



EML 2.2

EML Dev Committee
2018

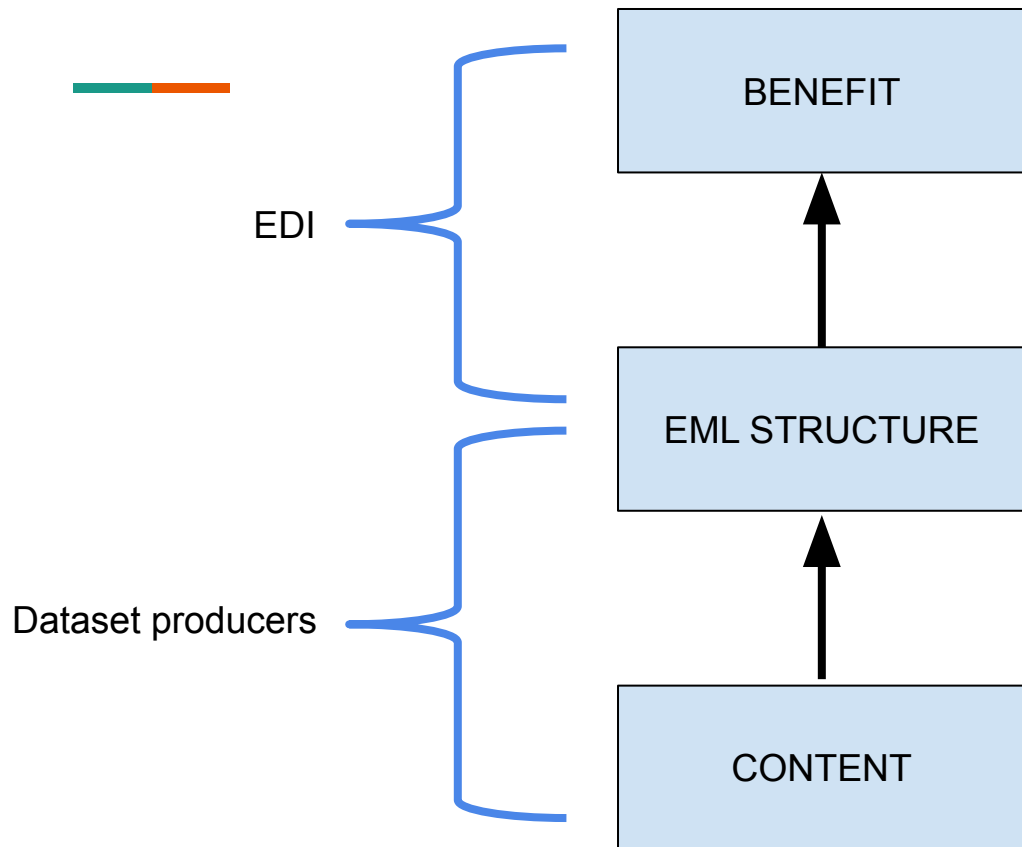


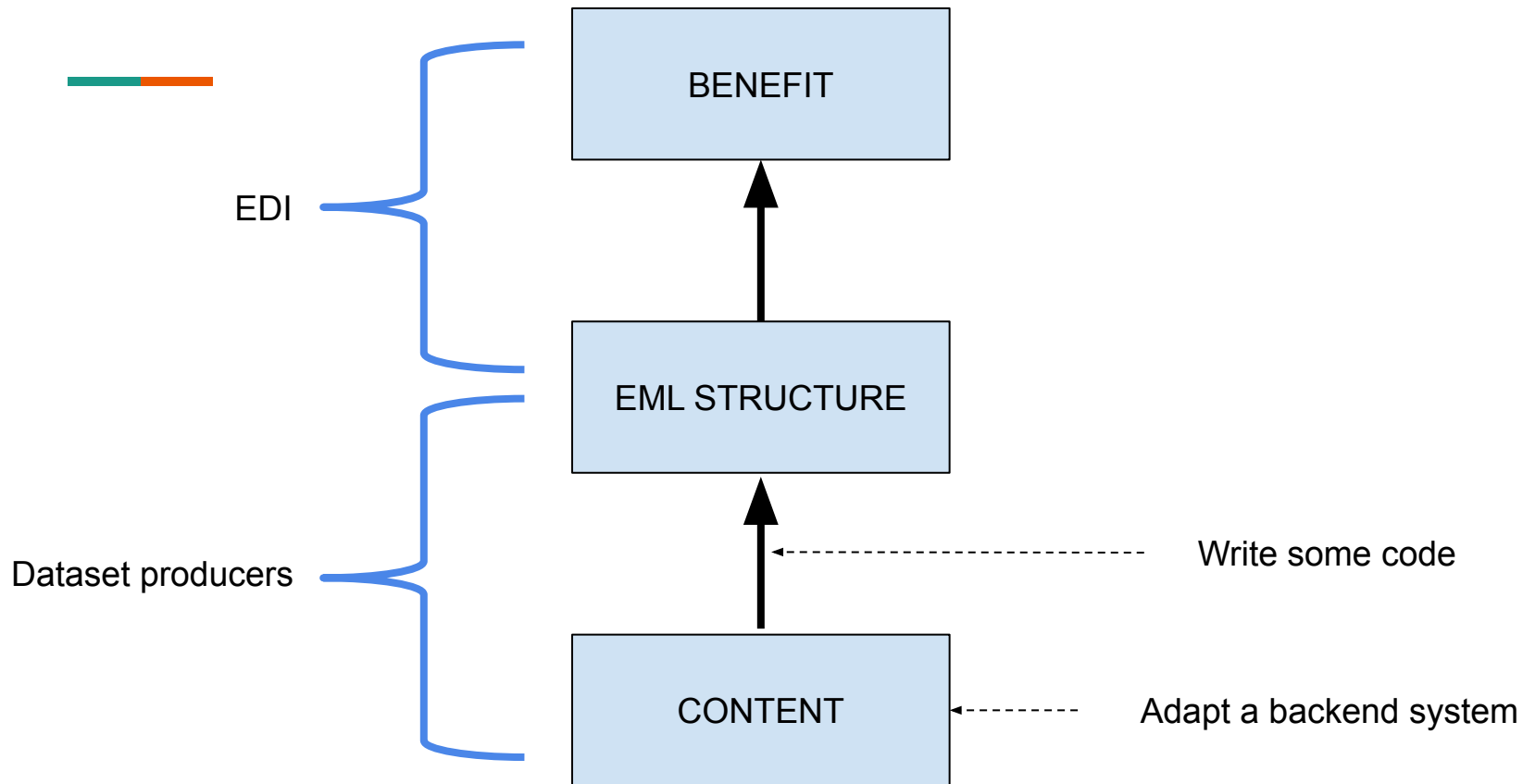
EML 2.2 New Features

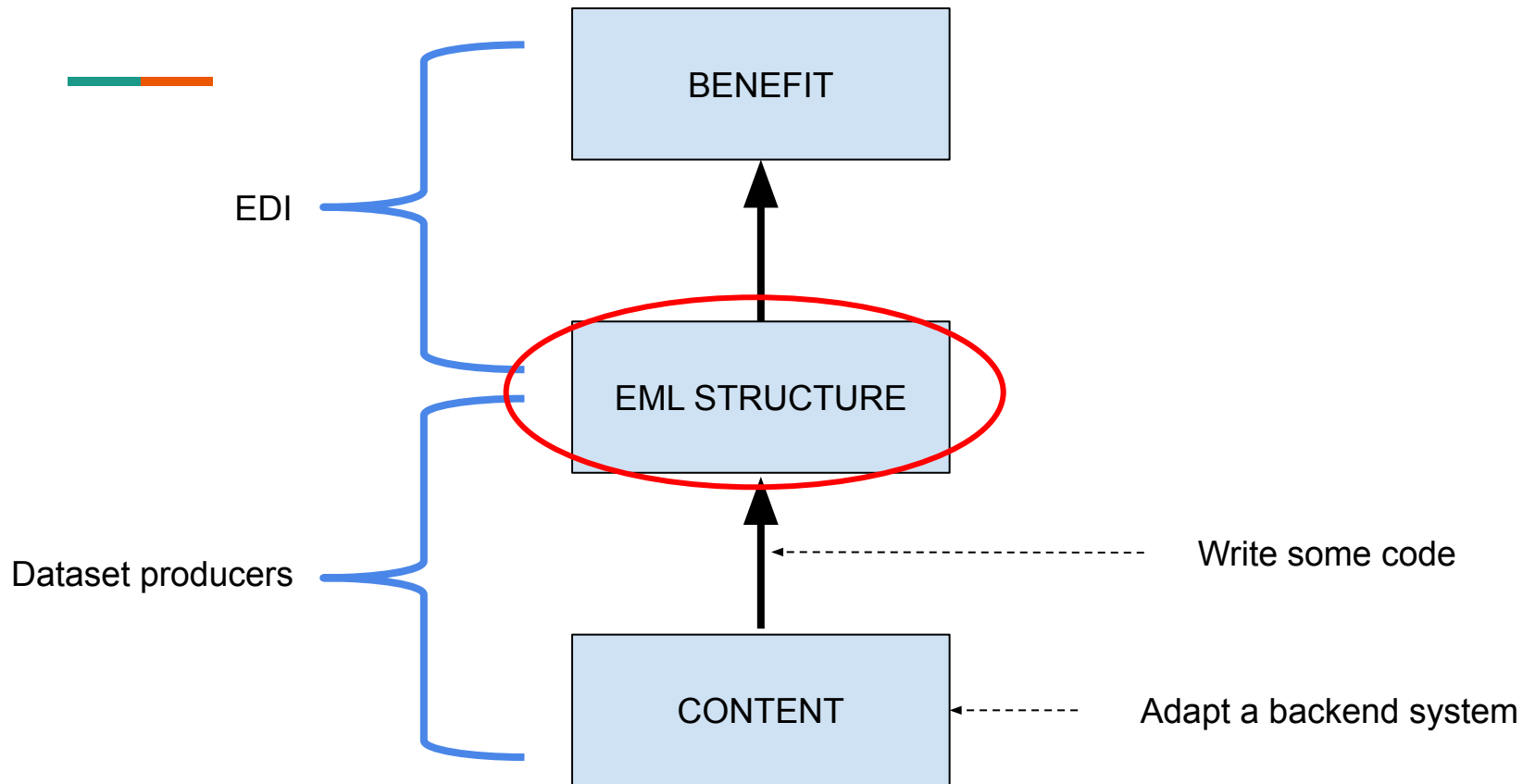


Backward compatible

EML you create now (2.1, 2.0) is fine as is
or re-label as EML 2.2







EML 2.2 New Features



Benefits:

- Data are easier to search
 - Allow targeted queries, e.g., funding codes
- Optional, simpler formats for some content
- EML datasets can approach completeness of data papers

EML 2.2 New Features



New features will take additional attention

Communities should plan

- Set priorities - science-driven

EML 2.2 New Features



- Choice of TextType or markdown
- Support for data papers
- Taxonomic classification has IDs
- Project tree includes more background
- Units Dictionary improved
- Annotations

EML 2.2 New Features



Backward compatible

EML you create now (2.1, 2.0) is fine as is
or re-label as EML 2.2

Root Element



```
<eml:eml xmlns:eml="https://eml.ecoinformatics.org/eml-2.2.0"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xmlns:stmml="http://www.xml-cml.org/schema/stmml-1.2" packageId="edi.999.1"
  xsi:schemaLocation="https://ecoinformatics.org/eml-2.2.0 https://nis.lternet.edu/schemas/EML/eml-2.2.0/xsd/eml.xsd"
  system="edi">
```

