

# Programming with the Semantic Web

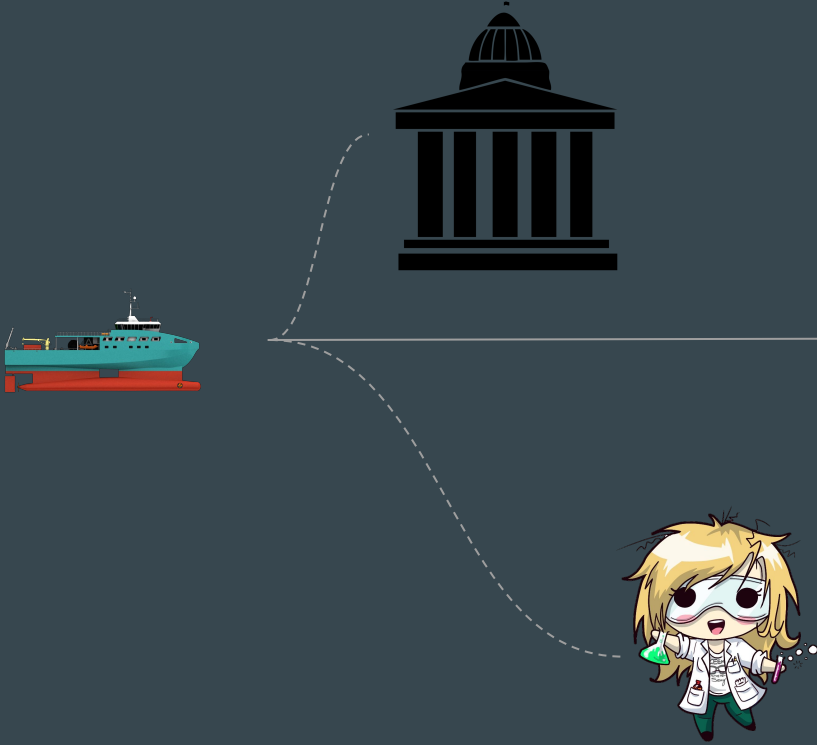
...

Adam Shepherd  
C. Chandler, R. Arko, D. Fils, D. Kinkade,  
M. Jones

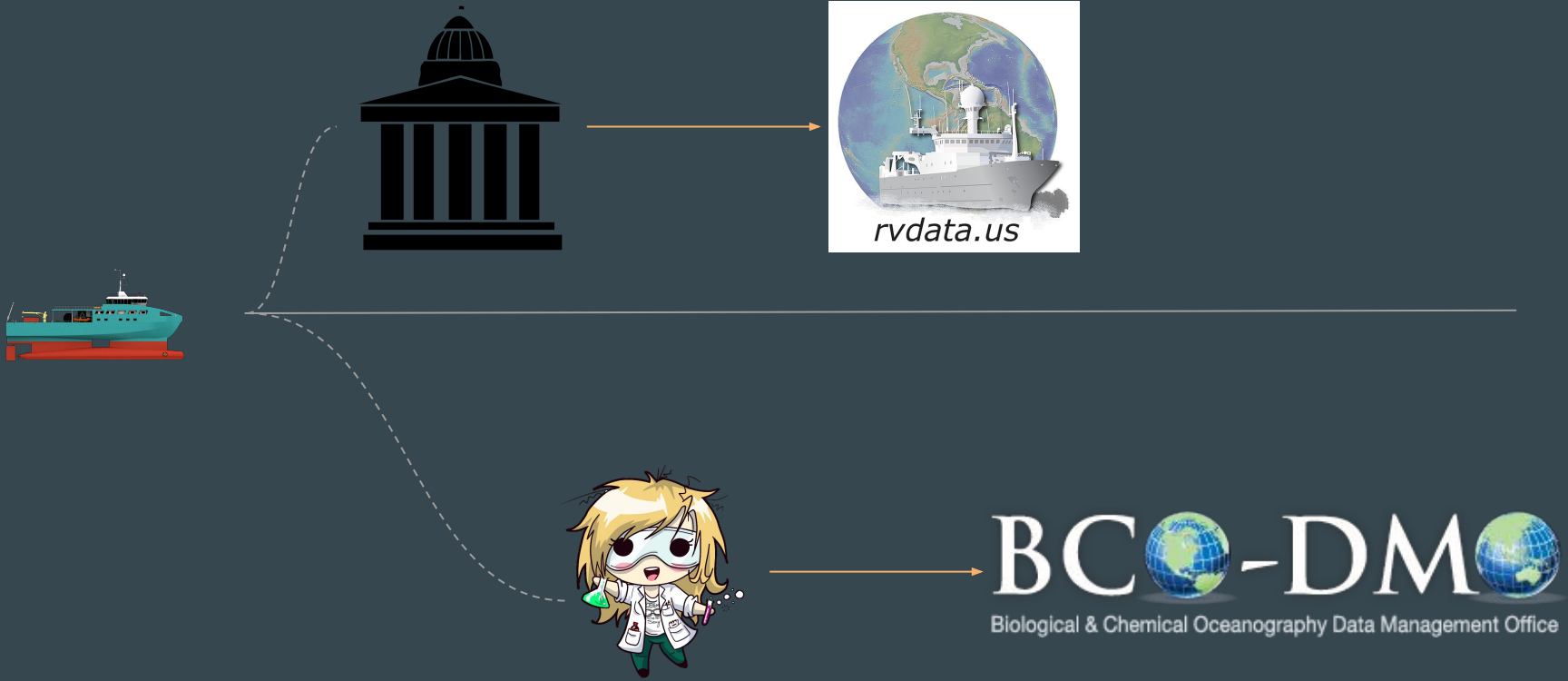
# Oceanographic Cruise: Linking to Authoritative Source



