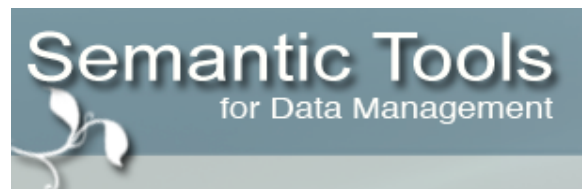


Semantic Metadata for Data Discovery and Integration

Ben Leinfelder,
Shawn Bowers, Margaret O'Brien, Matthew B. Jones, Mark Schildhauer

National Center for Ecological Analysis and Synthesis (NCEAS)
University of California Santa Barbara

EIM 2011
September 28, 2011



kelp

Search KNB


173 data packages found

Free-text Metadata Query

Performed on Title, Abstract, Keywords, Methods, etc...

- “find *all data packages containing the word ‘kelp’*”
 - does not match ‘**Macrocystis**’ keyword
- “find *all data packages containing the words ‘mass’ and ‘kelp’*”
 - data might record the ‘**mass**’ of a crab living in ‘**kelp**’
 - or ‘kelp’ from ‘**Mass**’achusetts

Metadata Inspection



The Knowledge Network for Biocomplexity




[Home](#)
[Data](#)
[People](#)
[Informatics](#)
[Biocomplexity](#)
[Education](#)

Data Set Citation

PISCO, Carr M of PISCO, and Caselle J of PISCO. **PISCO: Subtidal: Community Surveys: Fish Survey.**
[pisco_subtidal.150.2](http://knb.ecoinformatics.org/knb/metacat/pisco_subtidal.150.2/knb) (http://knb.ecoinformatics.org/knb/metacat/pisco_subtidal.150.2/knb).

Data Tables, Images, and Other Entities:

Metadata download: **Ecological Metadata Language (EML) File**

Data Table:	Taxonomic lookup table for subtidal data (View Metadata Download File )
Data Table:	Site table for subtidal survey data (View Metadata Download File )
Data Table:	pisco_fish_data.csv (View Metadata Download File )

Sampling Area: For a complete description see: <http://piscoweb.org/research/science-by-discipline/ecosystem-monitoring/kelp-forest-monitoring/subtidal-sampling-protocol>
 And Frequency:



‘kelp’

Semantic Query

Performed on the *observational data model*

- “find *all data sets containing observations of **Kelp***”
 - includes match for subclass ‘**Macrocystis**’ (recall)
- “find *all data sets containing **Mass** measurements of **Kelp***”
 - only data objects that record Kelp Mass (precision)

Metacat Semantic Query

Semantic Tools for Data Management

[Home](#)
[Login](#)
[Ontologies](#)
[BioPortal](#)

Semantic search

Entity, Characteristic, Protocol

Measurement

Options

Keywords

Spatial

Cart (0)

Find observations of

 ☒

- Entity
 - Organism
 - Algae
 - Kelp**
 - Macrocystis

with measurements of

 ☒

- Characteristic
 - PhysicalCharacteristic
 - Mass**
 - DryMass
 - WetMass
 - MassPerMass

and data values

using procedures outlined by

 ☐

- Protocol
 - ProtocolsByConstituent
 - ProtocolsByInstrumentation
 - ProtocolsByLaboratory
 - ProtocolsForElementalAnalysis
 - ProtocolsForOrganicTissue

