SeaDataCloud

SeaDataNet Vocabularies and Linked Data principles applied to SeaDataNet directories

Mark Hebden (BODC)

SeaDataCloud Mid-Term Review, Brussels, 6th December 2018
sdn-userdesk@seadatanet.org – www.seadatanet.org

Common vocabulary developments

- Collective aim of progressing the web of common controlled vocabularies underpinning the SeaDataNet infrastructure
- Adoption of prescribed, standardised terms **removes ambiguity** and encourages **greater interoperability** (including machine-to-machine exchange of information)
- SeaDataNet vocabularies provided by NERC Vocabulary Server (NVS), managed by BODC.

Following slides will address:

i) Improving user experience for adding and searching for parameter codes

ii) Enhancements to the vocab system, to improve robustness/resilience

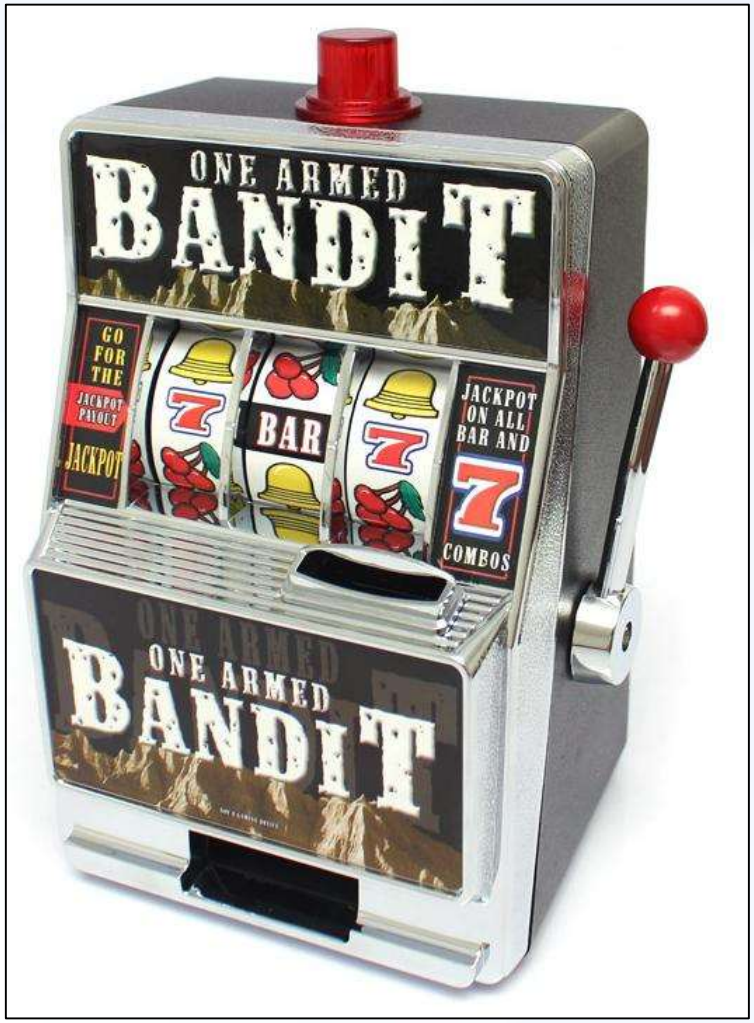
iii) Increase in vocabulary content.

Operationalising the Vocabulary Builder

- Exposing the semantic building blocks of key NVS Vocabulary Collections (e.g. P01) enables i) **improved search capability** and ii) vocabulary **content growth**, driven by community

- Prototype Vocabulary Builder developed under EMODnet Chemistry, exposing concepts under P01 chemical semantic model

- SeaDataCloud has enabled expansion to include biological and physical semantic models, following extensive re-modelling of underlying infrastructure.



