

# Semantics System Overview

The Semantics Team:

Matthew B. Jones, Mark Schildhauer, Ben Leinfelder, Xixi Luo, Deborah McGuinness, Lauren Walker, Chris Jones, Rachel Volentine, Margaret O'Brien, Sophie Hou, Chace LeCroy, Yaxing Wei, Steve Aulenbach, Christopher Schwalm

February 6, 2015



# Semantics: enabling improved discovery

---

Goal: Improve precision and recall of search

Annotating data with semantic types

Extending search systems to incorporate semantics

Display and edit semantic annotations in the web UI

# Semantics Use Cases

---

Use Case 50	<b>Automatic Annotation</b>	Automatic assignment of measurement types via algorithm
Use Case 51	<b>Manual Annotation</b>	Manual assignment of measurement types by people
Use Case 52	<b>Semantic Search</b>	Improved precision and recall of data search via semantics
Use Case 53	<b>Reject Annotation</b>	Curation and rejection of annotations by data owners
Use Case 54	<b>Annotation Conflicts</b>	Resolution of conflicts from multiple annotations on one data
Use Case 55	<b>Ontology Repository</b>	A repository to store, version, and access ontologies
Use Case 56	<b>Carbon Cycle Ontology</b>	Knowledge Models for the carbon cycle and primary productivity

# Semantics at-a-glance

---

New first-class objects in DataONE

Semantic Annotations

New Tools

New model for linking semantics to data

New KR models for Carbon Cycle and related areas

Ontology repository

Automated classifier to assign semantics to variables

Web UI for browsing and editing semantics



# Measurement Annotation Display

Number Of  
Records

3898583

Attribute Information

Variables

date

station

lat

lon

depth

temp

salinity

density

Name

density

Flag this tag as incorrect

Tags

Volumetric Density

Mass

Label

Definition

Ocean density

Storage Type

Measurement  
Type

ratio

Measurement  
Domain

Unit

kilogramPerCubicMeter

Type

real

1

3

+

# Annotate

Number Of Records3898583

Variables

date

station

lat

lon

depth

temp

salinity

density

Name

density

Tags

Volumetric Density

Label

Definition

Ocean density

Storage Type

Measurement Type

ratio

Measurement Domain

Unit

kilogram

Type

real

Missing Value Code

Accuracy Report

Accuracy Assessment

Coverage

1

2

Sarah Smith

April 26, 2014 5:24 PM

Added a tag

Density

Definition: Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur.

from the SWEET ontology

Bill Kertscher

May 25, 2014 5:24 PM

Added a tag

Volumetric Density

Definition: The ratio of the mass of any substance to the volume occupied by it (usually expressed in kilograms per cubic meter, but any other unit system may be used); the reciprocal of specific volume

from the SWEET ontology

Bill Kertscher

May 25, 2014 5:24 PM

I chose this tag because consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit

Help others find this dataset by adding tags and comments

Add a comment...

Volume

Volume

Volumetric

Volumetric Heat

Volumetric Heat Capacity

Volume Fraction

People and Associated Pa

Data Set

Creators

Indiv

Definition

Sed posuere consectetur est at lobortis. Aenean eu leo quam. Pellentesque ornare sem lacinia quam venenatis vestibulum.



# Search for Semantic Measurements

The screenshot displays the DataONE web interface. At the top, there's a navigation bar with 'DataONE', 'DATA', and 'ABOUT' links. A search bar labeled 'Search for data' is prominent. Below the search bar, a map of the North Pacific Ocean is shown with a heatmap overlay representing data density. The map includes labels for 'Bering Sea', 'Sea of Okhotsk', 'North Pacific Ocean', and 'South Pacific Ocean'. A 'Filter' panel on the left allows users to refine their search by 'Data attribute (density,)' with options like 'Temperature', 'Time', 'Type', 'Count', 'Length', and 'Proportion'. A 'Definition' pop-up window is visible, displaying a Latin phrase: 'Sed posuere consectetur est at lobortis. Aenean eu leo quam. Pellentesque ornare sem lacinia quam venenatis vestibulum.' On the right, a list of datasets is shown, including 'Historical CTD data from the Gulf of Alaska, 1970-2010' by Thomas J. (Tom) Weingartner, Thomas Royer, and Russell Hopcroft, and 'Relationship of Community Characteristics to Harvest Reporting: Comparative study of household surveys and harvest tickets in Alaska' by Jennifer I. Schmidt and Bonanza Creek LTER.

