



Annotation with EML 2.2

Margaret O'Brien	<u>0000-0002-1693-8322</u>
Steven Chong	<u>0000-0003-1264-1166</u>
Mark Schildhauer	<u>0000-0003-0632-7576</u>
Matt Jones	<u>0000-0003-0077-4738</u>



Outline



Review annotation in EML 2.2

RDF triples

Ontologies and the vocabulary spectrum

Annotations workflows, experiences so far

Upcoming EDI/ADC workshop

Why Annotate?

Amount of
data in
repositories is
increasing

Finding and
understanding
data are
difficult



Why Annotate?



Synonyms: CO₂ flux == Carbon dioxide flux

Homonyms: Litter <> Litter

Hierarchy: NPP *subClass of* Carbon flux

Why Annotate?



Annotations hold links to external vocabularies

Used to construct an “RDF Triple”

Machine readable

EML Annotation element



```
<annotation>  
  <propertyURI label="property label here">property URI here</propertyURI>  
  <valueURI label="value label here">value URI here</valueURI>  
</annotation>
```

Shared Definitions

Accessed through resolvable http URI's

this attribute ("umolesCO2") is_a " Carbon Dioxide Flux"

this attribute ("CO2_Flux_Rate") is_a " Carbon Dioxide Flux"

```
<annotation>  
  <propertyURI label="type">http://www.w3.org/1999/02/22-rdf-syntax-ns#rdf:type</propertyURI>  
  <valueURI label="Carbon Dioxide Flux">http://purl.dataone.org/odo/ECSO_00000536</valueURI>  
</annotation>
```

Carbon Dioxide Flux: http://purl.dataone.org/odo/ECSO_00000536

Annotations in datasets



- **top-level:**
 - an `<annotation>` element is a child of `<dataset>`
- **entity-level:**
 - an `<annotation>` element is a child of `<dataTable>`
- **attribute:**
 - an `<annotation>` element is a child of an entity's `<attribute>` element
- **eml/annotations:**
 - a container for a group of `<annotation>` elements using references
- **eml/additionalMetadata:**
 - `<annotation>` elements that reference a main-body element by its `id` attribute

Dataset Annotation

```
<keywordSet>
  <keyword>Forest</keyword>
  <keywordThesaurus>LTER Controlled Vocabulary V 1.0</keywordThesaurus>
</keywordSet>
```

EML ^
2.1

EML
2.2

V

```
</taxonomicCoverage>
</coverage>

<!-- dataset level, last element in the Resource Group, after coverage -->
<annotation>
  <propertyURI label="is about">http://purl.obolibrary.org/obo/IAO_0000136</propertyURI>
  <valueURI label="Forest Biome">http://purl.obolibrary.org/obo/ENVO_01000174</valueURI>
</annotation>

<contact>
```

Attribute

```
<attribute>
  <attributeName>NEE</attributeName>
  <attributeDefinition>net ecosystem exchange</attributeDefinition>
  <measurementScale>
    ...
  </measurementScale>
  <missingValueCode>
    ...
  </missingValueCode>
</attribute>
```

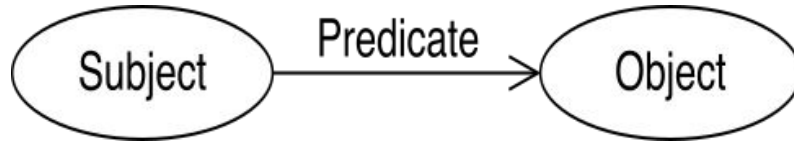
EML ^
2.1

EML
2.2

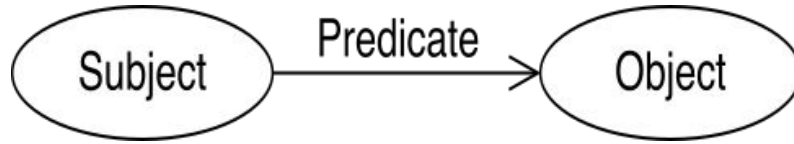
V

```
<attribute id="dsid_01.att_01">
  <attributeName>NEE</attributeName>
  <attributeDefinition>net ecosystem exchange</attributeDefinition>
  <measurementScale>
    ...
  </measurementScale>
  <missingValueCode>
    ...
  </missingValueCode>
  <annotation>
    <propertyURI label="contains measurements of type"
      >http://ecoinformatics.org/oboe/oboe.1.2/oboe-core.owl#containsMeasurementsOfType</propertyURI>
    <valueURI label="Net Ecosystem Exchange Carbon Flux"
      >http://purl.dataone.org/odo/ECSO_00000014</valueURI>
    </annotation>
  </attribute>
```

RDF Triple



RDF Triple

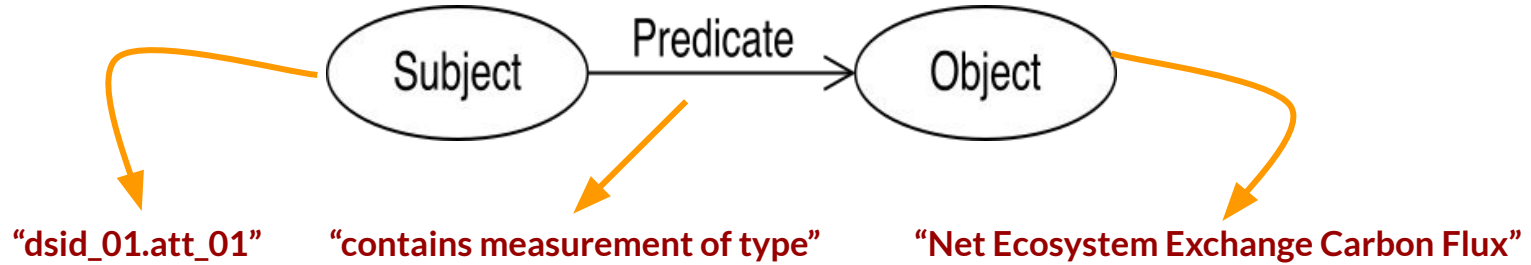


“dsid_01.att_01”