- P01 is a **compound vocabulary**
- Concept labels are created from concatenation of concept labels from other vocabularies…

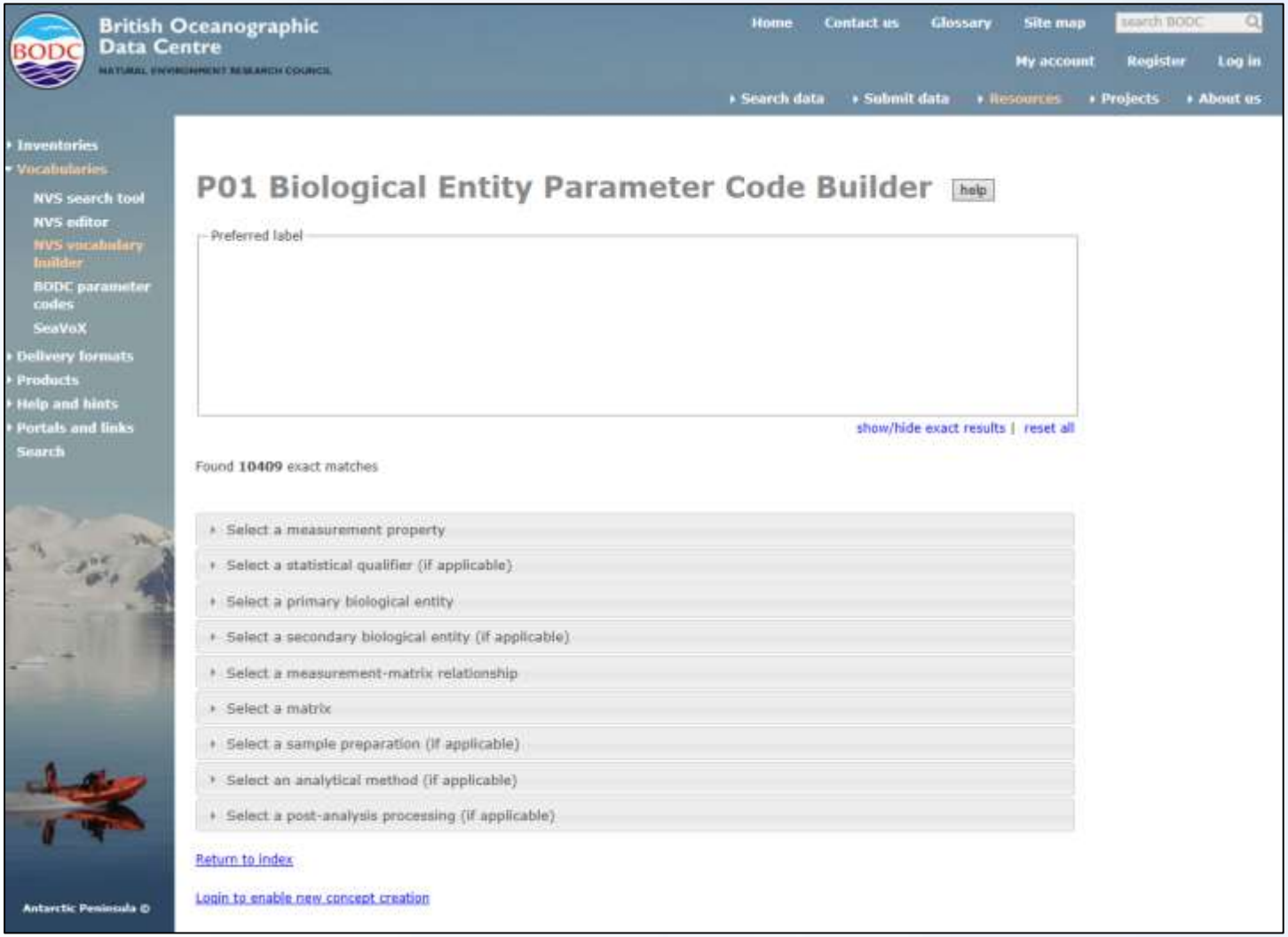
Measurement Property

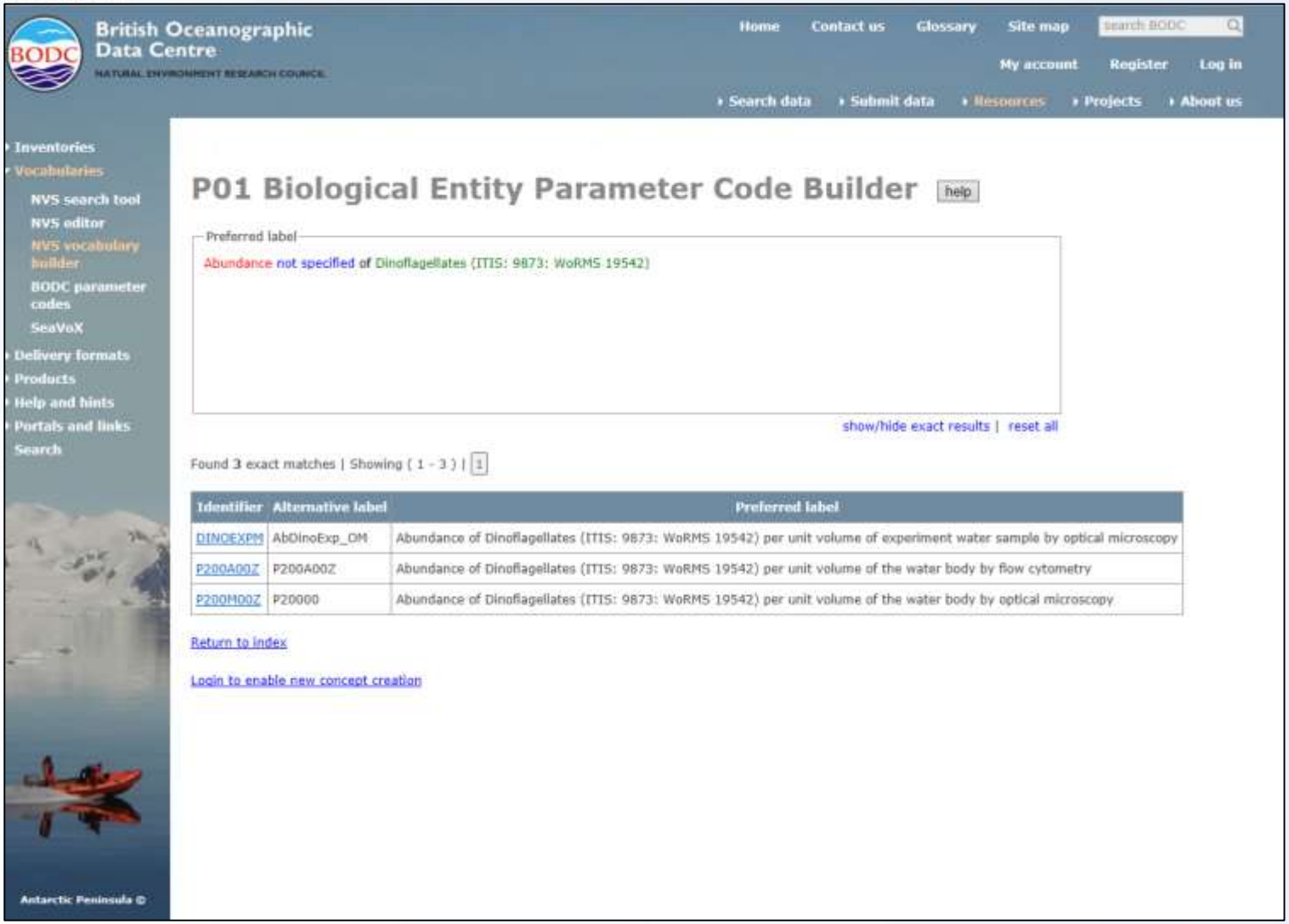
Chemical Substance

Measurement Matrix Relationship

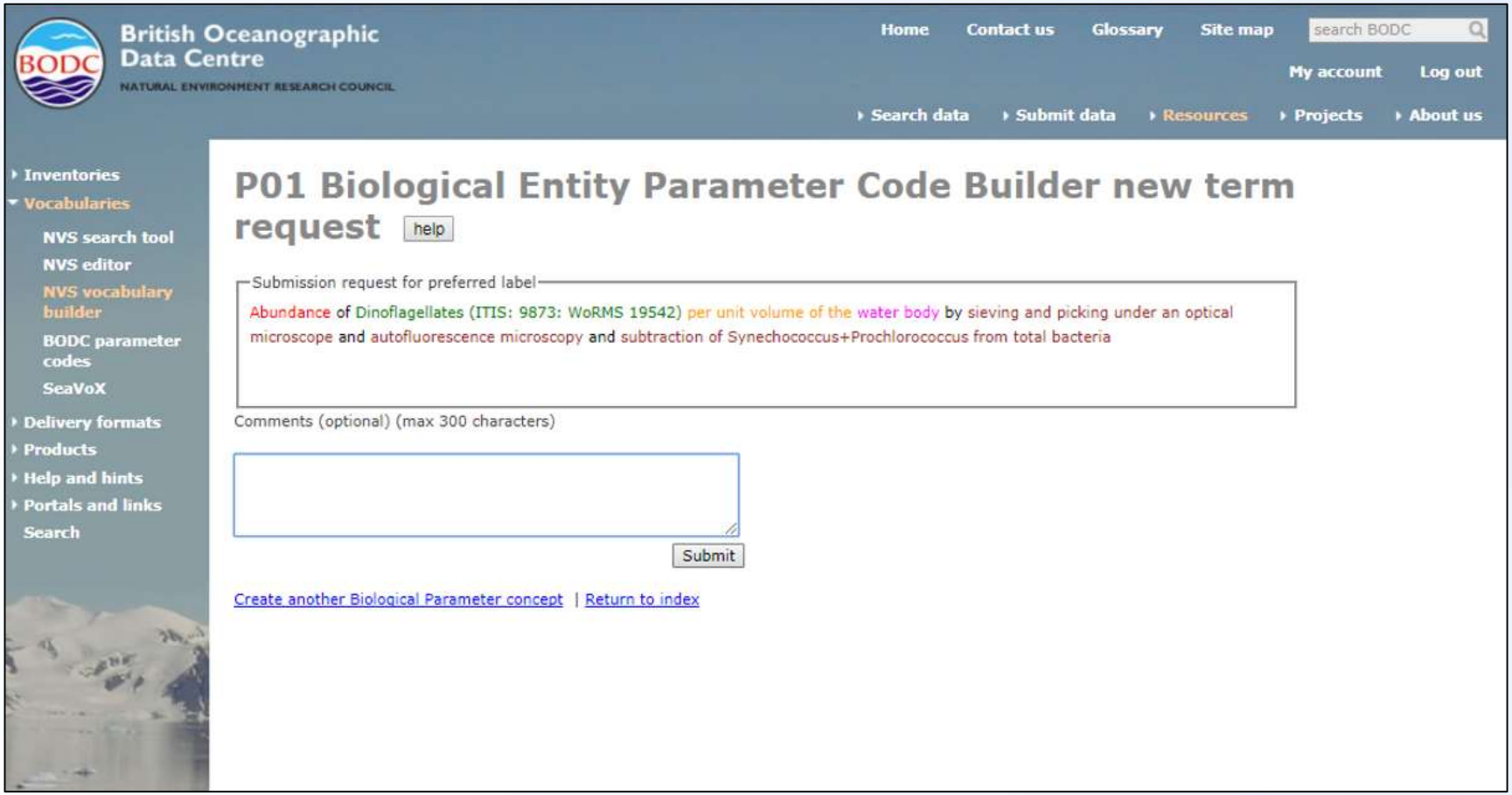
Matrix

For example:
Activity of americium-241 {241Am CAS 14596-10-2 } per unit dry weight of sediment


British Oceanographic Data Centre
NATURAL ENVIRONMENT RESEARCH COUNCIL
Home
Contact us
Glossary
Site map
search BODC
My account
Register
Log in
Search data
Submit data
Resources
Projects
About us
Inventories
Vocabularies
NVS search tool
NVS editor
NVS vocabulary builder
BODC parameter codes
SeaVoX
Delivery formats
Products
Help and hints
Portals and links
Search
Antarctic Peninsula ©
P01 Biological Entity Parameter Code Builder
help
Preferred label
show/hide exact results | reset all
Found 10409 exact matches
Select a measurement property
Select a statistical qualifier (if applicable)
Select a primary biological entity
Select a secondary biological entity (if applicable)
Select a measurement-matrix relationship
Select a matrix
Select a sample preparation (if applicable)
Select an analytical method (if applicable)
Select a post-analysis processing (if applicable)
Return to index
Login to enable new concept creation