Search for data

DataONE DATA ABOUT

Datasets 1 to 25 of 267

1 2 3 ... 11 Next

Sort by Most recent

Thomas J. (Tom) Weingartner, Thomas Royer, and Russell Hopcroft. 2014. **Historical CTD data from the Gulf of Alaska, 1970-2010.** (doi:10.5063/F1JM27JF)

Jennifer I. Schmidt, and Bonanza Creek LTER. 2014. **Relationship of Community Characteristics to Harvest Reporting: Comparative study of household surveys and harvest tickets in Alaska.** knb-lter-bnz.578.2

Mayumi Arimitsu. 2014. **seabird and mammal surveys in Kenai Fjords National Park, Alaska: 2006-2008, PWS: 2012.** df35b.219.11

Kristen L Manies, Jennifer W. Harden, and Bonanza Creek LTER. 2013. **Additional sites added to the Delta Junction fire-chronosequence soil database: carbon, nitrogen, bulk density, and other properties.** knb-lter-bnz.540.3

Edward A.G. Schuur, Christina Schadel, and Bonanza Creek LTER. 2013. **The impact of permafrost thaw on ecosystem carbon balance: Seasonal water table depth from 2012-2013.** knb-lter-bnz.554.3

Edward A.G. Schuur, Caitlin Elizabeth Hicks-Bries, and Bonanza Creek LTER. 2013. **The**

Filter

Anything

Data attribute (density,)