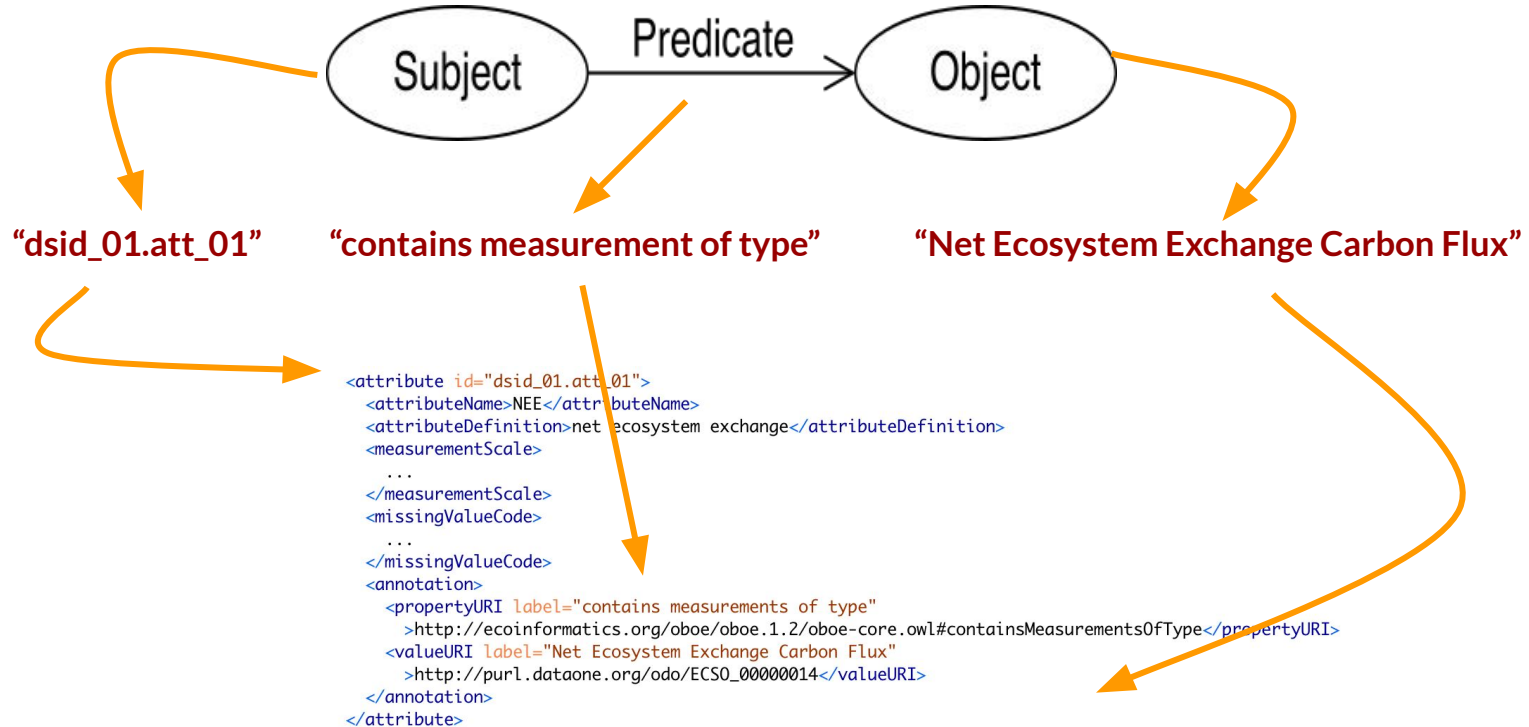
“contains measurement of type”

“Net Ecosystem Exchange Carbon Flux”

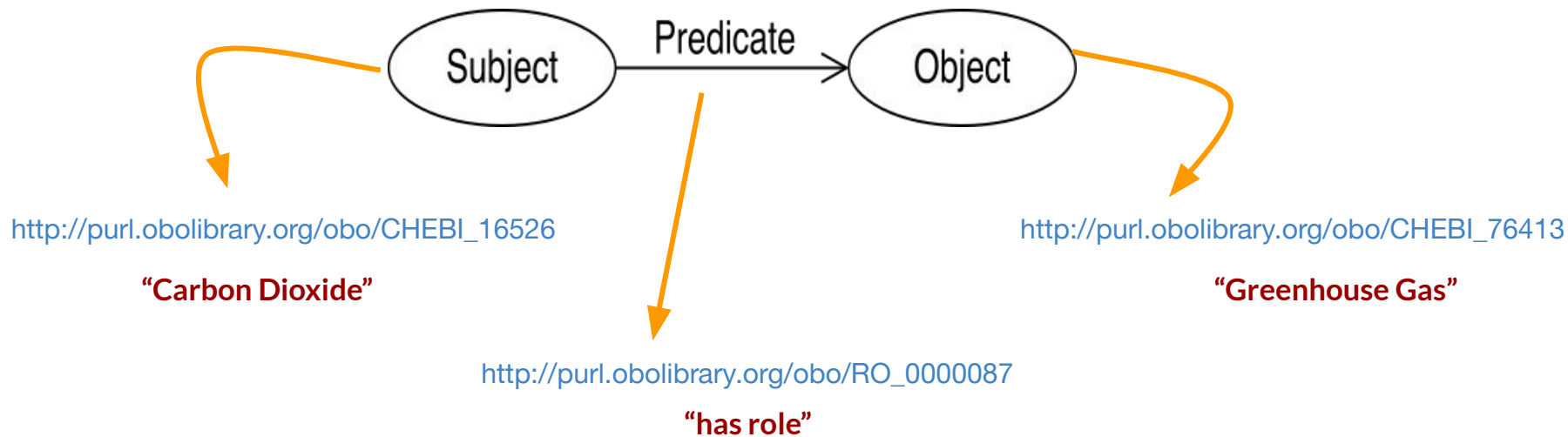
RDF Triple



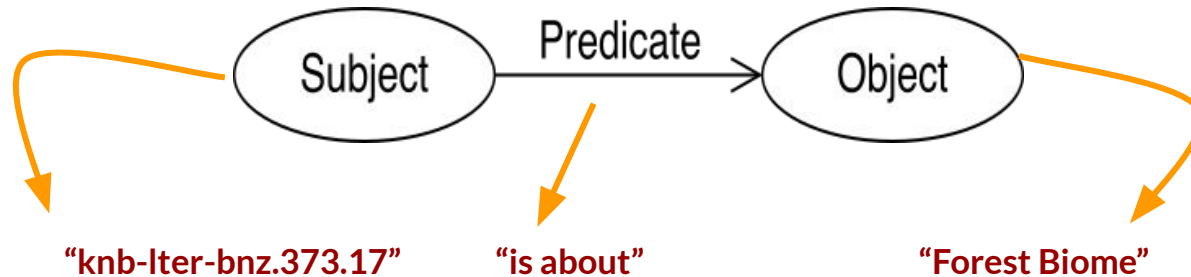
Annotation - attribute



What the computer uses



Annotation - dataset



```
</taxonomicCoverage>  
</coverage>
```

```
<!-- dataset level, last element in the Resource Group, after coverage -->
```

```
<annotation>  
  <propertyURI label="is about">http://purl.obolibrary.org/obo/IAO_0000136</propertyURI>  
  <valueURI label="Forest Biome">http://purl.obolibrary.org/obo/ENVO_01000174</valueURI>  
</annotation>
```

```
<contact>
```


Summary

```
<annotation>  
  <propertyURI label="property label here">property URI here</propertyURI>  
  <valueURI label="value label here">value URI here</valueURI>  
</annotation>
```