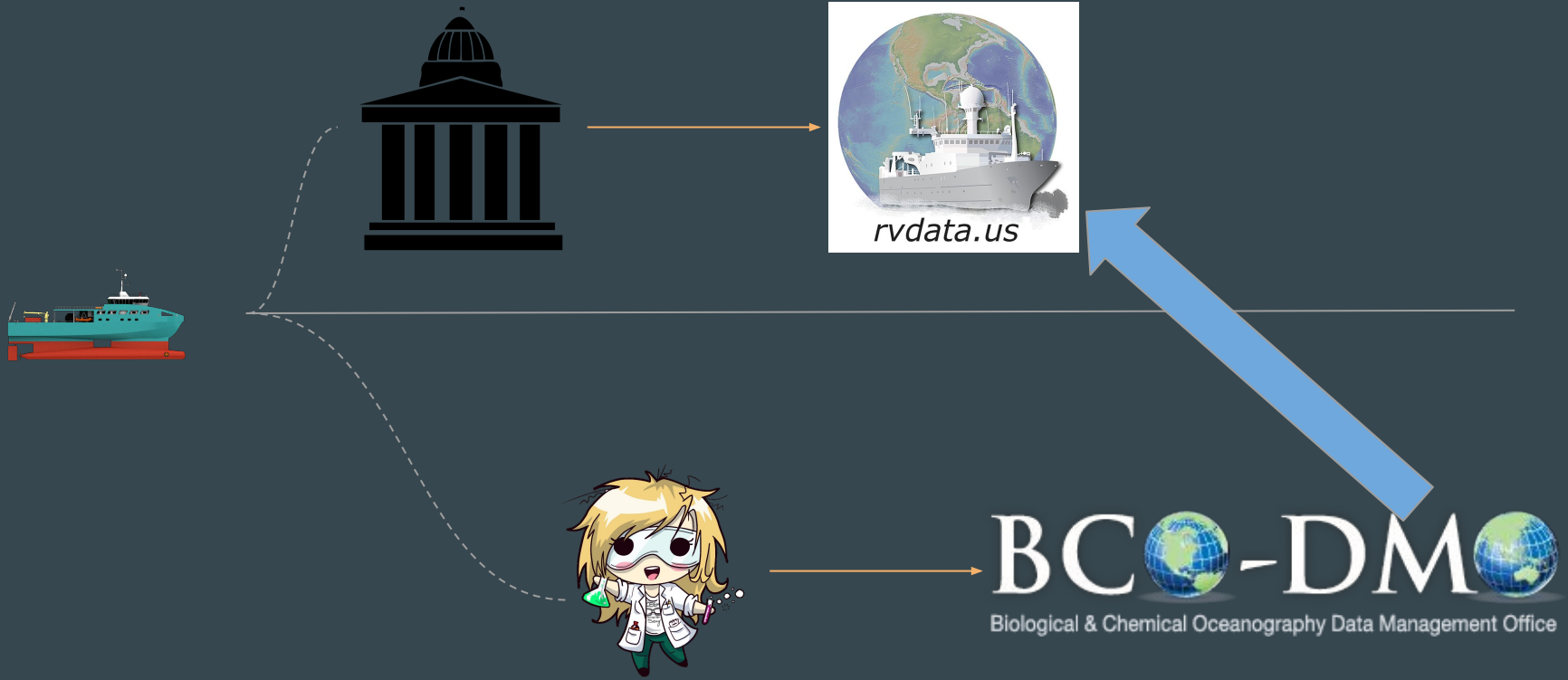
# Oceanographic Cruise: Linking to Authoritative Source



# Oceanographic Cruise: Linking to Authoritative Source



# Oceanographic Cruise: Linking to Authoritative Source





- Cruises:
  - 2,000 BCO-DMO
  - 6,400 R2R
- Data stored in two different models, local ontologies w. some shared vocabulary use (Dublin Core, etc.)