Search criteria

Search Results

1 total records found

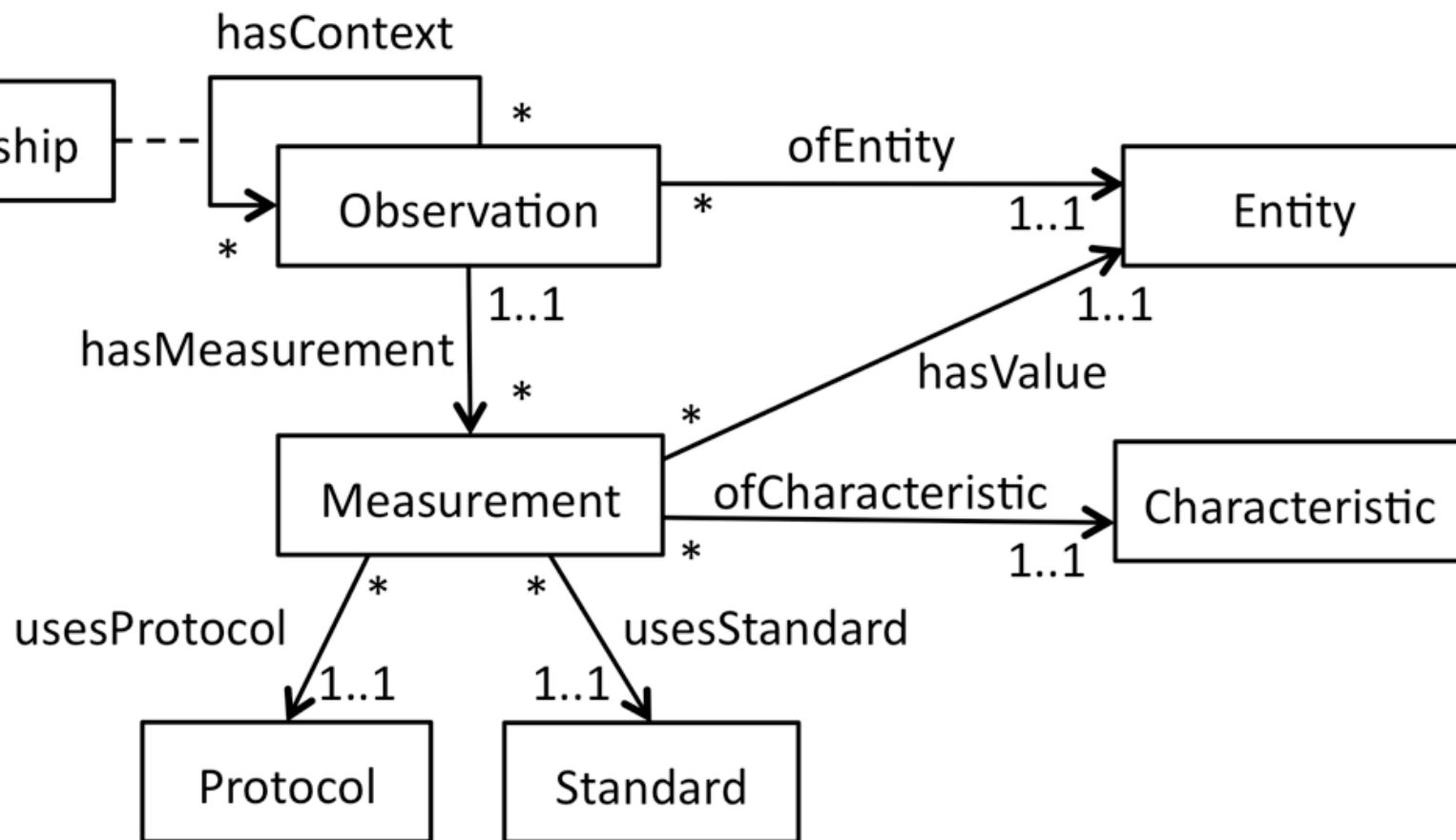
+ Reed, Reed. 2009. **SBCLTER Monthly Kelp CHN**. benriver.310.4.
(<http://bespin.nceas.ucsb.edu/knb/metacat/benriver.310.4/semtools>)

OBOE Core Concepts

A scientific *Observation* is the

- *Measurement* of the *Value*
- of a *Characteristic*
- of some *Entity*
- in a particular *Context*

OBOE Model



SBC-LTER Domain Ontology

- Extends OBOE using OWL-DL
 - Entities
 - *Kelp, Macrocystis*
 - Characteristics
 - *WetMass, DryMass*
 - Standards
 - *Gram*
 - Protocol
 - *AcousticDopplerCurrentProfileProtocol*
 - Context/Relationship
 - *Within*



Column Annotation
Context Annotation
Full Annotation
Graph Annotation

Selected Attribute: **WET**
☐ Is Distinct?
☐ Is Key?
[Help](#)
Edit

Use pre-configured measurement template:

The Characteristic

of the Entity was recorded

using the Standard

and the Protocol

WetMass

Macrocystis

Gram

Measurements were taken on the same sample or individual as:

Macrocystis [WET, DRY, _N, DRY_WET]

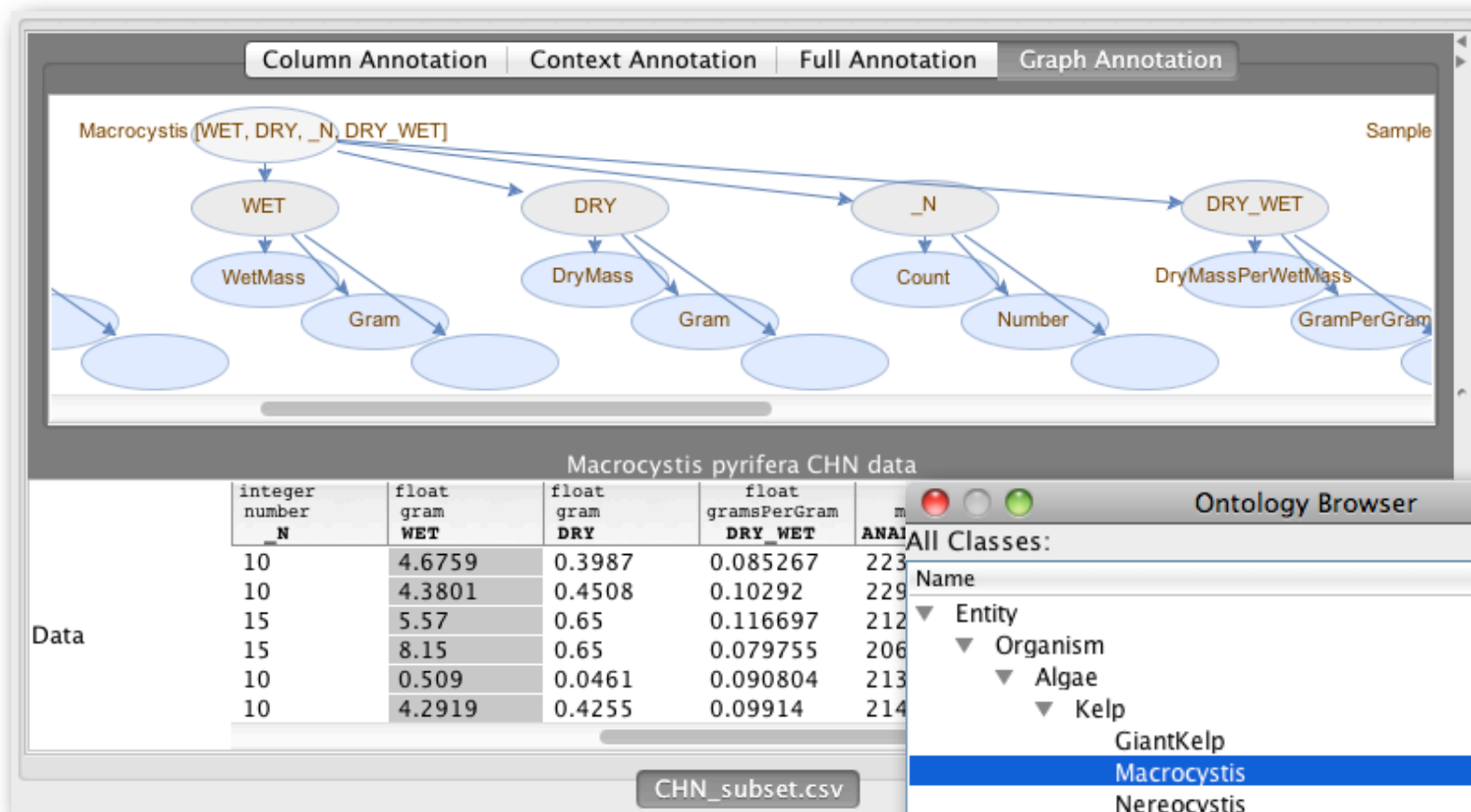
Macrocytis pyrifera CHN data

	string	integer	integer	string	integer	float	float
	SITE	YEAR	MONTH	DATE	number _N	gram WET	gram DRY
Data	ABUR	2002	6	06/12/2002	10	4.6759	0.3987
	ABUR	2002	7	07/12/2002	10	4.3801	0.4508
	AQUE	2008	11	11/06/2008	15	5.57	0.65