Namespaces

EML: xmlns: <https://eml.ecoinformatics.org/eml-2.2.0>

STMML: <http://www.xml-cml.org/schema/stmml-1.2>

Schema location

<https://nis.lternet.edu/schemas/EML/eml-2.2.0/eml.xsd>



TextType

*Benefit: For some dataset builders, markdown
will be easier to handle*

TextType or markdown

EML 2.2

TextType
or
markdown

->

EML 2.1

TextType

->

```
<abstract>
<para>These data describe the abundance of giant kelp located at SBC LTER long-term kelp
removal experiment sites. The number of fronds on each plant > 1 m tall within 40 m x 2 m
transects in each plot is recorded in selected 20 m2 sub-sections
of each transect. In continuous removal plots plants are removed after being sampled. </para>
<para>The kelp removal experiment was initiated in 2008 at selected reef sites along the
mainland coast of the Santa Barbara Channel. The following three giant kelp manipulations
are maintained in sampling plots at each site: </para>
<para>- Control: <em>Macrocystis pyrifera</em> is left intact in a 2000
m2 plot. </para>
<para>- Annual removal: <em>Macrocystis pyrifera</em> is removed from an adjacent
2000 m2 plot once each winter and allowed to re-colonize over the
course of the year. </para>
<para>- Continuous removal: <em>Macrocystis pyrifera</em> is removed from a 200
m2 area within the 2000 m2 plot on
each sampling date. </para>
</abstract>
```

```
<abstract>
```

```
<markdown><![CDATA[
```

```
Some intro text in abstract, then break into subsections.
```

```
## Level 2 heading
```

```
We use a level 2 heading because Level 1 would be at the same level as
the main sections of the paper.
```

```
## Another level 2 heading
```

```
With some information.
```

```
Plus, it can include all of the other features of
```

```
[Github Flavored Markdown (GFM)](https://github.github.com/gfm/).
```

```
Note that this version of GFM is a superset of CommonMark, and is
intended to eventually be an official extension of CommonMark.
```

```
]]>
```

```
</markdown>
```

```
</abstract>
```



CitationType

*Benefit: EML includes more structure than the
ASCII text documents typically used for data
papers*

Citations - Summarized

Element Name	parent	appeared in EML	How to use
referencePublication	dataset	2.2	when citing THE primary paper that describes how the dataset was generated
literatureCited	dataset	2.2	list of articles which were referenced in the dataset or its associated metadata
citation	methodStep	2.1	reference a literature resource for a method
citation	project	2.1	reference a paper describing a project
usageCitation	dataset	2.2	subsequent uses of the dataset e.g., in meta-analyses

CitationType



Citations appear at

- `/eml/dataset/referencePublication`
- `/eml/dataset/literatureCited`
- `/eml/dataset/usageCitation`
- `/eml/dataset/methods/methodStep/citation`
- `/eml/dataset/project/designDescription/citation`

Citations can be
structured as either
XML or as bibtex

CitationType XML elements

- Authors: `//citation/creator`
 - `responsiblePartyType`
- Year: `//citation/pubDate`
 - consistent with `/eml/dataset/pubDate`
- Title: `//citation/title`
- Journal: `//citation/article/journal`
- Volume: `//citation/article/volume`
- Issue: `//citation/article/issue`
- Page Range: `//citation/article/pageRange`

literatureCited (bibtex option)

Holds a list
of citation
elements

```
</dataTable>
<literatureCited>
  <citation>
    <bibtex>
      @article{
        title={Major shifts at the range edge of marine forests: the combined effects of climate changes and limited dispersal},
        author={Assis, J and Berecibar, E and Claro, B and Alberto, F and Reed, DC and Raimondi, PT and Serrao, EA},
        year={2017},
        journal={Scientific Reports},
        volume={7},
        pages={44348},
        doi={10.1038/srep44348},
      }
    </bibtex>
  </citation>
  <citation>
    <bibtex>
      @article{
        title={Assessing controls on cross-shelf phytoplankton and suspended particle distributions using repeated bio-optical glider surveys},
        author={Hendrikx Freitas, F and Siegel, DA and Washburn, L and Halewood, S and Stassinis, E},
        year={2016},
        journal={Journal of Geophysical Research - Oceans},
        volume={121},
        pages={7776-7794},
        doi={10.1002/2016JC011781},
      }
    </bibtex>
  </citation>
</literatureCited>
</dataset>
```


Fields to Support Data Papers

EML 2.1 includes...

- `/eml/dataset/title`
- `/eml/dataset/creator`
- `/eml/dataset/keywordSet`
- `/eml/dataset/coverage/geographicCoverage`
- `/eml/dataset/coverage/temporalCoverage`
- `/eml/dataset/abstract`
- `/eml/dataset/methods/sampling/studyExtent/description`
- `/eml/dataset/methods/sampling/samplingDescription`
- `/eml/dataset/methods/methodStep`
- `/eml/dataset/methods/qualityControl`
- `/eml/dataset/creator/@userId`
- `/eml/dataset/creator/electronicMailAddress`
- `/eml/dataset/creator/address`

EML 2.2 adds...