- Cruises:  
2,000 BCO-DMO  
6,400 R2R
- Data stored in two different models,  
local ontologies w. some shared  
vocabulary use (Dublin Core, etc.)



## Entity Matching Ontologies (EMO)

- Built on top of PROV-O
- Describes comparable entities  
and the process of comparing "things"  
for potential linkage
- Each type of comparison defines the  
properties to compare and their  
datatype

# Entity Matching Ontologies

## GENERAL CLASSES

MatchingRoutine

Criteria

PotentialMatchCollection

PotentialMatch

EntityProperty

EntityPropertyType

...

## CRUISE CASE EXTENSIONS

PossibleCruiseMatch (PotentialMatch)

Cruise ID

Start/End Date

Chief Scientist Name

Vessel Name

Vessel Code (ICES)

Vessel Linked Data Identifier

<https://github.com/BCODMO/Entity-Matching-Ontologies>



## Check all BCO-DMO Cruises

### CruiseID Match Threshold \*

Set the floor on matching cruiseID scores. Valid values are between 1 and 100. The cruise ID matching score is a percentage where 100% represents a perfect match and 0% not matching at all

### Calculate the combined cruise property match score

Weigh differences in cruise properties. The total of all weights should equal 100%.

**Start Date \***  25%

**End Date \***  25%

**Chief Scientist \***  25%

**Vessel Code \***  15%

**Vessel Name \***  5%

**Community URI \***  5%

Total: 100 %

### Calculate the total score

Weigh the Cruise ID score with the combined cruise property score. The total of all weights should equal 100%.

**Cruise ID Score Weight \***  25%

**Combined Property Score Weight \***  75%

Total: 100 %

### Set the match scale

For each category, set the range of scores for classification.

**Green Light \***  [default: 97 to 100 %] Confident that a match with this score is a exact match.

**Yellow Light \***  [default: 75 to 97 %] Possible that a match with this score may be the same cruise, but needs investigation.

**Red Light \***  [default: 0 to 75 %] Confident that a match with this score is not a valid match.

**Run**

- ☒ Show cruises **without** an existing match
  - ☐ Only show cruises **with** an existing match
- FILTER RESULTS**
- ☐ Cruise ID
  - ☐ Start Date
  - ☐ End Date
  - ☐ Vessel Code
  - ☐ Chief Scientist
  - ☐ Community URI

Potential Matches: 26 of 491

Cruise ▲	Matches	Best Score	Existing Match
AT11-20	13	87%	
BH05-05	1	40%	
BH05-13	1	40%	
BH05-18	1	40%	
BH13-13	17	72%	
CB1007	1	81%	
CB1009	1	81%	
FK010	1	93%	
HRS1324	3	70%	
KM1416	1	100%	
KOK1115	4	77%	
LMG0102	19	100%	
LMG0414	6	77%	
LMG1312	5	100%	
MGL1216	11	96%	
NBB10-02	1		

# LMG0102 - Match to R2R Cruise

[^ back to results listing ^](#)

[<< Previous \(KOK1205\)](#)

[Next \(LMG0103\) >>](#)

BCO-DMO Cruise: LMG0102		
LMG0102	score: 100.00%	
LMG0105	score: 74.07%	
LMG0106	score: 74.07%	
LMG0002	score: 73.93%	
LMG0802	score: 73.30%	
LMG0902	score: 73.30%	
LMG0104	score: 72.86%	
LMG0502	score: 72.32%	
LMG1102	score: 72.32%	
LMG0103	score: 72.25%	
LMG0602	score: 72.05%	
LMG0402	score: 71.88%	
LMG0109	score: 70.32%	
LMG0101	score: 70.18%	
LMG0107	score: 70.18%	
LMG0302	score: 68.63%	
LMG0108	score: 68.30%	
LMG0702	score: 66.43%	
LMG0202	score: 64.55%	

## LMG0102 - Match to R2R Cruise

[^ back to results listing ^](#)

[<< Previous \(KOK1205\)](#)

[Next \(LMG0103\) >>](#)

### BCO-DMO Cruise: LMG0102

LMG0102 score: 100.00% 

#### Create Match

##### Match Type \*

- Select a value -

##### Match Notes

##### Match Differences

Create Match

R2R Cruise Identifier: <http://data.rvdata.us/id/cruise/LMG0102>

BCO-DMO Cruise Identifier: <http://lod.bco-dmo.org/id/deployment/671723>(view in OSPREY)