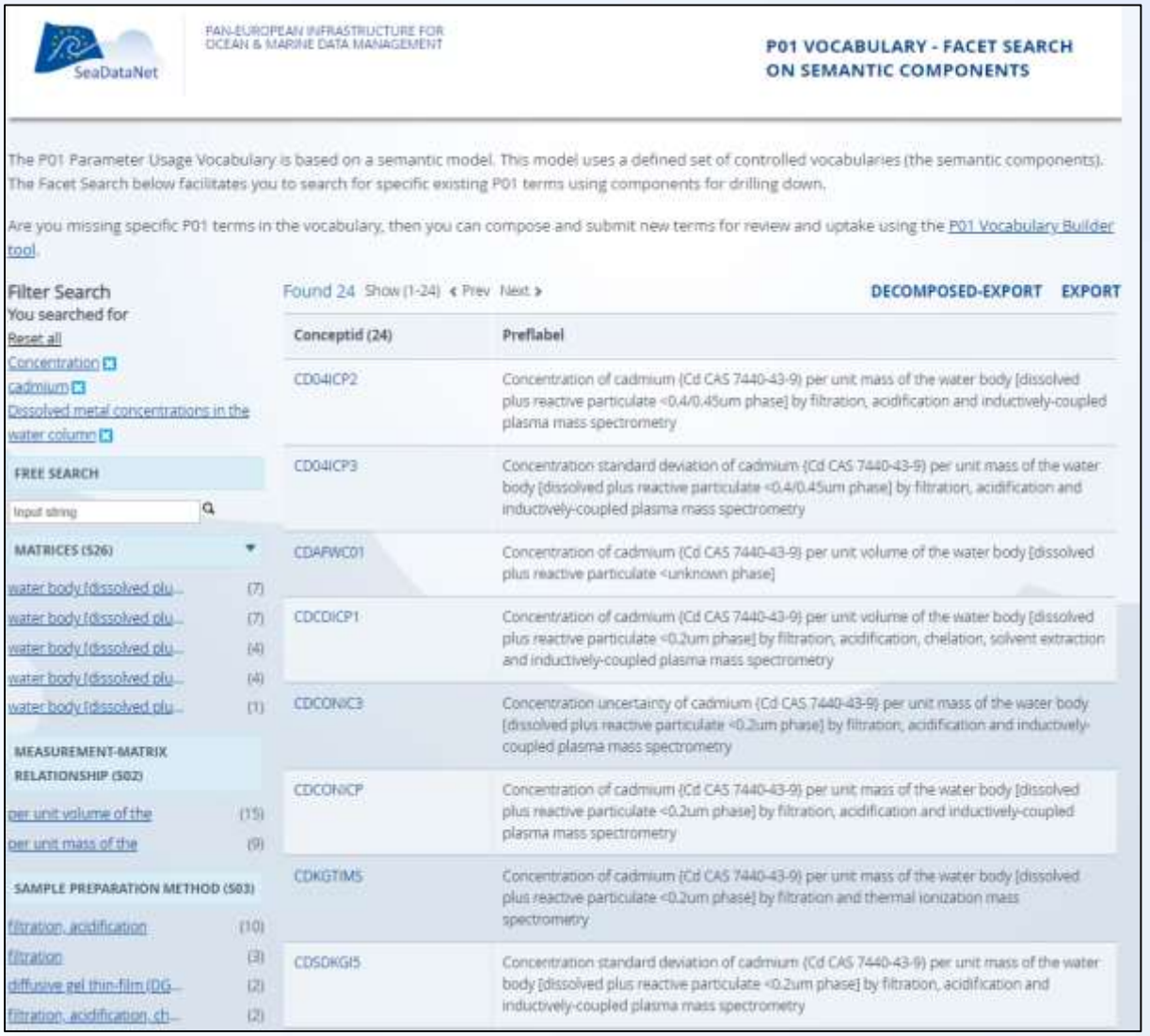

British Oceanographic Data Centre
NATURAL ENVIRONMENT RESEARCH COUNCIL
Home
Contact us
Glossary
Site map
search BODC
My account
Register
Log in
Search data
Submit data
Resources
Projects
About us
Inventories
Vocabularies
NVS search tool
NVS editor
NVS vocabulary builder
BODC parameter codes
SeaVoX
Delivery formats
Products
Help and hints
Portals and links
Search
Antarctic Peninsula ©
P01 Biological Entity Parameter Code Builder
help
Preferred label
Abundance not specified of Dinoflagellates (ITIS: 9873: WoRMS 19542)
show/hide exact results | reset all
Found 3 exact matches | Showing (1 - 3) | 1
Identifier
Alternative label
Preferred label
DINOEXPM
AbDinoExp_OM
Abundance of Dinoflagellates (ITIS: 9873: WoRMS 19542) per unit volume of experiment water sample by optical microscopy
P200A00Z
P200A00Z
Abundance of Dinoflagellates (ITIS: 9873: WoRMS 19542) per unit volume of the water body by flow cytometry
P200M00Z
P20000
Abundance of Dinoflagellates (ITIS: 9873: WoRMS 19542) per unit volume of the water body by optical microscopy
Return to index
Login to enable new concept creation


British Oceanographic Data Centre
Home
Contact us
Glossary
Site map
search BODC
My account
Register
Log in
Search data
Submit data
Resources
Projects
About us
Inventories
Vocabularies
NVS search tool
NVS editor
NVS vocabulary builder
BODC parameter codes
SeaVoX
Delivery formats
Products
Help and hints
Portals and links
Search
P01 Biological Entity Parameter Code Builder
help
Preferred label