	AQUE	2008	12	12/04/2008	15	8.15	0.65
	MOHK	2002	5	05/07/2002	10	0.509	0.0461
	MOHK	2002	6	06/10/2002	10	4.2919	0.4255

CHN_subset.csv

“The *Characteristic* **WetMass** of the *Entity* **Macrocystis** was recorded using the *Standard* **Gram**”

Morpho Widgets



Ontology Browser

All Classes:

Name	Ontology
Entity	obo-core
Organism	OBOE SBC...
Algae	OBOE SBC...
Kelp	OBOE SBC...
GiantKelp	OBOE SBC...
Macrocytis	OBOE SBC...
Nereocystis	OBOE SBC...

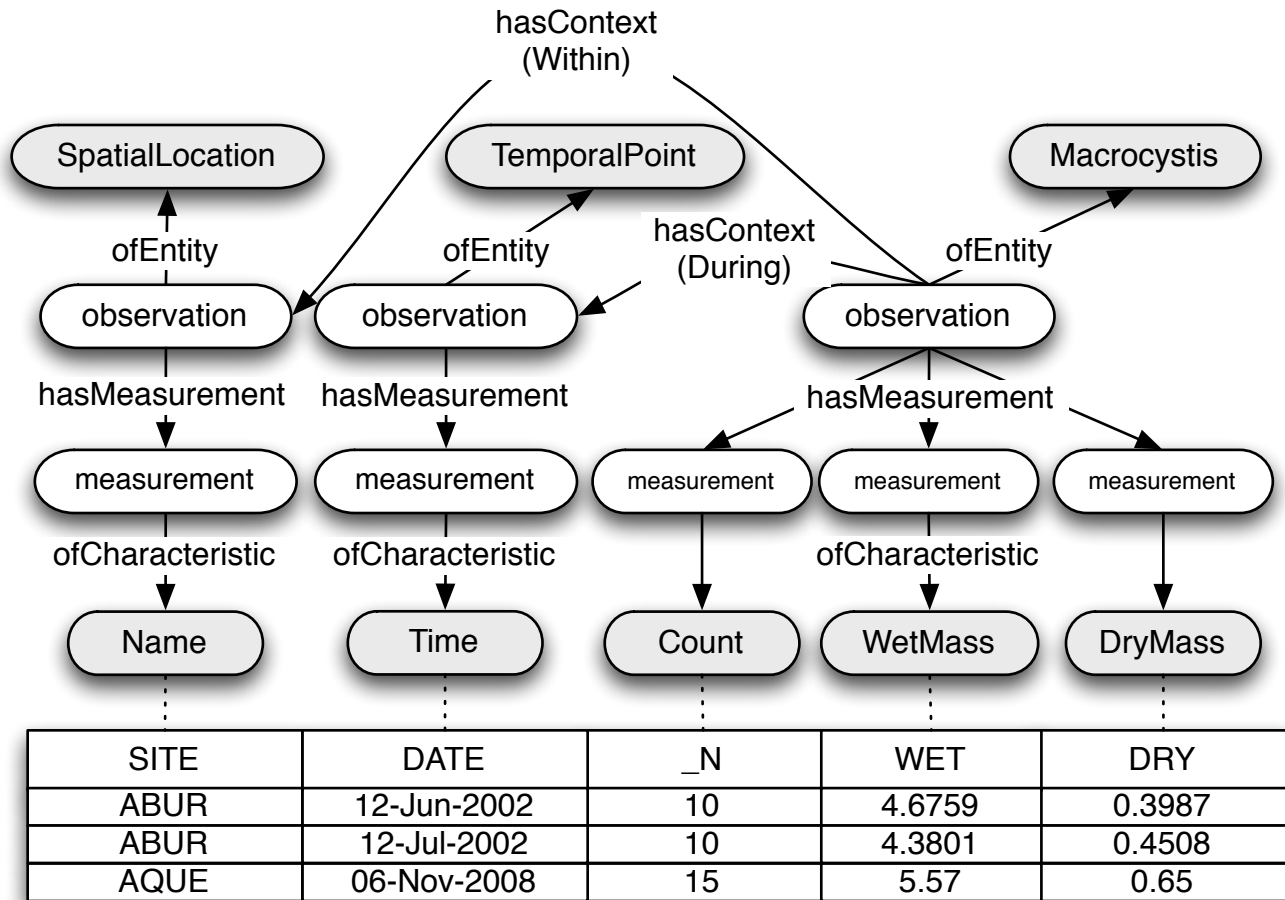
Description:

kelp

Search

Select

Semantic Annotation (OBOE)



Data-level Query

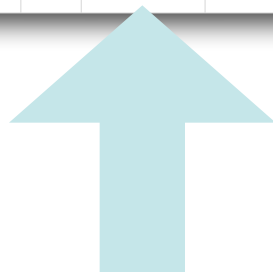
Performed on *observational data values*

- “find *all data sets containing **Mass** measurements of **Kelp** with **values less than or equal to 5 Grams***”
 - match only data values for kelp mass ≤ 5 g (precision)



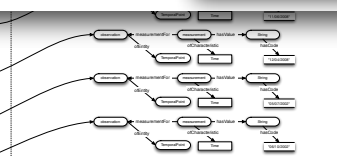
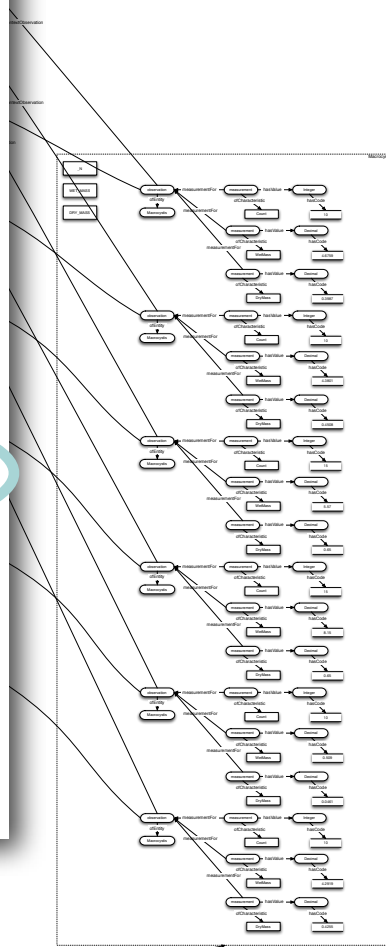
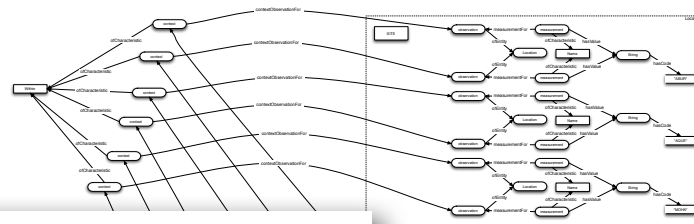
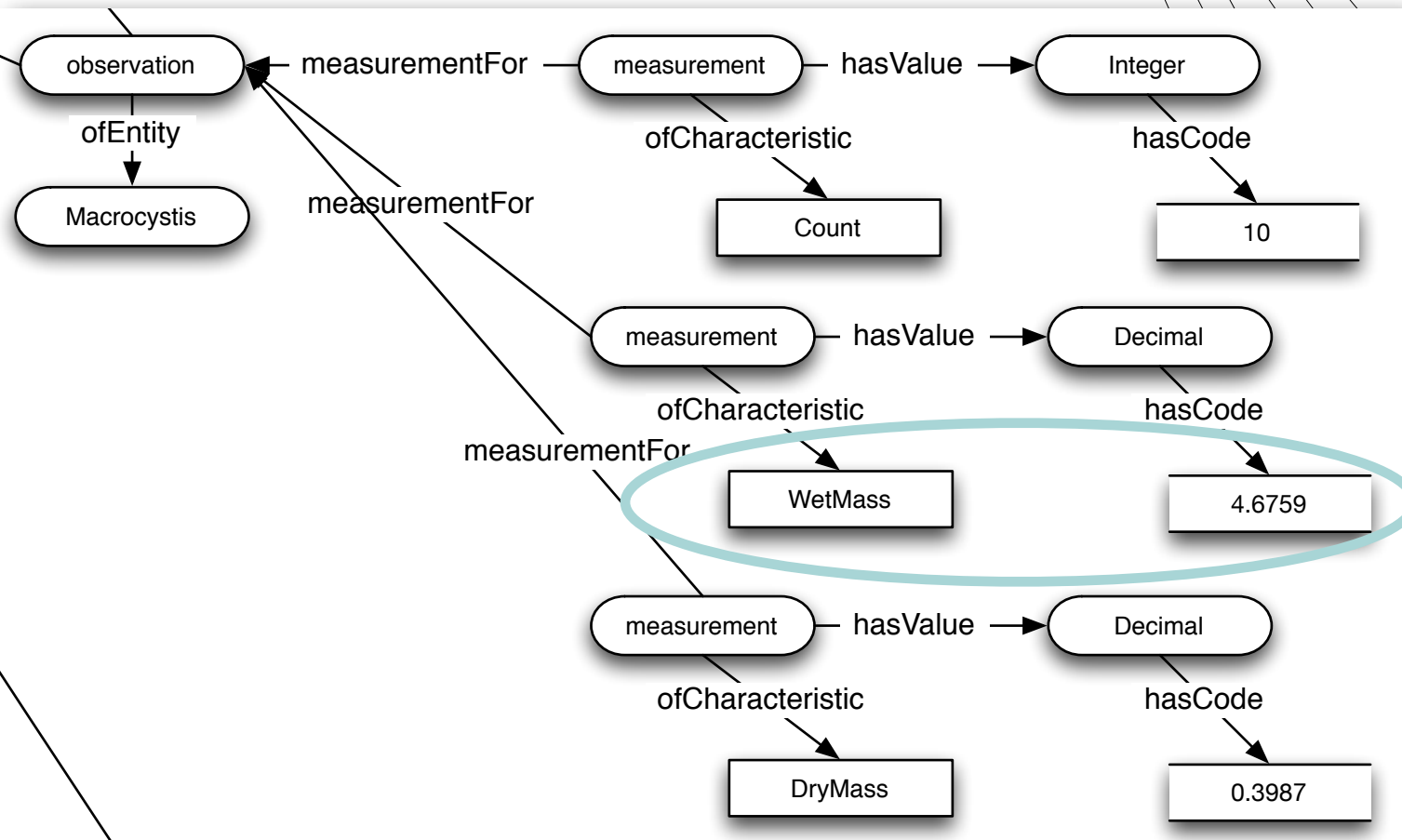
Macrocystis Data Set

SITE	YEAR	MONTH	DATE	_N	WET	DRY	DRY_WET	ANALYTICAL_WT	C	H	N	CN_RATIO
ABUR	2002	6	Jun 12, 2002	10	4.6759	0.3987	0.085267	2234	33.275	4.71	3.015	11.065
ABUR	2002	7	Jul 12, 2002	10	4.3801	0.4508	0.10292	2290.5	31.17	4.78	2.125	14.69
AQUE	2008	11	Nov 6, 2008	15	5.57	0.65	0.116697	2126.5	30.975	4.76	1.385	22.365
AQUE	2008	12	Dec 4, 2008	15	8.15	0.65	0.079755	2068	36.255	5.285	2.175	16.68
MOHK	2002	5	May 7, 2002	10	0.509	0.0461	0.090804	2135	22.57	3.2	2.8	8.06
MOHK	2002	6	Jun 10, 2002	10	4.2919	0.4255	0.09914	2142	24.13	3.4	1.78	13.57