- **TextType**
 - Introduction `/eml/dataset/introduction`
 - Data synopsis `/eml/dataset/purpose`, for tables, figs
 - Getting started `/eml/dataset/gettingStarted`
 - Acknowledgements `/eml/dataset/acknowledgements`
- **CitationType**
 - `/eml/dataset/referencePublication`
 - `/eml/dataset/literatureCited`
 - `/eml/dataset/usageCitation`



Taxonomic IDs

Benefit: Links to external taxonomic registries

Taxonomic Registries



Example	Coverage
ITIS	~10% of described species on earth
USDA	Terrestrial plants
Catalog of Life	> 100 expert taxonomic DBs
WoRMS	Temperate marine
GBIF Backbone Taxonomy	Aggregates several databases

Taxonomic Classification

EML
2.1

->

```
<taxonomicClassification>
  <taxonRankName>genus</taxonRankName>
  <taxonRankValue>Macrocystis</taxonRankValue>
  <taxonomicClassification>
    <taxonRankName>species</taxonRankName>
    <taxonRankValue>Macrocystis pyrifera</taxonRankValue>
    <taxonomicClassification>
      <taxonRankName>commonName</taxonRankName>
      <taxonRankValue>MAPY</taxonRankValue>
    </taxonomicClassification>
  </taxonomicClassification>
</taxonomicClassification>
```

EML
2.2

->

```
<taxonomicClassification>
  <taxonRankName>genus</taxonRankName>
  <taxonRankValue>Macrocystis</taxonRankValue>
  <taxonomicClassification id="taxon_MAPY">
    <taxonRankName>species</taxonRankName>
    <taxonRankValue>Macrocystis pyrifera</taxonRankValue>
    <commonName>Giant Kelp</commonName>
    <taxonId provider="https://itis.gov">11274</taxonId>
    <taxonId provider="https://www.ncbi.nlm.nih.gov/taxonomy">35122</taxonId>
  </taxonomicClassification>
</taxonomicClassification>
```



Project Tree

Benefit: Fine-grained queries for research project information and support

Project Tree - Additions

EML
2.1

->

```
<funding>
  <section>
    <para>wholly or in part by NSF Awards OCE-1233283, 1233288, 1233839</para>
  </section>
</funding>
```

EML
2.2

->

```
<project>
  ...
  <funding><para>Funding is from a grant from the National Science Foundation.</para></funding>
  <award>
    <funderName>National Science Foundation</funderName>
    <funderIdentifier>https://doi.org/10.13039/00000001</funderIdentifier>
    <awardNumber>1546024</awardNumber>
    <title>Scientia Arctica: A Knowledge Archive for Discovery and Reproducible Science in the Arctic</title>
    <awardUrl>https://www.nsf.gov/awardsearch/showAward?AWD_ID=1546024</awardUrl>
  </award>
</project>
```



Unit Dictionary

*Benefit: List is easier to peruse; convert using
UDUNITS tools*

Unit Dictionary



Summary

Unit Types:

- Several renamed to better reflect their dimensions

Units added to “standard list”

- based on lists assembled from LTER and ADC

- for consistent spelling of unitId and unitName

- inconsistent form marked as deprecatedInFavorOf

UDUNITS conversion

- UDUNITS synonym included to assist with conversion using that package

- (<https://www.unidata.ucar.edu/software/udunits/>)

Unit Dictionary

EML
2.1

->

```
<unit id="bushel" name="bushel" unitType="volume" parentSI="liter" multiplierToSI="0.035239" abbreviation="b">
  <description>1 bushel = 35.23907 liters</description>
</unit>

<unit id="cubicInch" name="cubicInch" parentSI="liter" unitType="volume" multiplierToSI="0.000016387064" abbreviation="in³">
  <description>cubic inch</description>
</unit>
```

EML
2.2

->

```
<unit id="bushel" name="bushel" unitType="volume" parentSI="liter" multiplierToSI="35.23907"
  abbreviation="b" udunitsSynonym="bushel">
  <description>1 bushel = 35.23907 liters</description>
</unit>

<unit id="cubicInch" name="cubicInch" parentSI="liter" unitType="volume"
  multiplierToSI="0.01638706" abbreviation="in³" deprecatedInFavorOf="inchCubed"
  udunitsSynonym="international_inch^3">
  <description>cubic inch</description>
</unit>
```