Property	BCO-DMO	R2R	Score
Cruise ID	LMG0102	LMG0102	100.00%
Start Date	2001-02-20	2001-02-20	100.00%
End Date	2001-03-14	2001-03-14	100.00%
Chief Scientist	Smith, Craig	Smith, Craig	100.00%
Vessel Name	Laurence M. Gould	Laurence M. Gould	100.00%
Vessel Code	33LG	33LG	100.00%
Community URI	<a href="http://vocab.nerc.ac.uk/collection/C17/current/33LG/">http://vocab.nerc.ac.uk/collection/C17/current/33LG/</a>	<a href="http://vocab.nerc.ac.uk/collection/C17/current/33LG/">http://vocab.nerc.ac.uk/collection/C17/current/33LG/</a>	100.00%


## LMG0102 - Match to R2R Cruise

[^ back to results listing ^](#)

[<< Previous \(KOK1205\)](#)

[Next \(LMG0103\) >>](#)

### BCO-DMO Cruise: LMG0102

LMG0102 score: 100.00% 

#### Create Match

##### Match Type \*

✓ - Select a value -

rdfs:seeAlso - Use if two things are related but not really the same thing.

odo:matches - Use if something is similar but not exactly the same.

owl:sameAs - Two things are exactly the same and should inherit its properties. !Use with caution!


#### Match Differences

Create Match


R2R Cruise Identifier: <http://data.rvdata.us/id/cruise/LMG0102>

BCO-DMO Cruise Identifier: [http://lod.bco-dmo.org/id/deployment/671723\(view in OSPREDY\)](http://lod.bco-dmo.org/id/deployment/671723(view%20in%20OSPREDY))

Property	BCO-DMO	R2R	Score
Cruise ID	LMG0102	LMG0102	100.00%
Start Date	2001-02-20	2001-02-20	100.00%
End Date	2001-03-14	2001-03-14	100.00%
Chief Scientist	Smith, Craig	Smith, Craig	100.00%
Vessel Name	Laurence M. Gould	Laurence M. Gould	100.00%
Vessel Code	33LG	33LG	100.00%
Community URI	<a href="http://vocab.nerc.ac.uk/collection/C17/current/33LG/">http://vocab.nerc.ac.uk/collection/C17/current/33LG/</a>	<a href="http://vocab.nerc.ac.uk/collection/C17/current/33LG/">http://vocab.nerc.ac.uk/collection/C17/current/33LG/</a>	100.00%

LMG0105 score: 74.07% 

LMG0106 score: 74.07% 

LMG0002 score: 73.93% 

Shout out to  
Cyndy  
Chandler

## AT11-20 - Match to R2R Cruise

[^ back to results listing ^](#)

[<< Previous \(AT11-17\)](#)

[Next \(AT11-30\) >>](#)

### BCO-DMO Cruise: AT11-20

, [AT11-20](#) *score: 87.11%* 

, [AT11-21](#) *score: 74.87%* 

, [AT11-28](#) *score: 68.54%* 

, [AT11-30](#) *score: 67.37%* 

, [AT11-10](#) *score: 66.90%* 

, [AT15-20](#) *score: 65.22%* 

, [AT11-22](#) *score: 64.55%* 

, [AT11-23](#) *score: 64.32%* 

, [AT11-26](#) *score: 63.62%* 

, [AT11-27](#) *score: 62.44%* 

, [AT11-24](#) *score: 61.85%* 

, [AT11-25](#) *score: 60.80%* 

, [AT11-29](#) *score: 58.93%* 

# AT11-20 - Match to R2R Cruise

[^ back to results listing ^](#)

[<< Previous \(AT11-17\)](#)

[Next \(AT11-30\) >>](#)

BCO-DMO Cruise: AT11-20

AT11-20 **score: 87.11%** 

Create Match

Match Type \*

- Select a value - 

Match Notes


[Match Differences](#)


Create Match


R2R Cruise Identifier: <http://data.rvdata.us/id/cruise/AT11-20>


BCO-DMO Cruise Identifier: [http://lod.bco-dmo.org/id/deployment/654275\(view in OSPREY\)](http://lod.bco-dmo.org/id/deployment/654275(view%20in%20OSPREY))

Property	BCO-DMO	R2R	Score
Cruise ID	AT11-20	AT11-20	100.00%
Start Date	2004-11-05	2004-11-05	100.00%
End Date	2004-11-26	2004-11-26	100.00%
Chief Scientist	Edwards, Katrina	Lilley, Marvin	31.25%
Vessel Name	Atlantis	Atlantis	100.00%
Vessel Code	33AT	33AT	100.00%
Community URI	<a href="http://vocab.nerc.ac.uk/collection/C17/current/33AT">http://vocab.nerc.ac.uk/collection/C17/current/33AT</a>	<a href="http://vocab.nerc.ac.uk/collection/C17/current/33AT/">http://vocab.nerc.ac.uk/collection/C17/current/33AT/</a>	100.00%