Predicate

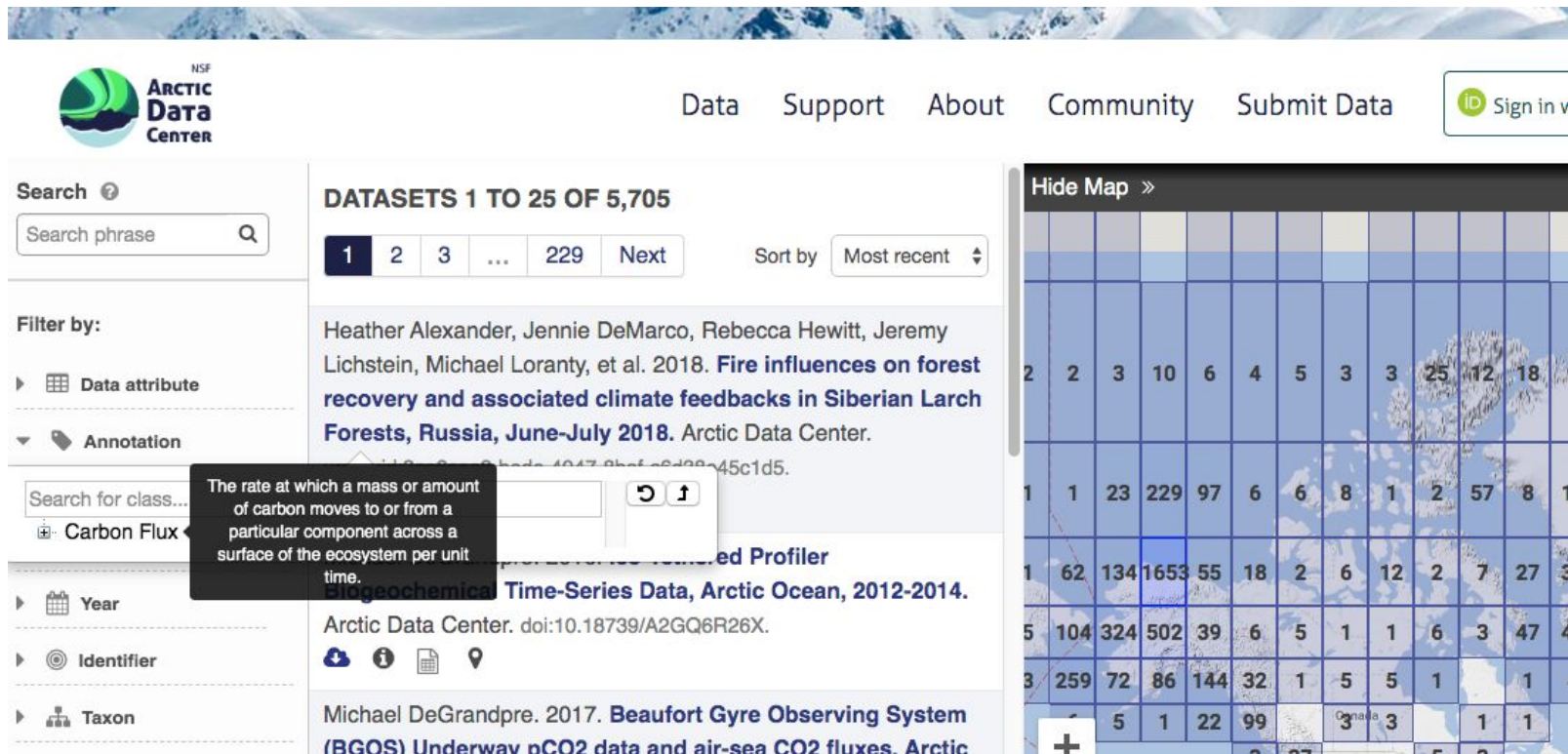
Object

Subject

Predicate


Object


Support at Arctic Data Center





The screenshot displays the Arctic Data Center website interface. At the top, the NSF Arctic Data Center logo is on the left, and navigation links for Data, Support, About, Community, and Submit Data are in the center. A sign-in button is on the right. The main content area shows search results for "DATASETS 1 TO 25 OF 5,705". A search bar on the left contains the text "Search phrase". Below it, a "Filter by:" section includes "Data attribute" and "Annotation". A "Search for class..." dropdown is set to "Carbon Flux". A tooltip explains: "The rate at which a mass or amount of carbon moves to or from a particular component across a surface of the ecosystem per unit time." The search results list two datasets: 1. "Fire influences on forest recovery and associated climate feedbacks in Siberian Larch Forests, Russia, June-July 2018." by Heather Alexander et al. 2. "Beaufort Gyre Observing System (BGOS) Underway pCO2 data and air-sea CO2 fluxes." by Michael DeGrandpre. On the right, a "Hide Map" button is above a map of the Arctic region with a grid overlay showing numerical data values.

NSF
ARCTIC
Data
Center

Data Support About Community Submit Data  Sign in w

Search ?
Search phrase 

Filter by:
▶  Data attribute
▼  Annotation
Search for class...
Carbon Flux

DATASETS 1 TO 25 OF 5,705

1 2 3 ... 229 Next Sort by Most recent ▾

Heather Alexander, Jennie DeMarco, Rebecca Hewitt, Jeremy Lichstein, Michael Loranty, et al. 2018. **Fire influences on forest recovery and associated climate feedbacks in Siberian Larch Forests, Russia, June-July 2018.** Arctic Data Center.
doi:10.18739/A2GQ6R26X.

The rate at which a mass or amount of carbon moves to or from a particular component across a surface of the ecosystem per unit time.

Biogeochemical Time-Series Data, Arctic Ocean, 2012-2014.
Arctic Data Center. doi:10.18739/A2GQ6R26X.

Michael DeGrandpre. 2017. **Beaufort Gyre Observing System (BGOS) Underway pCO2 data and air-sea CO2 fluxes.** Arctic Data Center. doi:10.18739/A2GQ6R26X.

Hide Map »

2 2 3 10 6 4 5 3 3 25 12 18
1 1 23 229 97 6 6 8 1 2 57 8 1
1 62 134 1653 55 18 2 6 12 2 7 27
5 104 324 502 39 6 5 1 1 6 3 47
3 259 72 86 144 32 1 5 5 1 1
5 1 22 99 3 3 1 1

Support at DataONE

Filter
by
annotation



More at: [January 2020 DataONE Webinar \(Schildhauer/O'Brien\)](#)

Navigation: About News Participate Resources Education Data

DATAONE SEARCH: Search Summary Jump to: DOI or ID Go

Clear all filters

Search

Search phrase

My Search

- Annotation: Measurement Type
- Data source: Environmental Data Initiative
- Data source: LTER Network Member Node

Filter by:

- Data attribute
- Annotation
- Data files
- Member Node
 - ☐ Arctic Data Center
 - ☐ ARM - Atmospheric R...
 - ☐ Biological and Chemic...
 - ☐ Cary Institute of Ecos...

Datasets 1 to 5 of 5

Sort by: Most recent

1. Santa Barbara Coastal LTER, Daniel C Reed, and Shannon Harrer. 2020. **SBC LTER: Kelp Removal Experiment: Hourly photon irradiance at the surface and seafloor.** LTER Network Member Node. <https://pasta.lternet.edu/package/metadata/eml/knb-lter-sbc/36/15>.

2. Santa Barbara Coastal LTER, J Clint Nelson, Daniel C Reed, Shannon Harrer, and Robert J Miller. 2020. **SBC LTER: Reef: Coefficients for estimating biomass from body size or percent cover for kelp forest species.** LTER Network Member Node. <https://pasta.lternet.edu/package/metadata/eml/knb-lter-sbc/127/2>.

