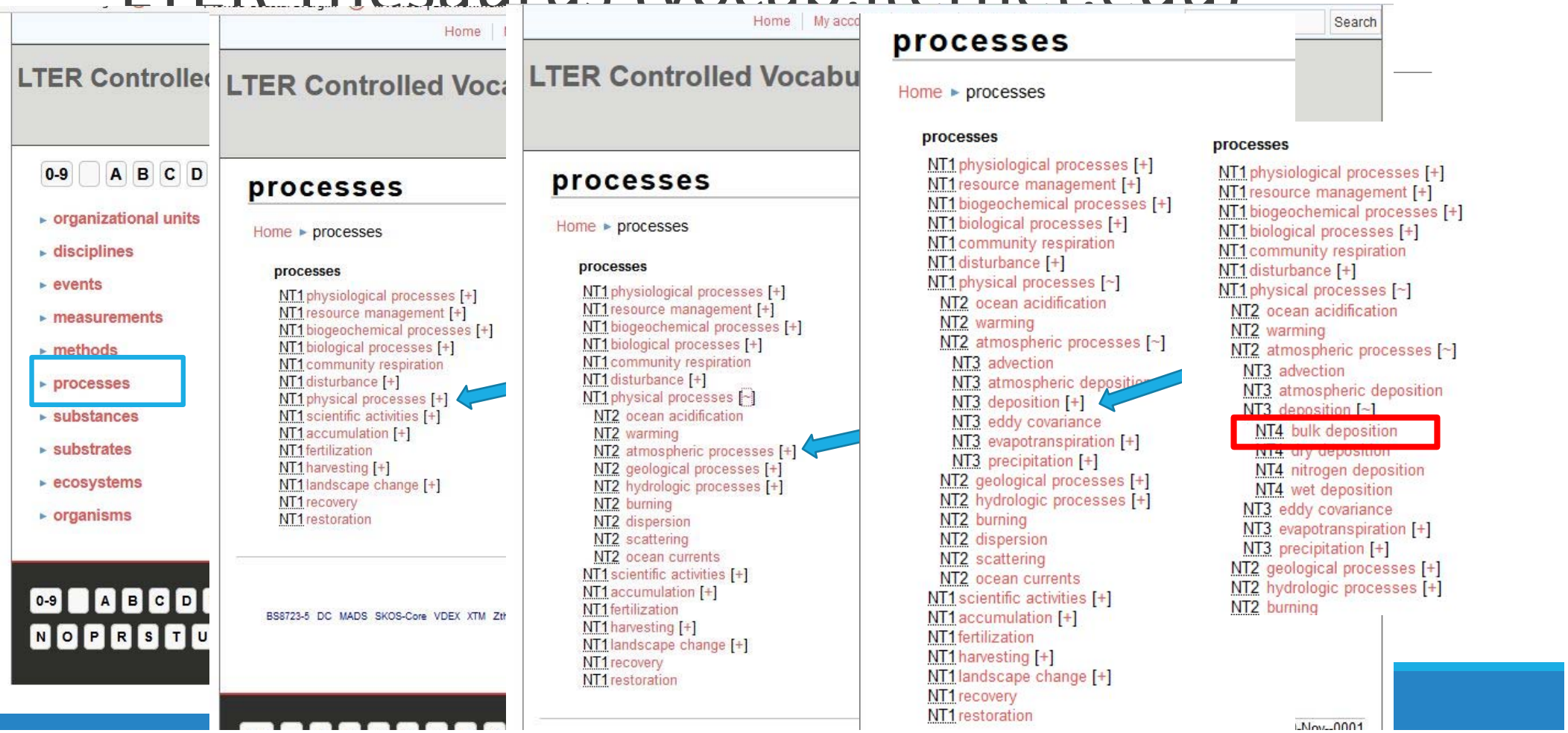
Use a thesaurus or a controlled vocabulary for keywords whenever possible