Abundance not specified of Dinoflagellates (ITIS: 9873: WoRMS 19542) per unit volume of water body by sieving and picking under an optical microscope and autofluorescence microscopy and subtraction of Synechococcus+Prochlorococcus from total bacteria
show/hide close match results | reset all
Found 0 exact matches
Found 2 close matches
✓ Select a measurement property
✓ Select a statistical qualifier (if applicable)
✓ Select a primary biological entity
Select a secondary biological entity (if applicable)
✓ Select a measurement-matrix relationship
✓ Select a matrix
✓ Select a sample preparation (if applicable)
✓ Select an analytical method (if applicable)
✓ Select a post-analysis processing (if applicable)
Approved
Antarctic Peninsula ©


British Oceanographic Data Centre
NATURAL ENVIRONMENT RESEARCH COUNCIL
Home
Contact us
Glossary
Site map
search BODC
My account
Log out
Search data
Submit data
Resources
Projects
About us
Inventories
Vocabularies
NVS search tool
NVS editor
NVS vocabulary builder
BODC parameter codes
SeaVoX
Delivery formats
Products
Help and hints
Portals and links
Search
P01 Biological Entity Parameter Code Builder new term request
help
Submission request for preferred label
Abundance of Dinoflagellates (ITIS: 9873: WoRMS 19542) per unit volume of the water body by sieving and picking under an optical microscope and autofluorescence microscopy and subtraction of Synechococcus+Prochlorococcus from total bacteria
Comments (optional) (max 300 characters)
Submit
Create another Biological Parameter concept | Return to index

MARIS search facet:

- Alternative (and complementary) SeaDataNet-branded interface for lookup of SeaDataNet common vocabularies
- Drill down catalogue enabling lookup of terms