3. Beaufort Lagoon Ecosystems LTER, Core Program. 2019. **Dissolved organic carbon (DOC) and total dissolved nitrogen (TDN) from river, lagoon, and open ocean sites along the Alaska Beaufort Sea coast, 2018-ongoing.** LTER Network Member Node. <https://pasta.lternet.edu/package/metadata/eml/knb-lter-ble/2/3>.

4. Beaufort Lagoon Ecosystems LTER and Michael Rawlins. 2019. **Model simulated hydrological estimates for the North Slope drainage basin, Alaska, 1980-2010.** LTER Network Member Node. <https://pasta.lternet.edu/package/metadata/eml/knb-lter-ble/5/1>.

5. Moorea Coral Reef LTER, Deron Burkepile, and Tom Adam. 2019. **MCR LTER: Coral Reef: Coral bleaching with nitrogen and heat stress: 2016 data in support of Donovan et al.**

Map: Hide Map >

Reusable Measurements



Patrick Sullivan. 2016. Light Saturated Needle Photosynthesis and Stomatal Conductance. Arctic Data Center. doi:10.18739/A28911R0D.



http://purl.dataone.org/odo/EC SO_00003084

Description Measurements of instantaneous mid-day light saturated net photosynthesis and stomatal conductance of white spruce needles at sites on the Agashashok, Kugururok, Dietrich and Wind Rivers in the Brooks Range, Alaska.

Attribute Information

VARIABLES

- site
- date
- tree
- photo_umol_m-2_s-1
- cond_mmol_m-2_s-1

Name cond_mmol_m-2_s-1

Annotations contains measurements of type stomatal conductance

Label cond_mmol_m-2_s-1

Definition Stomatal conductance

Storage Type integer

Measurement Type nominal

Measurement Domain Definition unit: mmol H₂O m⁻² s⁻¹

Missing Value Code

Code	Explanation
NA	Missing Value

Annotation Support at EDI

Phase 1:

HTML - display and link at 2 levels

- Dataset
- Attribute

☐ +/- Data Package Usage Rights

☒ +/- External Annotations

With link(s) out to external vocabularies

Dataset is about [marine biome](#)

Dataset is about **kelp forest**

Dataset is about **algal material**

Dataset is about **photosynthesis**

☐ +/- Keywords

Phase 2:

TBD

[Dataset-level: knb-Iter-sbc.57.9](#)

[Attribute-level: knb-Iter-ble.5.2](#)

Table Column Descriptions

GHAAS river basin ID		Date	River discharge/export	River cumulative discharge/export
Column Name:	ghaas_basin_id	date	export	export_cumulative
Definition:	Basin ID in the Global Hydrologic Archive and Analysis System (GHAAS)	Date YYYY-MM-DD	Daily model estimated river discharge/export	Estimated cumulative river discharge/export from start of temporal domain (1980-01-01) to corresponding date
External Measurement Definition, Link:			containsMeasurementsOfType water discharge	containsMeasurementsOfType water discharge
Storage Type:	integer	date	float	float
Measurement Type:	nominal	dateTime	ratio	ratio
Measurement Values Domain:	Definition any text	Format YYYY-MM-DD	Unit kilometerCubed Type real	Unit kilometerCubed Type real
Missing Value		Precision 1		

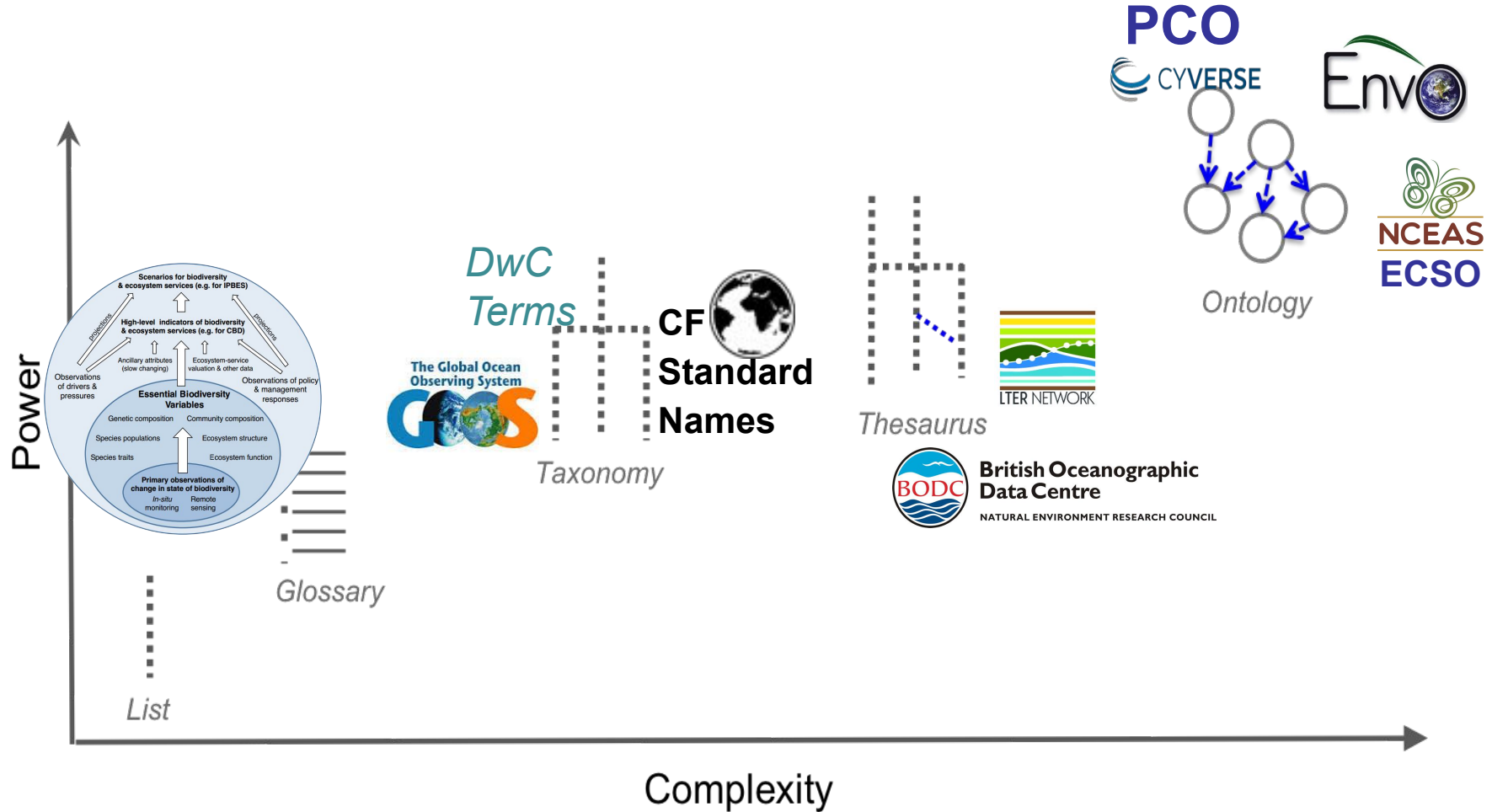
URIs in a Darwin Core Archive



Measurement URIs can populate the measurementID field, here using a defined measurement for volumetric abundance in a **Darwin Core Archive for GBIF**

eventID	date	Taxon	taxonID	measurementOrFact	measurementID	measurementValue	measurementUnit
80116_1_5m	16-Jan-2008	Pseudo-nitzschia	ITIS:584561	cells_liter	http://purl.dataone.org/odo/ECSO_00001201	5960	numberPerLiter
80116_1_5m	16-Jan-2008	Leptocylindrus	ITIS:2394	cells_liter	http://purl.dataone.org/odo/ECSO_00001201	3040	numberPerLiter
80116_1_5m	16-Jan-2008	Thalassiosira	ITIS:2484	cells_liter	http://purl.dataone.org/odo/ECSO_00001201	2680	numberPerLiter
80116_1_5m	16-Jan-2008	Chaetoceros	ITIS:2758	cells_liter	http://purl.dataone.org/odo/ECSO_00001201	4740	numberPerLiter