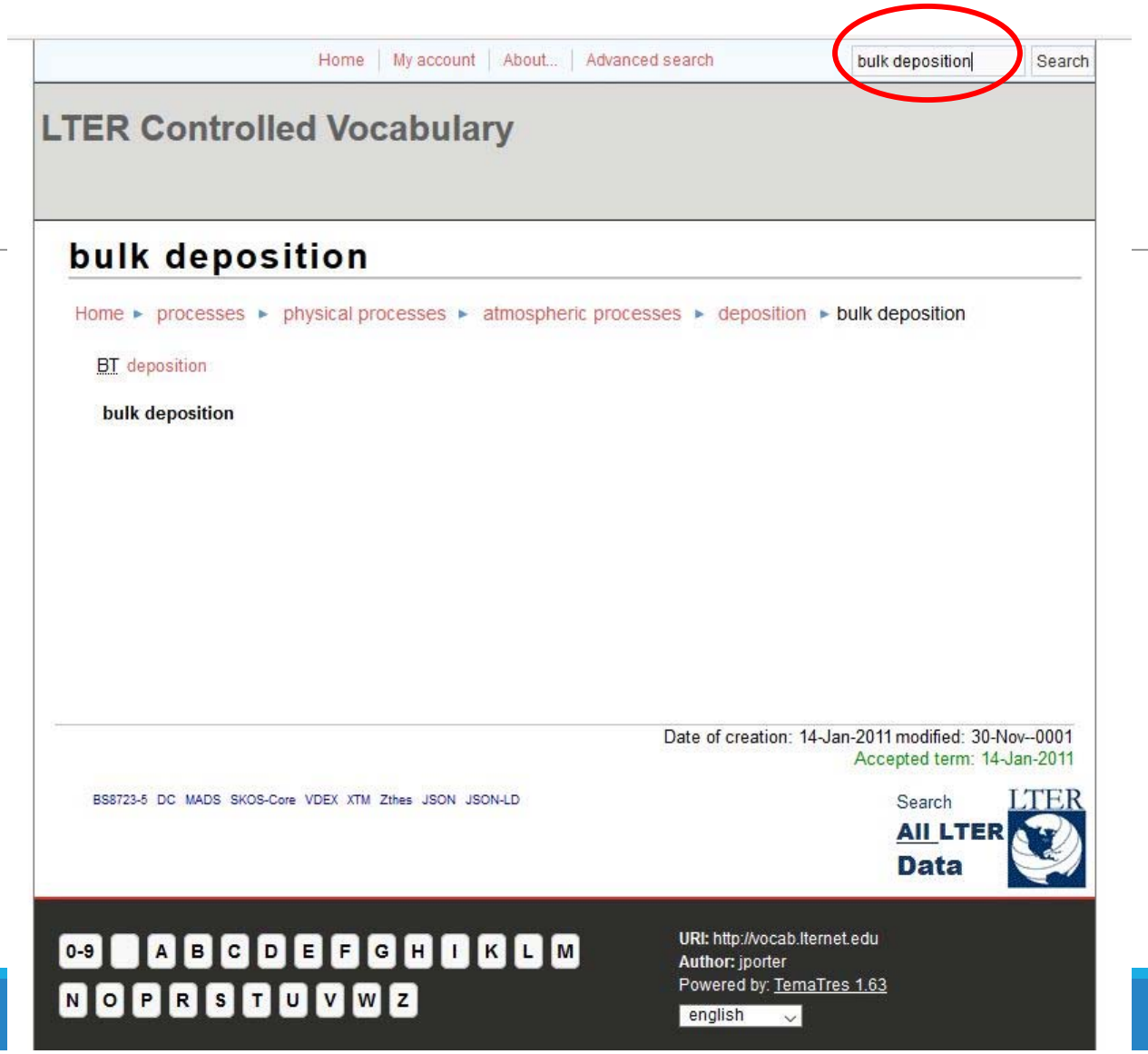
CC image by Marco Arment on Flickr

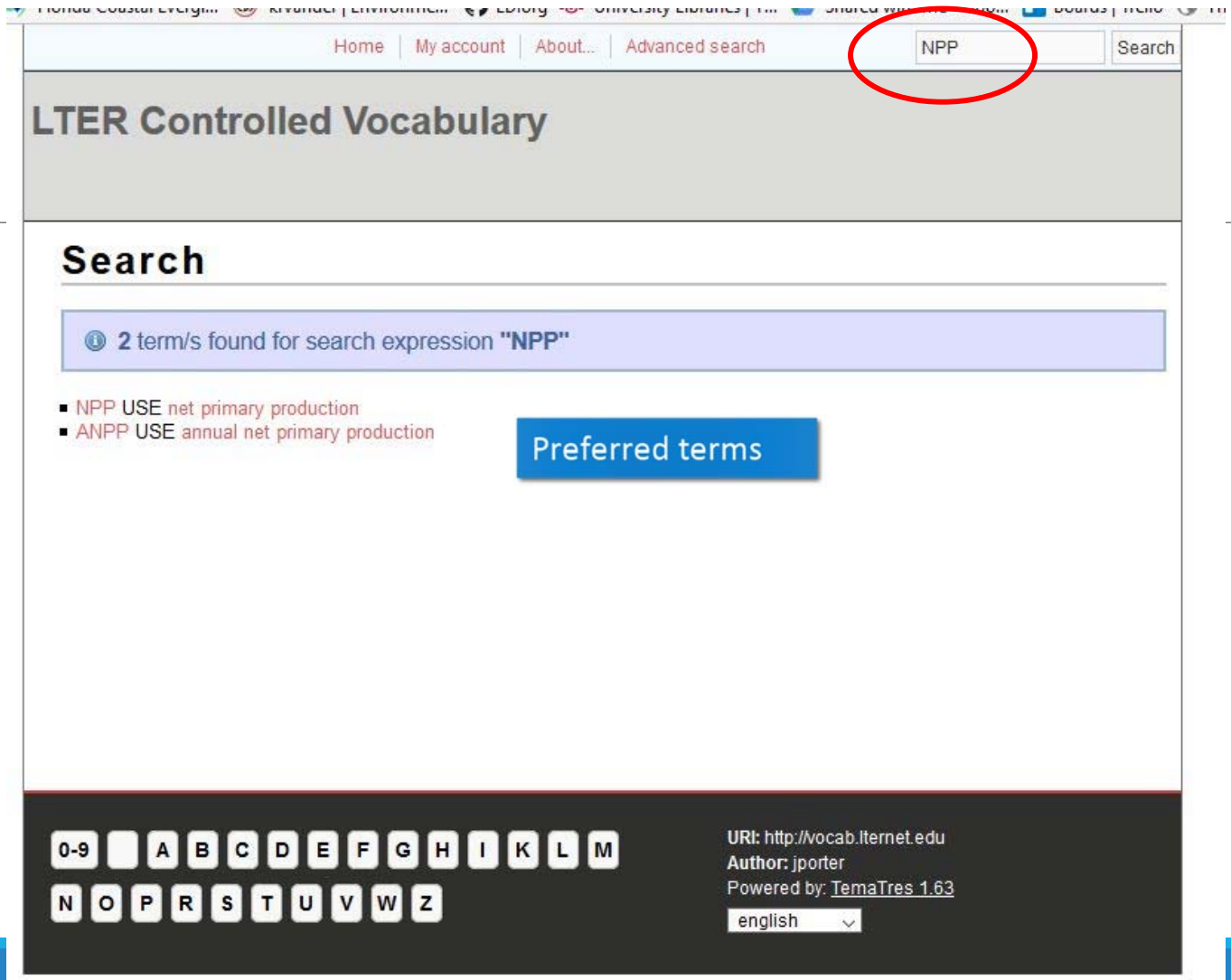
A thesaurus:

- Restricted list of words or terms
- Has a hierarchical structure
- Has broader than, narrower then, and related to terms
- Has preferred terms
- Is used to organize, categorize and index information for subsequent retrieval.