AT11-21 **score: 74.87%** 

AT11-28 **score: 68.54%** 

AT11-30 **score: 67.37%** 

AT11-10 **score: 66.90%** 

AT11-20 **score: 65.22%** 

# Keep track of provenance: PROV-O

For each asserted match,

- Keep a copies of **original Cruise RDF** at the time of match (BCO-DMO & R2R)
- Keep copies of the **transformed cruises** into the EMO ontology
- Keep a copy of the **user specified criterion** for matching routine
- Keep track of **algorithm and all scores** (property matches & aggregations)
- Keep track of **user, software agents & activities** from conversion to link assertion

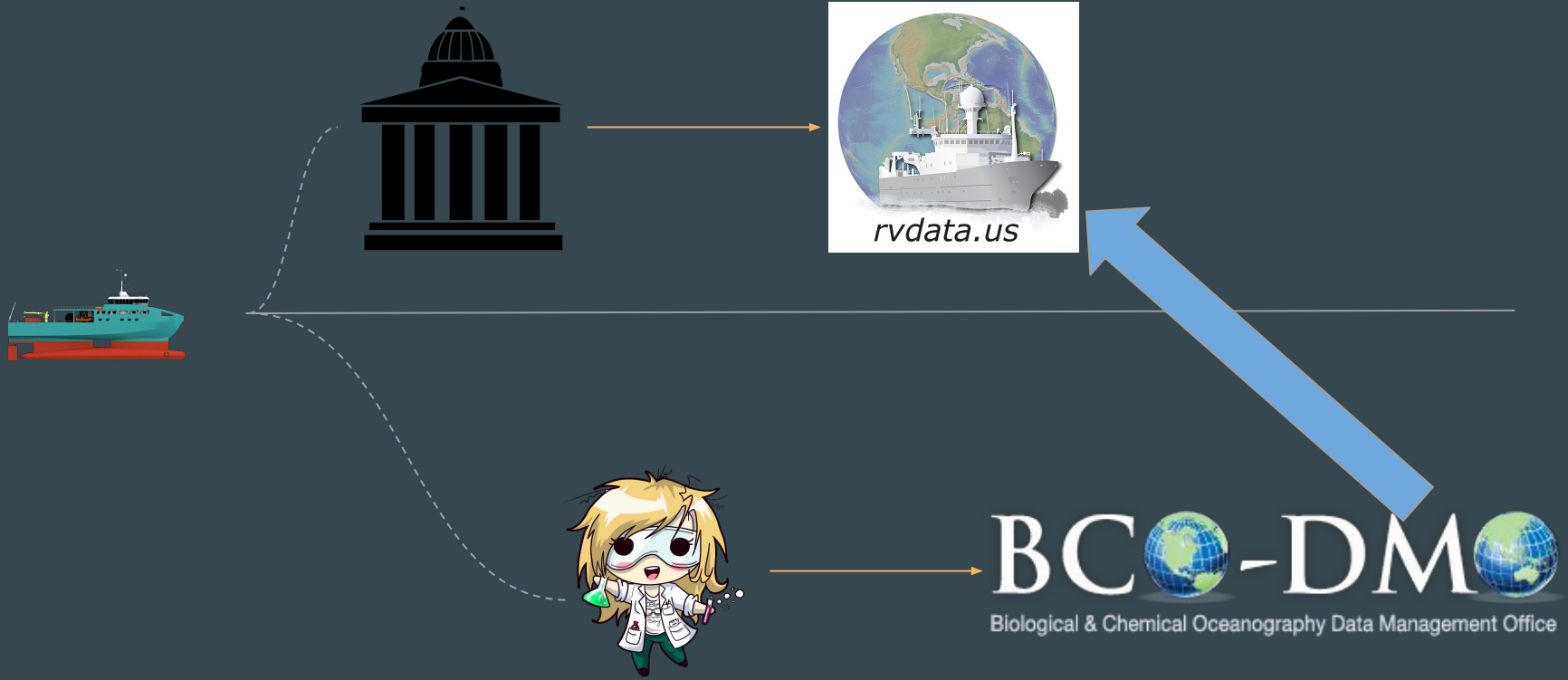


AE1314	R2R Cruise: AE1314		100.00%	owl:sameAs	<a href="#">delete</a>
AE1319	R2R Cruise: AE1319		100.00%	owl:sameAs	<a href="#">delete</a>
AE1322	AE1322: R2R Cruise: AE1322		100.00%	owl:sameAs	<a href="#">delete</a>
AE1409	AE1409: R2R Cruise: AE1409	End Date	98.13%	odo:matches	<a href="#">delete</a>
AE1410	AE1410: R2R Cruise: AE1410	Start Date, End Date	94.38%	odo:matches	<a href="#">delete</a>
AE1505	AE1505: R2R Cruise: AE1505		100.00%	owl:sameAs	<a href="#">delete</a>
AE1516	AE1516: R2R Cruise: AE1516		100.00%	owl:sameAs	<a href="#">delete</a>
AT11-04	AT11-04: R2R Cruise: AT11-04	Chief Scientist	94.08%	odo:matches	<a href="#">delete</a>
AT11-07	AT11-07: R2R Cruise: AT11-07		100.00%	owl:sameAs	<a href="#">delete</a>
AT11-17	R2R Cruise: AT11-17		100.00%	owl:sameAs	<a href="#">delete</a>
AT11-30	R2R Cruise: AT11-30	Start Date	98.13%	odo:matches	<a href="#">delete</a>
AT15-06	AT15-06: R2R Cruise: AT15-06		98.83%	owl:sameAs	<a href="#">delete</a>
AT15-25	R2R Cruise: AT15-25	Chief Scientist	85.94%	odo:matches	<a href="#">delete</a>
AT15-35	AT15-35: R2R Cruise: AT15-35		100.00%	owl:sameAs	<a href="#">delete</a>
AT15-38	AT15-38: R2R Cruise: AT15-38		100.00%	owl:sameAs	<a href="#">delete</a>
AT15-40	R2R Cruise: AT15-40		100.00%	owl:sameAs	<a href="#">delete</a>

# GeoLink: Federated Linked Data & Semantics

**What if** a federation of data facilities agreed to publish their metadata as Linked Data using a common vocabulary?

# GeoLink: Leveraging Cruise Links for Broader Discovery



## Deployment: AT11-07

Map It