Vocabulary Spectrum



SKOS and Ontology Compared



Semi-formal KOS	Ontology
<p>Semi-formal, for thesauri and classification schemes</p> <p>may also be arranged in structures (e.g. hierarchies); relationships are built-in</p> <p>structures have no formal semantics</p>	<p>Represents a logical view of a domain</p> <p>Asserts axioms or facts; designers must create these</p> <p>Automated reasoning</p> <p>Import other ontologies</p>

ENVO - Environment Ontology

Hierarchies for
Biomes
Environmental Features
Environmental Materials

International

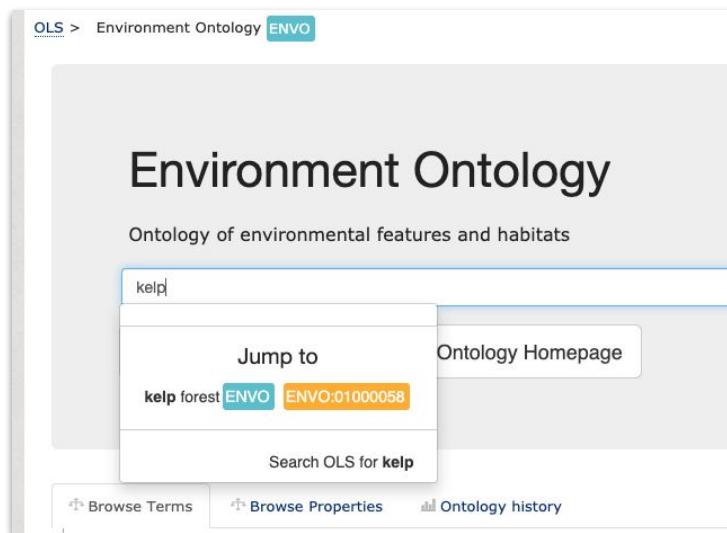
Cross-references LTER vocabulary terms



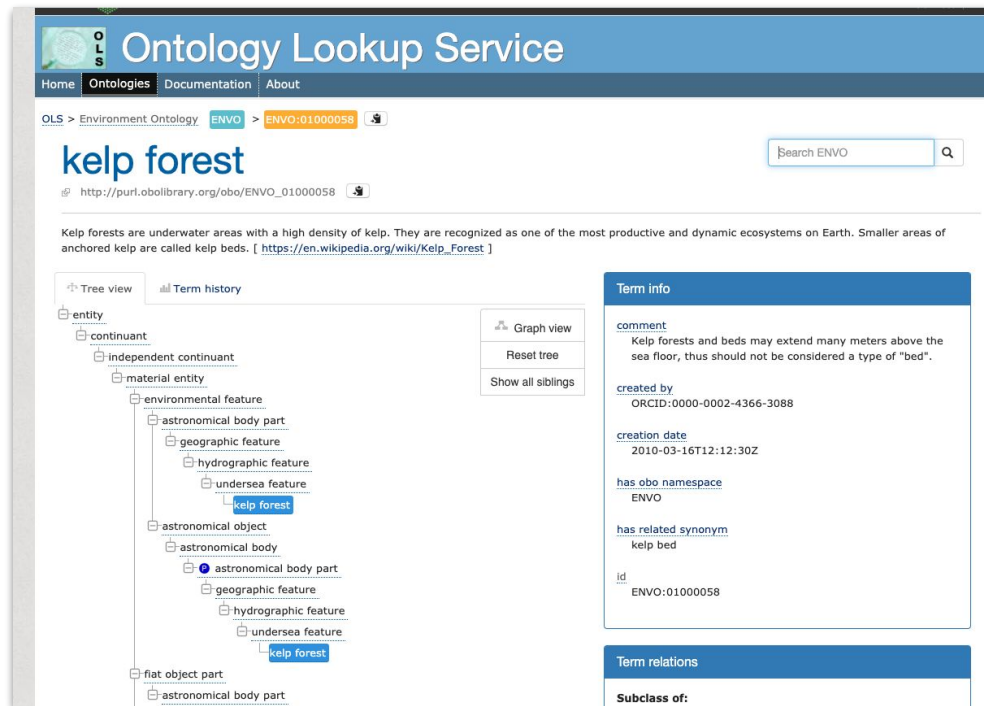
<http://environmentontology.org/>



ENVO - Environment Ontology



<http://www.environmentontology.org/Browse-EnvO>



[Concept "Kelp Forest" \(ENVO:01000058\) in Ontology Lookup Service](#)

ENVO - Environment Ontology

Ontology Lookup Service

Home | **Ontologies** | Documentation | About

OLS > Environment Ontology | ENVO > **ENVO:01000174**

forest biome

http://purl.obolibrary.org/obo/ENVO_01000174

A forest biome is a terrestrial biome which includes, across its entire spatial extent, densely packed vegetation which strongly limits light penetration to the forest floor.

Tree view | **Term history**

- entity
 - continuant
 - independent continuant
 - material entity
 - environmental feature
 - astronomical body part
 - biosphere
 - biome
 - terrestrial biome
 - forest biome**

Graph view
Reset tree
Show all siblings

Term info

database cross reference

- SPIRE:Forest
- LTER:212

comment
Preliminary definition.

created by
ORCID:0000-0002-4366-3088

creation date
2013-04-07T14:26:46Z

Term relations

Subclass of:

- [terrestrial biome](#)
- [has part some forest ecosystem](#)

Cross-reference to term
“Forest” in the LTER
Controlled Vocabulary

[Concept "Forest Biome" \(ENVO:01000174\) in Ontology Lookup Service](#)