Temperature

Time

Type

Count

Length

Proportion

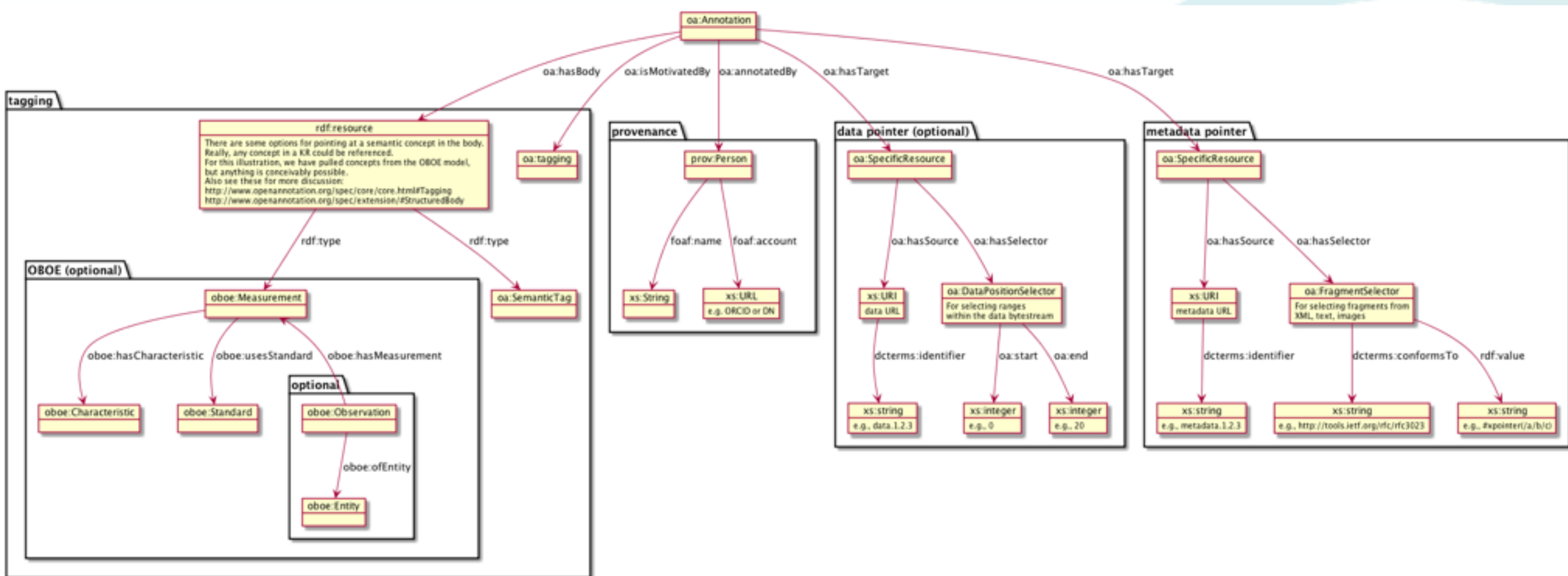
Definition

Sed posuere consectetur est at lobortis. Aenean eu leo quam. Pellentesque ornare sem lacinia quam venenatis vestibulum.