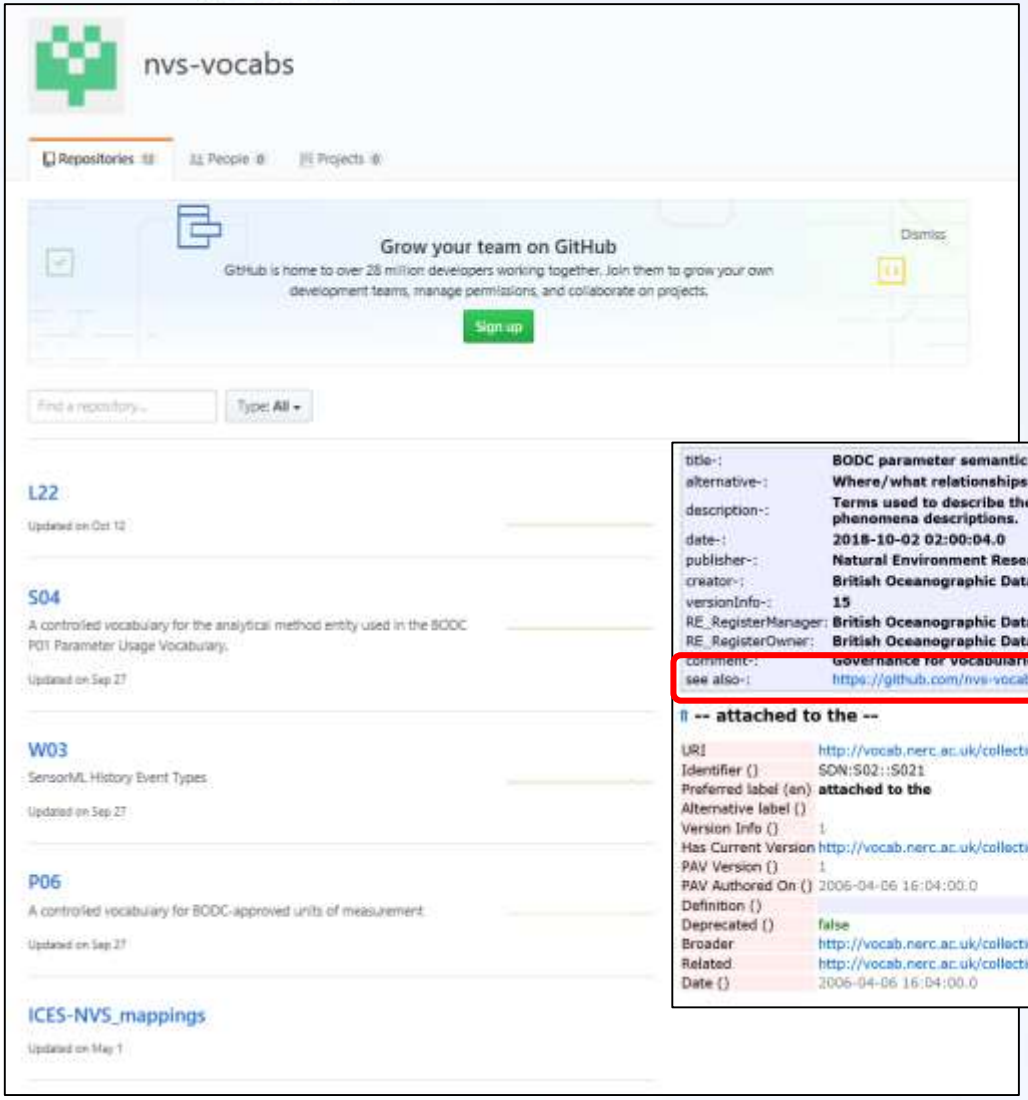


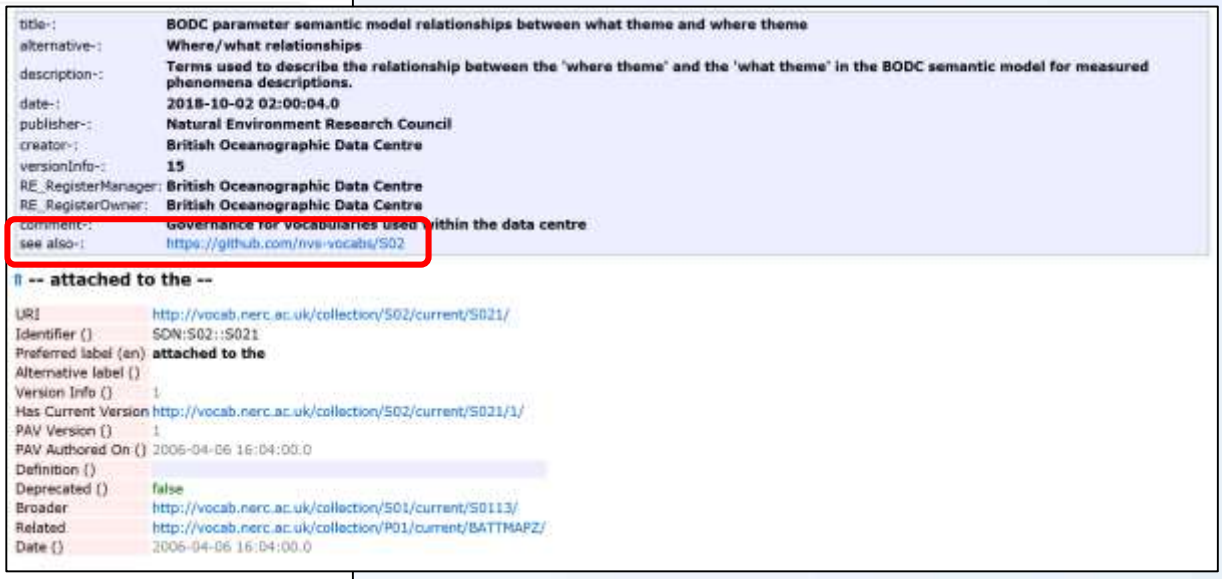
Improving the transparency of the vocabulary governance model

- Need highlighted by Research Data Alliance (RDA) Vocabularies Services Interest Group (VSIG)
- Will help ensure **ongoing confidence** in NVS vocabulary services, essential for mark up of SeaDataNet information assets

Solution:

1) Creation of publically-visible **GitHub repositories** for key NVS vocabularies, capturing discussions and decisions
2) Extension of the NVS database to formally link to these
3) Exposure of the extended NVS infrastructure by means of NVS RESTful, SOAP and SPARQL public services.


nvs-vocabs
Repositories
People
Projects
Grow your team on GitHub
GitHub is home to over 28 million developers working together. Join them to grow your own development teams, manage permissions, and collaborate on projects.
Sign up
Dismiss
Type: All
L22
S04
A controlled vocabulary for the analytical method entity used in the BODC P01 Parameter Usage Vocabulary.
W03
SensorML History Event Types
P06
A controlled vocabulary for BODC-approved units of measurement
ICES-NVS_mappings


title-: BODC parameter semantic model relationships between what theme and where theme
alternative-: Where/what relationships
description-: Terms used to describe the relationship between the 'where theme' and the 'what theme' in the BODC semantic model for measured phenomena descriptions.
date-: 2018-10-02 02:00:04.0