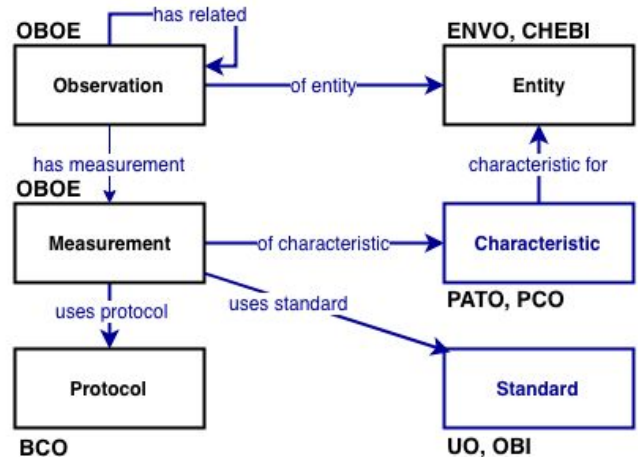
ECSO - EcoSystem Ontology

Defines “Measurement Types”

Imports existing ontologies into an observation framework, “OBOÉ ”

Initial Focus on Carbon Cycling

<http://bioportal.bioontology.org/ontologies/ECSO>



Interoperable Measurements

http://purl.dataone.org/odo/ECSO_00003084

Above HTTP URI dereferences
to a term in the ECSO ontology

“stomatal conductance”
measurements are a type of
“carbon flux” in the ECSO ontology

- Calcium Carbonate
- Carbon Cycle Component
- Chlorophyll
- entity
- Inorganic Carbon
- KilogramPerMeterSquaredPerSecond
- Leaf_Area_Index
- Measurement Type
 - 00 Unclassified Measurement Type
 - absorbance measurement type
 - age_MeasurementType
 - amount_MeasurementType
 - angle measurement type
 - Area Measurement Type
 - Areal Density Measurement Type
 - attenuation measurement type
 - categorical data measurement type
 - conductivity
 - Count Measurement Type
 - data quality assessment measurement type
 - direction measurement type
 - error MeasurementType
 - Fluorescence_measurementType
 - Flux Measurement Type
 - Carbon Flux
 - Carbon Dioxide Flux
 - Carbon Dioxide Diffusion Flux
 - Oceanic Carbon Dioxide Flux
 - soil carbon dioxide flux
 - stomatal conductance**
 - carbon monoxide flux
 - Dissolution Carbon Flux
 - Dissolved Organic Carbon Flux

| | |
|----------------|---|
| Preferred Name | stomatal conductance |
| Definitions | The rate of carbon dioxide entering or water vapor exiting through the stomata of a leaf. |
| ID | http://purl.dataone.org/odo/ECSO_00003084 |
| creator | http://orcid.org/0000-0003-1264-1166 |
| date | 2019-07-19T00:06:38Z |
| definition | The rate of carbon dioxide entering or water vapor exiting through the stomata of a leaf. |
| label | stomatal conductance |
| prefixIRI | odo:ECSO_00003084 |
| prefLabel | stomatal conductance |
| subClassOf | Carbon Dioxide Flux |

http://purl.dataone.org/odo/ECSO_00000014

Jump to:

- Carbon Cycle Component
- Characteristic Qualifier
- Collection
- Concept
- Concept
- Concept Scheme
- Continuant
- Entity
- frozen
- Measurement
- Measurement Type
 - 00 Unclassified Measurement Type
 - Area Measurement Type
 - Area Density Measurement Type
 - Biomass Measurement Type
 - Concentration Measurement Type
 - Count Measurement Type
 - Fluorescence_measurementType
 - Flux Measurement Type
 - Carbon Dioxide Flux
 - Carbon Flux
 - Dissolution Carbon Flux
 - Dissolved Organic Carbon Flux
 - Fire Carbon Flux
 - Fire Emissions MOV
 - Net Ecosystem Exchange Carbon Flux**
 - Primary Production Carbon Flux
 - Respiration Carbon Flux
 - Methane Flux
 - Momentum Flux
 - Net Assimilation Rate Flux
 - Net Primary Production Biomass Flux
 - Radiative flux
 - Specific Flux

Details Visualization Notes (0) Class Mappings (0)

| | |
|------------------------|--|
| Preferred Name | Net Ecosystem Exchange Carbon Flux |
| Synonyms | NEE |
| Definitions | The magnitude of carbon sources and sinks is defined as the vertical exchange of CO2 between the surface (land or ocean) and the atmosphere. |
| ID | http://purl.dataone.org/odo/ECSO_00000014 |
| alternative label | NEE |
| definition | The magnitude of carbon sources and sinks is defined as the vertical exchange of CO2 between the surface (land or ocean) and the atmosphere. |
| definition_Contributor | Chase LeCroy, orcid.org/0000-0002-1338-9436 |
| definition_Source | Hayes, D.J. et al. Reconciling estimates of the contemporary North American carbon balance among terrestrial biosphere models, atmospheric inversions, and a new approach for estimating net ecosystem exchange from inventory-based data. Global Change Biology. V 8, i 4. pp. 1282-1299. April 2012. DOI: 10.1111/j.1365-2486.2011.02627.x |
| measuresEntity | carbon atom |
| preferred label | Net Ecosystem Exchange Carbon Flux |
| prefixIRI | odo:ECSO_00000014 |
| rdfs:label | Net Ecosystem Exchange Carbon Flux |
| seeAlso | NEE |
| subClassOf | Carbon Flux |

Attribute, with annotation

```
<attribute id="dsid_01.att_01">
  <attributeName>NEE</attributeName>
  <attributeDefinition>net ecosystem exchange</attributeDefinition>
  <measurementScale>
    <interval>
      <unit>
        <customUnit>micromolsPerSquareMeterPerSecond</customUnit>
      </unit>
      <numericDomain>
        <numberType>real</numberType>
      </numericDomain>
    </interval>
  </measurementScale>
  <missingValueCode>
    <code>-9999</code>
    <codeExplanation>NA</codeExplanation>
  </missingValueCode>
  <annotation>
    <propertyURI label="contains measurements of type">http://ecoinformatics.org/oboe/oboe.1.2/oboe-core.owl#containsMeasurementsOfType</propertyURI>
    <valueURI label="Net Ecosystem Exchange Carbon Flux">http://purl.dataone.org/odo/ECSO_00000014</valueURI>
  </annotation>
</attribute>
```