Annotations

Benefit: Links to external vocabularies, which themselves are highly structured

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>
```

Example - Annotation at the ADC

The screenshot displays the Arctic Data Center (ADC) 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". The "Annotation" filter is expanded, showing a search for "Carbon Flux" with a tooltip that reads: "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: "Fire influences on forest recovery and associated climate feedbacks in Siberian Larch Forests, Russia, June-July 2018" and "Beaufort Gyre Observing System (BGOS) Underway pCO2 data and air-sea CO2 fluxes". On the right, a map shows the Arctic region with a grid overlay. A "Hide Map" button is at the top of the map area.

NSF ARCTIC Data Center

Data Support About Community Submit Data

Sign in with ORCID

Search ?

Search phrase

Filter by:

Data attribute

Annotation

Search for class...

Carbon Flux

Year

Identifier

Taxon

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

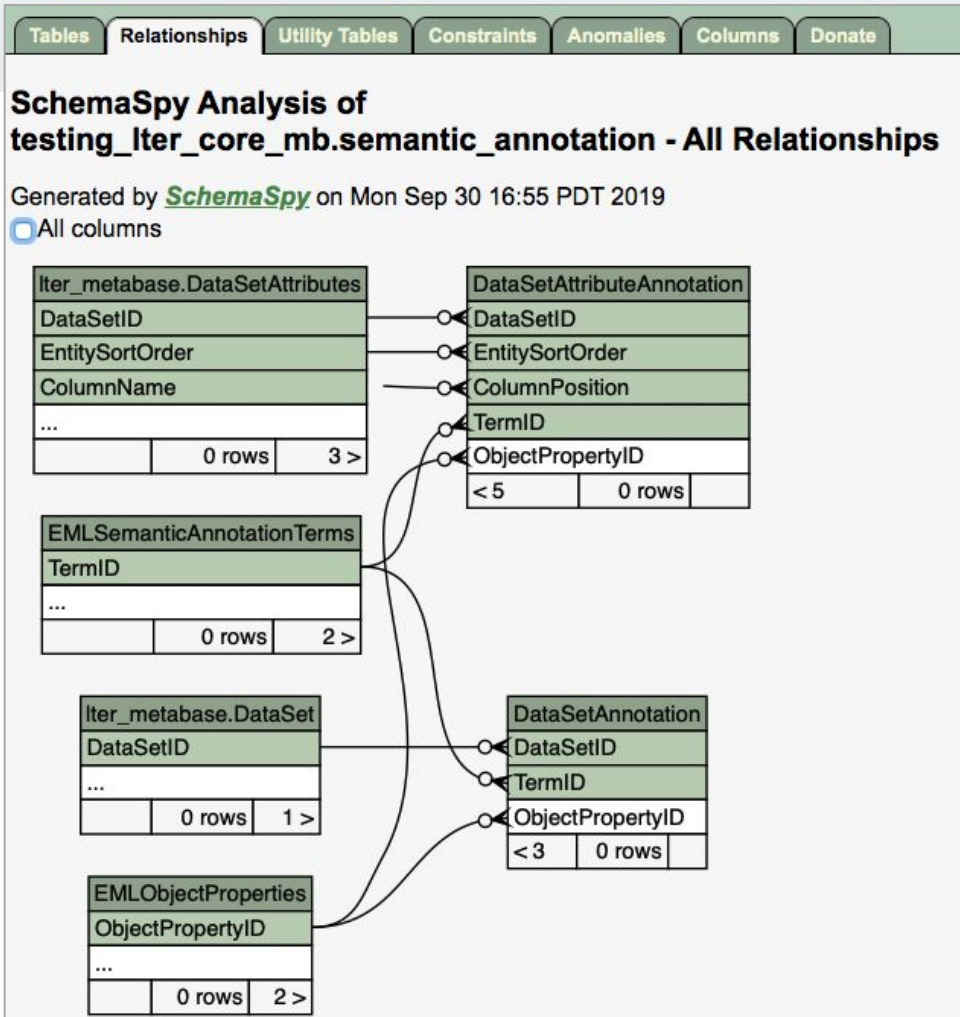
3 259 72 86 144 32 1 5 5 1 1

5 1 22 99 3 3 1 1

LTER Core Metabase

Tables for annotation

- Dataset-level
- Attribute-level



Discussion and Questions

- EML Best Practices updated
 - Volunteers, please
- How can we help users find datasets?
 - Improve precision and recall
 - With features of EML 2.2?
- How best can we encourage to use of consistent terms
 - whether in keywords or annotations
- Advanced features will take additional attention
 - guidelines for classifying datasets vis-a-vis EML features



Appendix

Slides included for specific questions or further discussion

- EDI support for EML 2.2
- Using external vocabularies
- Annotation background
- Precision and recall in searches