**Deployment:** AT11-07

**Chief Scientist:** [Dr Hans Schouten](#) (Woods Hole Oceanographic Institution, WHOI)

**Platform:** [R/V Atlantis](#)

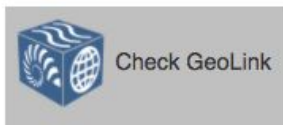
**Platform Type:** vessel

**Start Date:** 01/28/2004

**End Date:** 02/24/2004

**Location:** East Pacific Rise, Pacific

### ▾ [GeoLink](#)



### ▾ [Datasets](#)

Data from this Deployment only	Full Dataset (multiple deployments)
<a href="#">Assembled metagenome sequences &gt; 500 bp (AT11-07)</a>	<a href="#">Assembled metagenome sequences &gt; 500 bp</a>

Location: East Pacific Rise, Pacific

GeoLink



GeoLink Resource Information for: **AT11-07**

Cruise Information

Cruise: AT11-07  
Identifier: [doi:10.7284/900008](https://doi.org/10.7284/900008)  
Start Port: San Diego  
End Port: Puntarenas

Cruise DOI  
(via R2R)

Hard rock geology  
(via SESAR)

Physical Samples (441)

Sample	Description	People	Dataset	Dates
1. <a href="#">IGSN:MGD000AVU</a>	Elevation: -2512 Meters Feature of Interest: <a href="#">EPR:9N</a> Feature Type: <a href="#">SpreadingCenterSegment</a> Geometry: POINT(-104.291693 9.840713)	Edwards, Katrina Collector	<a href="#">AT11-07,</a> <a href="#">Sampler&gt;IncubationChamber</a>	Collected: 2004-02-01 to Registered: 2012-03-11 Published: 2012-03-10
2. <a href="#">IGSN:MGD000B0A</a>	Elevation: -2512 Meters Feature of Interest: <a href="#">EPR:9N</a> Feature Type: <a href="#">SpreadingCenterSegment</a>	Edwards, Katrina Collector	<a href="#">AT11-07,</a> <a href="#">Sampler&gt;IncubationChamber</a>	Collected: 2004-02-01 to Registered: 2012-03-11 Published: 2012-03-10

# GeoLink Widget: Web Component

```
<gldeployment-element  
  params={'"host": "GeoLink",  
          "hosturl": "http://geolink.org",  
          "url": "http://lod.bco-dmo.org/geolink/id/deployment/616332",  
          "label": "AT11-07"}'  
/>
```

Shout out to Doug Fils

# GeoLink Widget: Web Component using PIDs

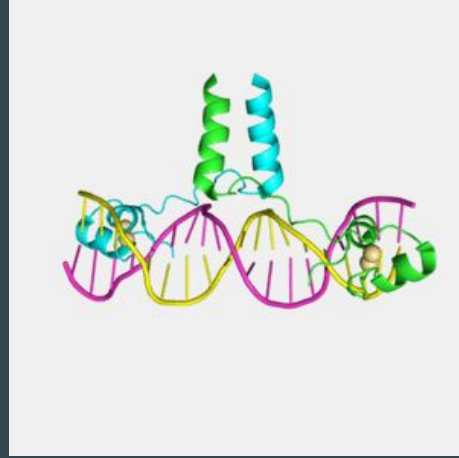
```
<gldeployment-element  
  params={'"host": "GeoLink",  
          "hosturl": "http://geolink.org",  
          "identifier-scheme": "http://purl.org/spar/datacite/doi",  
          "identifier-value": "10.7284/900008",  
          "label": "AT11-07"}  
>
```