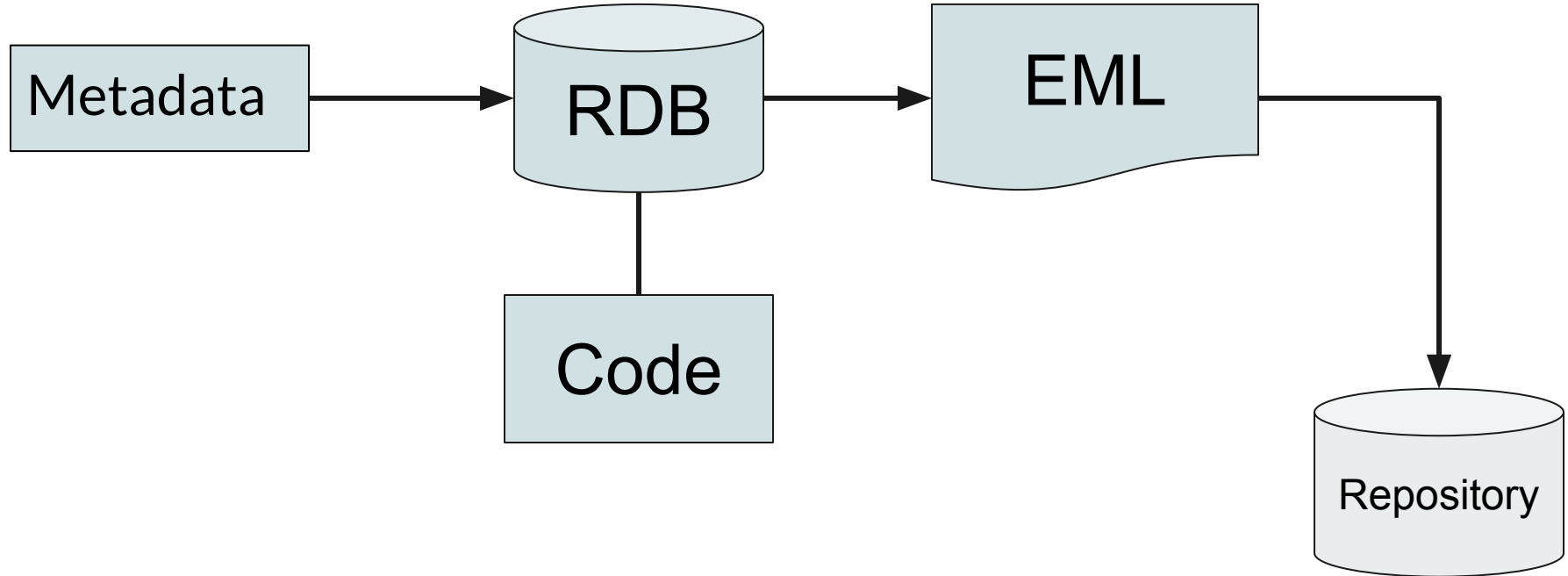


Annotation in a curation workflow



| | |
|-----------------------------|--|
| LTER | Metadata has multiple uses, stored centrally (e.g., RDB with export scripts) |
| Ad hoc EML assembly | Metadata collected for datasets only |
| Enhance existing EML | EML exists, let's annotate it |

LTER Workflow



LTER Core Metabase

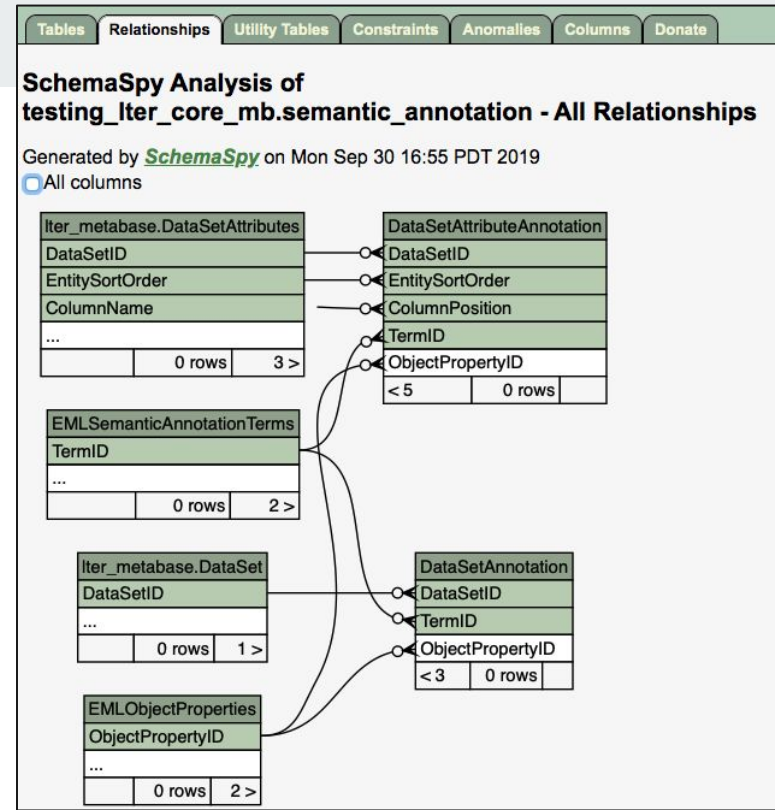
You will need

RDB Tables for annotation

- Dataset-level
- Attribute-level

Code to support

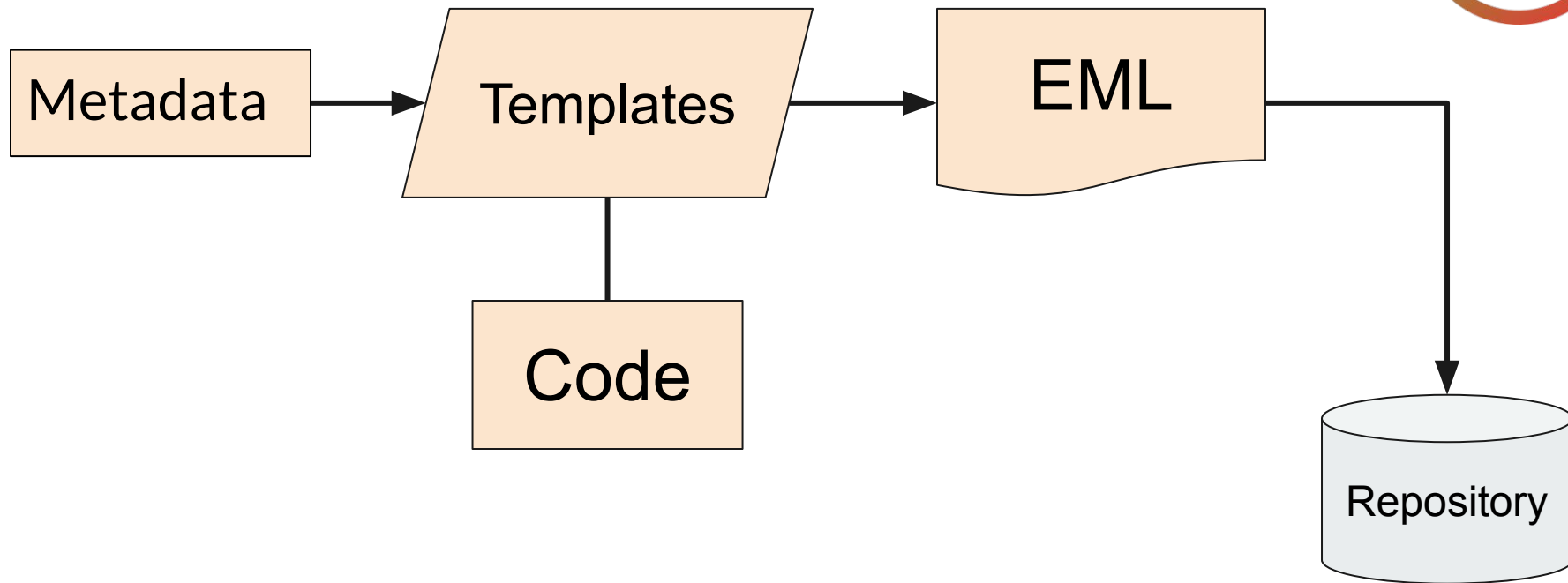
- EML 2.2
- Annotations



<https://github.com/lter/LTER-core-metabase>
<https://github.com/ble-lter/metaegress>