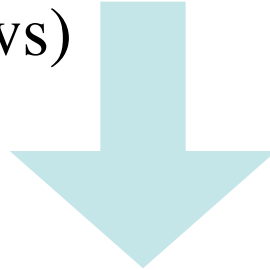
WetMass values in Grams

'Materialized' Data (OBOE)



Kelp Mass \leq 5 Grams

DryMass (6 rows)



SITE	YEAR	MONTH	DATE	_N	WET	DRY	DRY_WET	ANALYTICAL_WT	C	H	N	CN_RATIO
ABUR	2002	6	Jun 12, 2002	10	4.6759	0.3987	0.085267	2234	33.275	4.71	3.015	11.065
ABUR	2002	7	Jul 12, 2002	10	4.3801	0.4508	0.10292	2290.5	31.17	4.78	2.125	14.69
AQUE	2008	11	Nov 6, 2008	15	5.57	0.65	0.116697	2126.5	30.975	4.76	1.385	22.365
AQUE	2008	12	Dec 4, 2008	15	8.15	0.65	0.079755	2068	36.255	5.285	2.175	16.68
MOHK	2002	5	May 7, 2002	10	0.509	0.0461	0.090804	2135	22.57	3.2	2.8	8.06
MOHK	2002	6	Jun 10, 2002	10	4.2919	0.4255	0.09914	2142	24.13	3.4	1.78	13.57

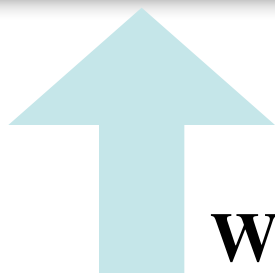


WetMass (4 rows)



Kelp Mass > 5 Grams

SITE	YEAR	MONTH	DATE	_N	WET	DRY	DRY_WET	ANALYTICAL_WT	C	H	N	CN_RATIO
ABUR	2002	6	Jun 12, 2002	10	4.6759	0.3987	0.085267	2234	33.275	4.71	3.015	11.065
ABUR	2002	7	Jul 12, 2002	10	4.3801	0.4508	0.10292	2290.5	31.17	4.78	2.125	14.69
AQUE	2008	11	Nov 6, 2008	15	5.57	0.65	0.116697	2126.5	30.975	4.76	1.385	22.365
AQUE	2008	12	Dec 4, 2008	15	8.15	0.65	0.079755	2068	36.255	5.285	2.175	16.68
MOHK	2002	5	May 7, 2002	10	0.509	0.0461	0.090804	2135	22.57	3.2	2.8	8.06
MOHK	2002	6	Jun 10, 2002	10	4.2919	0.4255	0.09914	2142	24.13	3.4	1.78	13.57

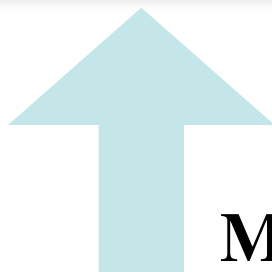


WetMass (2 rows)