# Ocean Protein Portal: Defining a Dataset Profile

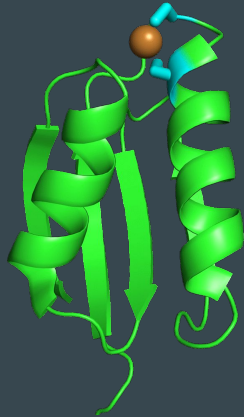
*Hemoglobin*



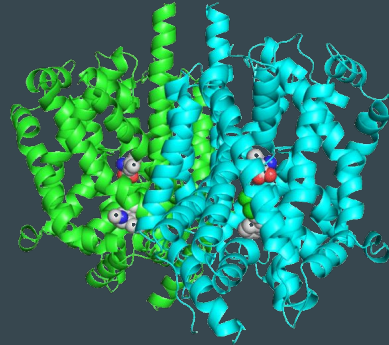
*DNA binding  
protein*



*Atx1 Cu chaperone*



*Leucine Transporter*





# Ocean Protein Portal: Searching for Proteins

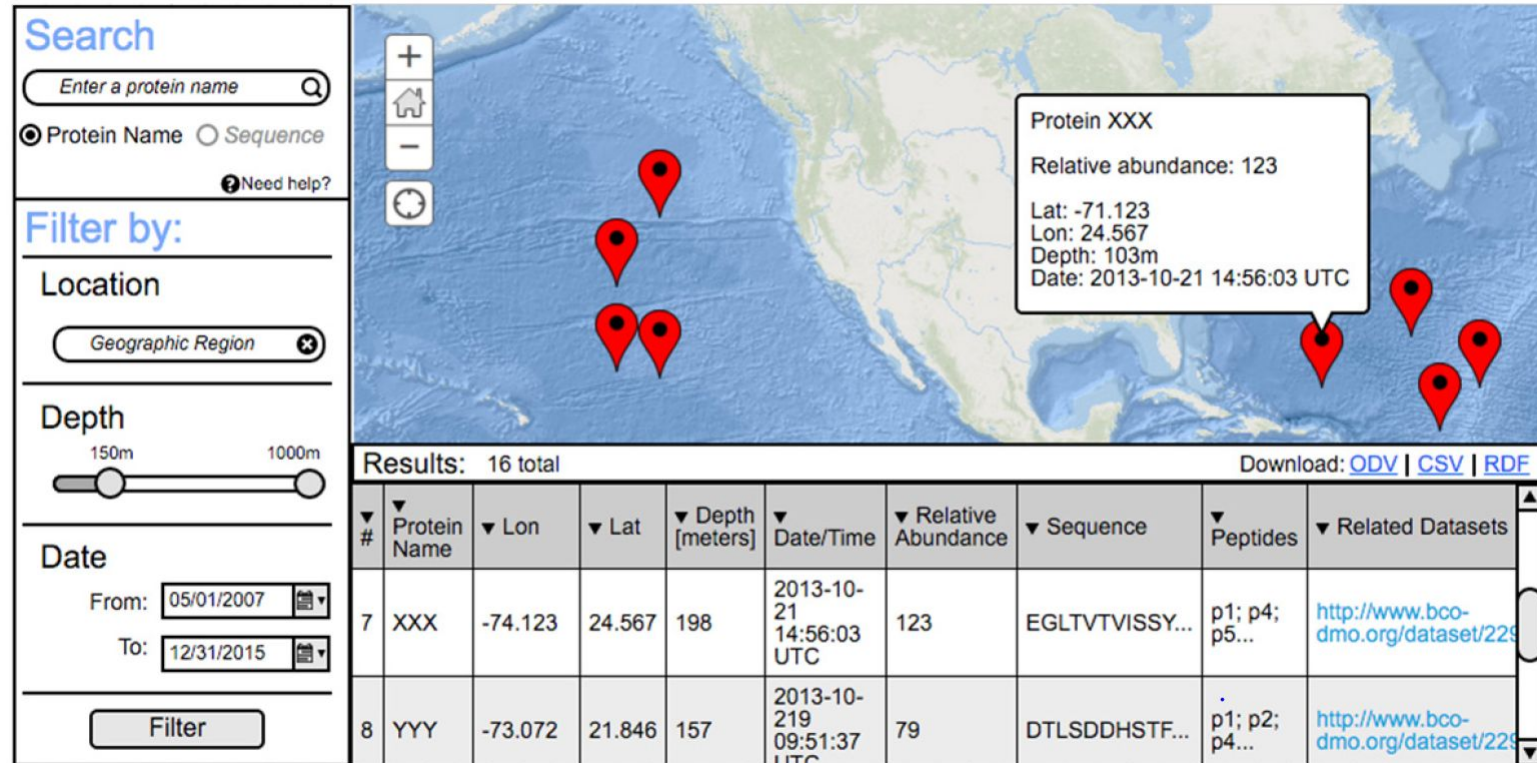


Figure 3. A mockup of the Ocean Protein Portal user interface.