Ad Hoc Workflow



EML Assembly Line



Option near end of an
EML-creation workflow

Or

Read in existing EML, add
annotations

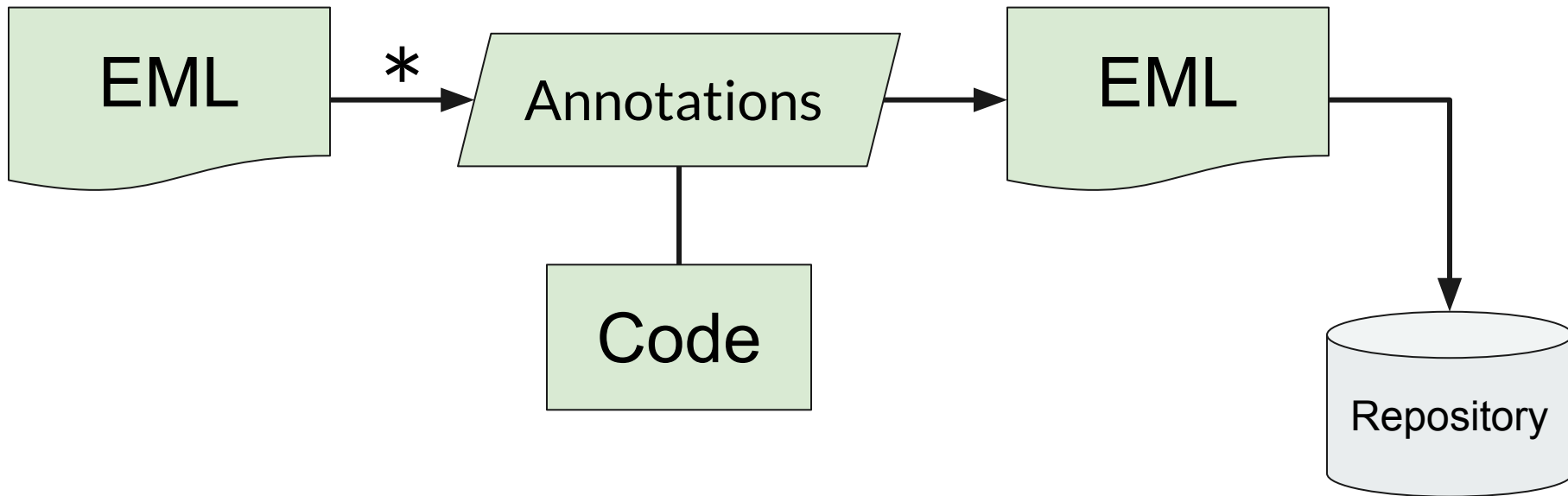
Annotation Template

- Single file, per other EAL templates

Output to

- Dataset level
- Entity level
- Attribute level
- Repeated at dataset/annotation

ADC Annotation Project, 2018-2019



ADC Annotation Project, 2018-2019



✱

Focus on attribute annotation

Identify datasets with "carbon cycling" measurements

(e.g. query carbon dioxide flux, CO₂ flux, carbon flux, NEE, Net Ecosystem Exchange, CH₄ flux, methane flux, gas flux ...)

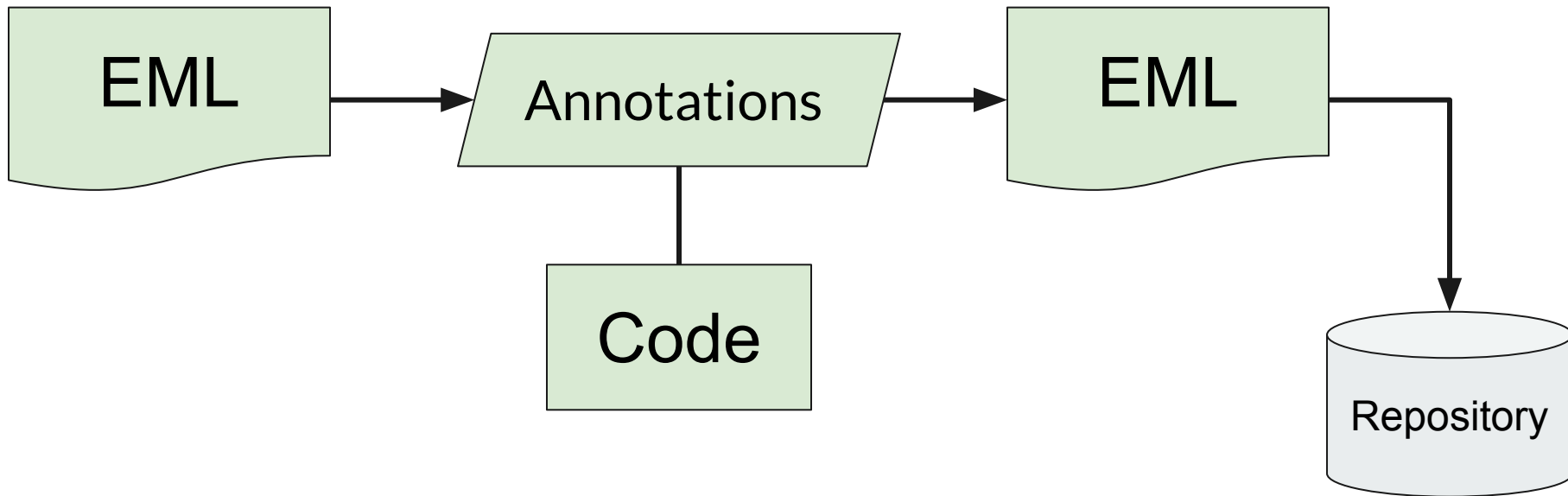
Extract attribute-level metadata from EML to a spreadsheet

(e.g., name, description, unit)

Build candidate terms for ECSO ontology

Add URIs for candidate annotations to the spreadsheet

ADC Annotation Project, 2018-2019



Lessons Learned



Consider a group of attributes together

Dataset-level and Attribute-level attotation have high return (discovery, understanding)

The ontology will be missing some terms

Annotation workshop



Late May 2020

EDI + ADC

Workshop activities

1. Examine, compare ontologies
2. Pick candidate concepts
3. Edit, refine with peer and expert input
4. Export EML in your preferred manner

Goal: annotations are added to *your* curation workflow

References



EML 2.2 Primer on semantic annotation in datasets

<https://eml.ecoinformatics.org/semantic-annotation-primer.html>

Background on URIs

<https://www.w3.org/TR/cooluris/>

<https://www.w3.org/2001/tag/awwsw/issue57/20110625/>

<https://www.gbif.org/document/80575/a-beginners-guide-to-persistent-identifiers>

ECSO presentation, ESIP 2017 (describes Dataone project):

https://docs.google.com/presentation/d/1Rh60ACyeaSew9ci2D3KSD4OjSCCL3biDHjPwK62j2uo/edit#slide=id.g23fe2b6583_0_41

Terms for the dataset level (URIs for LTER Controlled Vocabulary Terms):

<http://www.environmentontology.org/Browse-EnvO>





Thank you