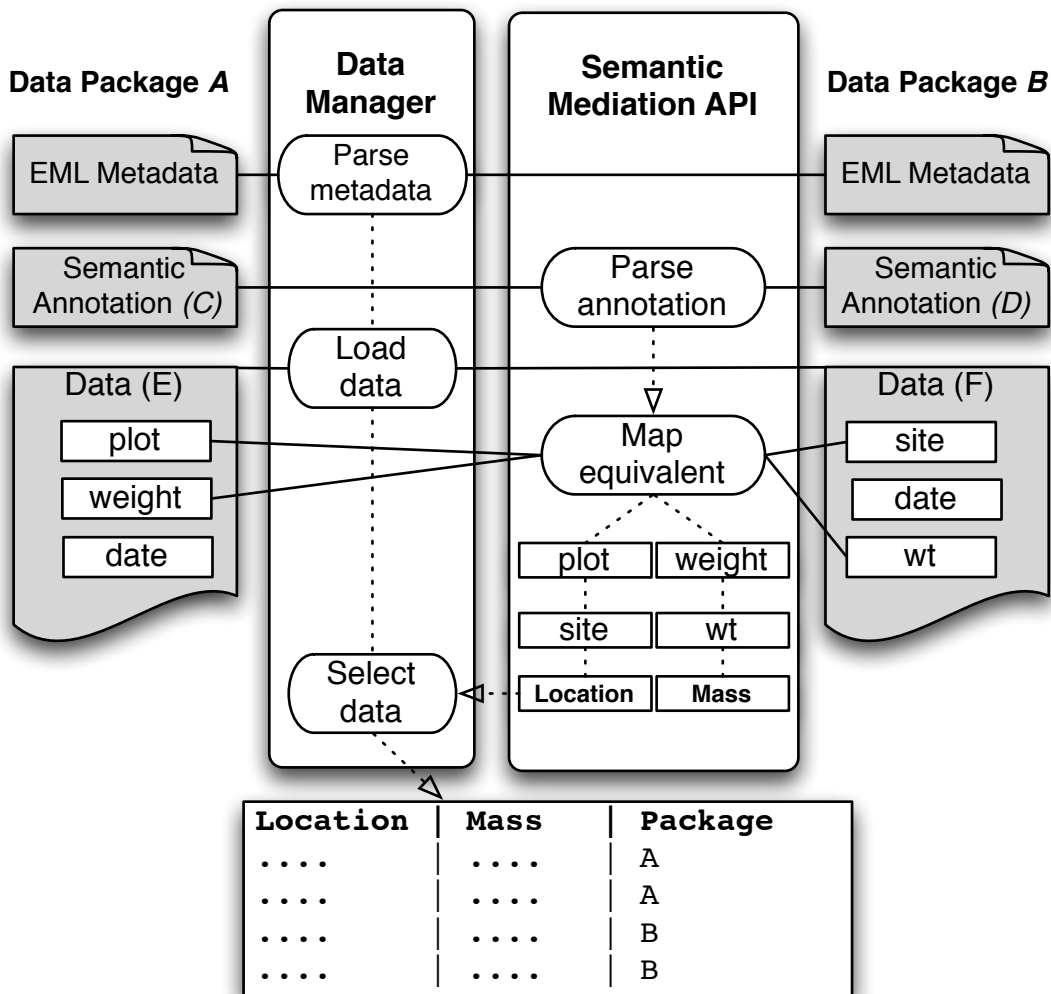
Kelp Mass > 10 Grams

SITE	YEAR	MONTH	DATE	_N	WET	DRY	DRY_WET	ANALYTICAL_WT	C	H	N	CN_RATIO
ABUR	2002	6	Jun 12, 2002	10	4.6759	0.3987	0.085267	2234	33.275	4.71	3.015	11.065
ABUR	2002	7	Jul 12, 2002	10	4.3801	0.4508	0.10292	2290.5	31.17	4.78	2.125	14.69
AQUE	2008	11	Nov 6, 2008	15	5.57	0.65	0.116697	2126.5	30.975	4.76	1.385	22.365
AQUE	2008	12	Dec 4, 2008	15	8.15	0.65	0.079755	2068	36.255	5.285	2.175	16.68
MOHK	2002	5	May 7, 2002	10	0.509	0.0461	0.090804	2135	22.57	3.2	2.8	8.06
MOHK	2002	6	Jun 10, 2002	10	4.2919	0.4255	0.09914	2142	24.13	3.4	1.78	13.57



Mass (0 rows)

Data Integration



Additional Information

- <http://semtools.ecoinformatics.org/>
- <http://sonet.ecoinformatics.org/>
- <http://knb.ecoinformatics.org/>
- <http://www.nceas.ucsb.edu/ecoinformatics/>



Acknowledgments

This material is based upon work supported by:

- The National Science Foundation.
- Collaborators include: NCEAS (UC Santa Barbara), University of New Mexico (Long Term Ecological Research Network Office), Santa Barbara Coastal LTER, UC Davis, Gonzaga University
- The National Center for Ecological Analysis and Synthesis, a Center funded by NSF (Grant Number 0072909), the University of California, and the UC Santa Barbara campus.