Q1: Where and when does a protein occur?

Text search

Sequence search

In silico digestion

Peptides

Query against precomputed datasets

Q2: Whose protein/peptide is this?

METATryp

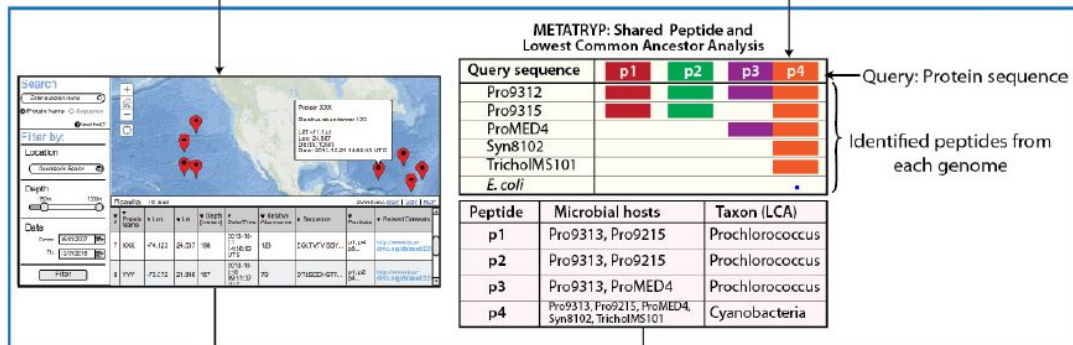
Microbial genomes

Metagenomes

Curated genomic database

Sample #1

Protein description	Accession	Abundance	Protein	Protein description
Alkaline phosphatase	Prochlorococcus marinus MIT 9312	2.5	GLUTAMYLASE	Prochlorococcus marinus MIT 9312
Alkaline phosphatase	Prochlorococcus marinus MIT 9315	1.7	GLUTAMYLASE	Prochlorococcus marinus MIT 9315
Alkaline phosphatase	Prochlorococcus marinus MIT 9315	1.7	GLUTAMYLASE	Prochlorococcus marinus MIT 9315
Alkaline phosphatase	Prochlorococcus marinus MIT 9315	1.7	GLUTAMYLASE	Prochlorococcus marinus MIT 9315
Alkaline phosphatase	Prochlorococcus marinus MIT 9315	1.7	GLUTAMYLASE	Prochlorococcus marinus MIT 9315
Alkaline phosphatase	Prochlorococcus marinus MIT 9315	1.7	GLUTAMYLASE	Prochlorococcus marinus MIT 9315
Alkaline phosphatase	Prochlorococcus marinus MIT 9315	1.7	GLUTAMYLASE	Prochlorococcus marinus MIT 9315
Alkaline phosphatase	Prochlorococcus marinus MIT 9315	1.7	GLUTAMYLASE	Prochlorococcus marinus MIT 9315
Alkaline phosphatase	Prochlorococcus marinus MIT 9315	1.7	GLUTAMYLASE	Prochlorococcus marinus MIT 9315
Alkaline phosphatase	Prochlorococcus marinus MIT 9315	1.7	GLUTAMYLASE	Prochlorococcus marinus MIT 9315



Export as ODV compatible text file

Export text file list of METATryp output

# Dataset Profile

Measurement	Units	Cardinality	Data Type
Latitude	decimal degrees	one	float
Longitude	decimal degrees	one	float
Depth	meters	one	float
Protein Name		one	string
Phylogenetic Name		zero to many	string
Spectral Count	unitless	one	float

# Ocean Protein Portal: Capital 'P' Proposed Metadata

- JSON-LD metadata file - (schema.org/Dataset + measurements)

```
"@type": "http://datapfiles.org/id/type/OceanProteinPortal",  
"measurements": [  
  {  
    "depth": "http://datapfiles.org/id/type/OceanProteinPortal/depth",  
    "units": "http://datapfiles.org/id/units/meters",  
  },  
  ...other measurements defined here...  
],  
"distribution": {  
  "encodingFormat" : "text/tab-separated-values",  
  "contentUrl" : "https://www.bco-dmo.org/dataset/472032/data/download",  
  "datePublished" : "2016-12-05"  
}
```