1900 2014

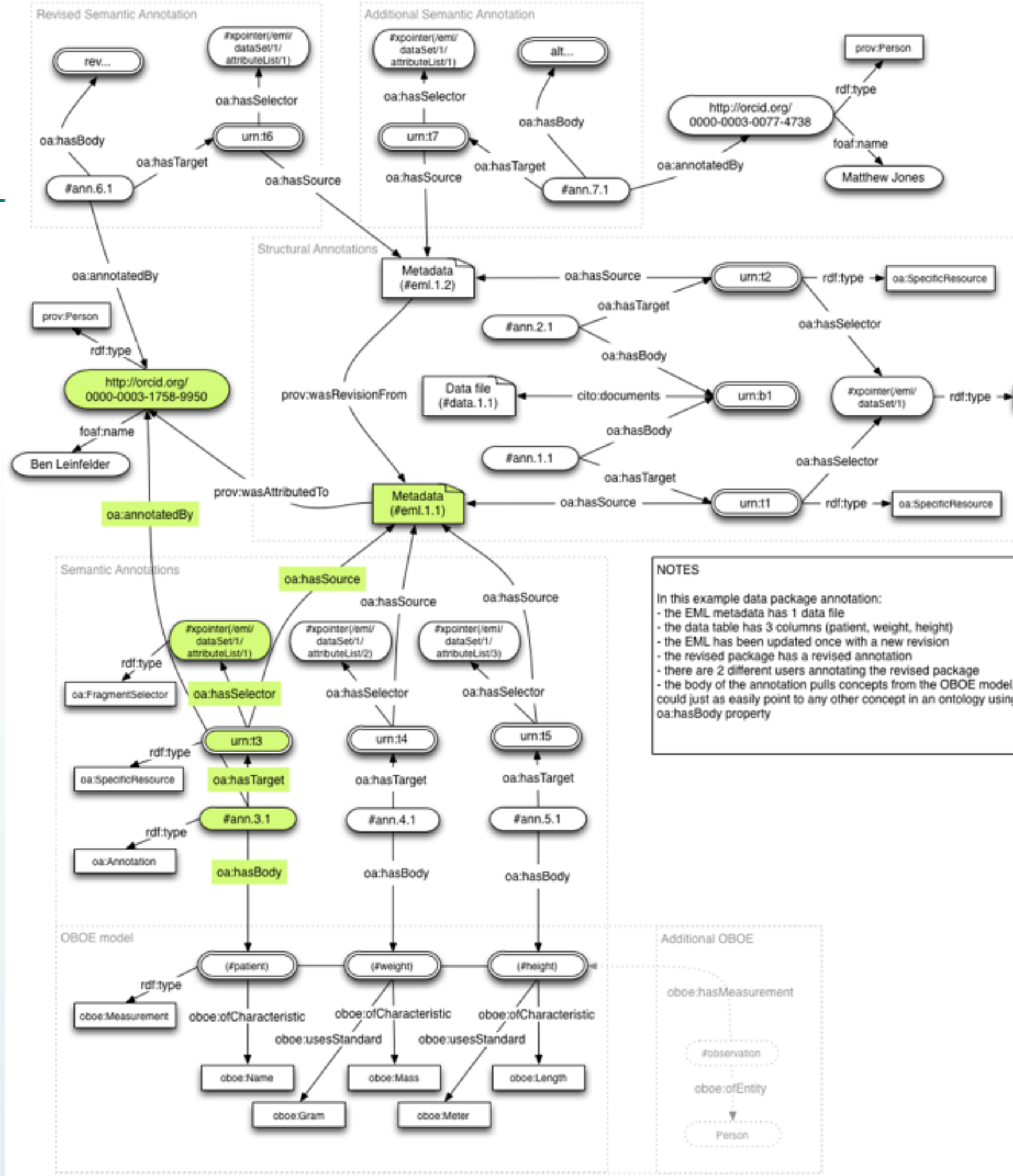
Data covers

# Semantic Annotation model





# Semantic Annotation Example



NOTES

- the EML metadata has 1 data file
- the data table has 3 columns (patient, weight, height)
- the EML has been updated once with a new revision
- the revised package has a revised annotation
- there are 2 different users annotating the revised package
- the body of the annotation pulls concepts from the OBOE mode

could just as easily point to any other concept in an ontology using `oa:hasBody` property

# Carbon Cycle Ontology

- ▼ ● CarbonFlux
  - FireCarbonFlux
  - DissolvedOrganicCarbonFlux
  - NetEcosystemExchangeCarbonFlux
  - ▼ ● PrimaryProductionCarbonFlux
    - ≡ GPP = GrossPrimaryProductionCarbonFlux = Gros
    - ≡ GrossPrimaryProductionCarbonFlux = GPP = Gros
    - ≡ GrossPrimaryProductivity = GPP = GrossPrimaryP
    - ▼ ≡ NetPrimaryProductionCarbonFlux = NetPrimaryPr
      - ▶ ● AnnualNetPrimaryProductivity
      - ≡ NetPrimaryProductivity = NetPrimaryProductionC
      - ≡ NPP = NetPrimaryProductivity = NetPrimaryProdu
  - ▶ ● RespirationCarbonFlux
- ▼ ● CarbonPool
  - AboveGroundWoodyBiomassCarbonPool
  - BenthicAlgaeCarbonPool
  - GrassCarbonPool
  - LeafCarbonPool
  - LeafLitterCarbonPool
  - PlanktonCarbonPool
  - SoilCarbonPool
  - TotalLivingCarbonPool
  - WoodCarbonPool
- Measurement
- Method
  - ▼ ● FieldMethod
    - AllometricMethod
    - Carbon14UptakeMethod
    - EddyCovarianceMethod
    - OxygenEvolutionLightDarkMethod

FireCarbonFlux

Description: FireCarbonFlux

SubClass Of +

● CarbonFlux ?

General class axioms +

SubClass Of (Anonymous Ancestor)

Members +

Target for Key +

Disjoint With +

● NetEcosystemExchange ?  
PrimaryProductionCarl  
DissolvedOrganicCarbo  
RespirationCarbonFlux

# Semantics status (as of Feb 1, 2015)

---

Design documentation

First draft completed

Annotation Model

Drafted, getting feedback

Indexing and Search

In progress

Web display

Mockups done, prototype development started

Carbon Cycle Model

Initial draft, in revision

Ontology repository

Using BioPortal