LTER Thesaurus (vocab.lternet.edu)




Home
My account
About...
Advanced search
bulk deposition
Search
LTER Controlled Vocabulary
bulk deposition
Home ▸ processes ▸ physical processes ▸ atmospheric processes ▸ deposition ▸ bulk deposition
BT deposition
bulk deposition
Date of creation: 14-Jan-2011 modified: 30-Nov--0001
Accepted term: 14-Jan-2011
BS8723-5 DC MADS SKOS-Core VDEX XTM Zthes JSON JSON-LD
Search
All LTER
Data
LTER
0-9 A B C D E F G H I K L M
N O P R S T U V W Z
URI: http://vocab.lternet.edu
Author: jporter
Powered by: TemaTres 1.63
english


Home
My account
About...
Advanced search
NPP
Search
LTER Controlled Vocabulary
Search
2 term/s found for search expression "NPP"
NPP USE net primary production
ANPP USE annual net primary production
Preferred terms
0-9
A B C D E F G H I K L M
N O P R S T U V W Z
URI: http://vocab.lternet.edu
Author: jporter
Powered by: TemaTres 1.63
english

Gazatteers

Geographic Names Information System

Getty Thesaurus of Geographic Names

USGS
Geographic Names Information System (GNIS)

Stop! Do not bookmark or copy/paste this URL before reading FAQs.

Query | Result | FAQs

Feature Query Results

Click the feature name for details and to access map services
Click any column name to sort the list ascending ▲ or descending ▼

Feature Name	ID	Class	County	State	Latitude	Longitude	Ele(ft)*	Map**	BGN Date	Entry Date
Port Everglades Junction	295555	Populated Place	Broward	FL	260449N	0801007W	10	Fort Lauderdale South	-	19-OCT-1979
Everglades Headwaters National Wildlife Refuge and Conservation Area	2765286	Park	Okeechobee	FL	273800N	0805748W	69	Fort Drum NW	-	08-MAR-2015
Everglades National Park	293666	Park	Monroe	FL	252220N	0805255W	0	Whitewater Bay East	-	28-AUG-1987
Everglades Wildlife Management Area Conservation Area 3A	307480	Park	Broward	FL	260410N	0804056W	7	Melaleuca Head	-	19-OCT-1979
Everglades Wildlife Management Area Conservation Area 3B	307481	Park	Miami-Dade	FL	254743N	0803841W	7	Long Island	-	19-OCT-1979
Everglades Wildlife Management Area Conservation Area Number 2A	307479	Park	Broward	FL	261416N	0802025W	10	Cooper City NE	-	19-OCT-1979
Everglades Wildlife Management Area Conservation Area Number 2B	294212	Park	Broward	FL	261118N	0802023W	10	Cooper City NE	-	19-OCT-1979
Everglades Wildlife Management Area Conservation Area Number 3	294213	Park	Miami-Dade	FL	255716N	0802826W	10	Pennsuco	-	19-OCT-1979
Camp Everglades	279898	Locale	Miami-Dade	FL	252306N	0803911W	3	Long Pine Key	-	19-OCT-1979
Camp Everglades	307452	Locale	Hendry	FL	262919N	0810654W	23	Rocky Lake Strand	-	19-OCT-1979
Everglades Archers Range	2078385	Locale	Miami-Dade	FL	253120N	0802812W	13	Goulds	-	08-NOV-2005
Everglades City Sewage Treatment Plant	2448876	Locale	Collier	FL	255110N	0812303W	7	Everglades City	-	21-JUL-2008
Everglades Experiment Station	282294	Locale	Palm Beach	FL	263953N	0803803W	13	Belle Glade	-	19-OCT-1979
Everglades Golf Course	2084622	Locale	Palm Beach	FL	264154N	0800224W	3	Palm Beach	-	25-FEB-2006
Golf Club of the Everglades	2448249	Locale	Martin	FL	270550N	0801124W	13	Gomez	-	21-JUL-2008

row(s) 16 - 30 of 62

Summary

A metadata record captures critical information about the content of a dataset

Metadata allow data to be discovered, accessed, and re-used

Metadata is of critical importance to data developers, data users, and organizations

Metadata completes a dataset