publisher-: Natural Environment Research Council
creator-: British Oceanographic Data Centre
versionInfo-: 15
RE_RegisterManager: British Oceanographic Data Centre
RE_RegisterOwner: British Oceanographic Data Centre
comment-: Governance for vocabularies used within the data centre
see also-: https://github.com/nvs-vocabs/S02
-- attached to the --
URI http://vocab.nerc.ac.uk/collection/S02/current/S021/
Identifier () SDN:S02::S021
Preferred label (en) attached to the
Alternative label ()
Version Info () 1
Has Current Version http://vocab.nerc.ac.uk/collection/S02/current/S021/1/
PAV Version () 1
PAV Authored On () 2006-04-06 16:04:00.0
Definition ()
Deprecated () false
Broader http://vocab.nerc.ac.uk/collection/S01/current/S0113/
Related http://vocab.nerc.ac.uk/collection/P01/current/BATTMAPZ/
Date () 2006-04-06 16:04:00.0

Vocabulary deprecation

- Establishing rules for handling deprecation in SeaDataNet
- Each concept has a URI, which must be preserved in the event of that term being deprecated
- Software needs to know if handling a deprecated term and take appropriate action

Agreed rules:

- Always having a ‘replace by’ term before deprecation is permitted
- User experience should not be compromised – automatic replacement of deprecated terms by SeaDataNet Download Manager, with report back to Data Centre to update local filestock

-- mass_concentration_of_pm10_ambient_aerosol_in_air --

URI	http://vocab.nerc.ac.uk/collection/P07/current/DIEJIDDA/
Identifier ()	SDN:P07::DIEJIDDA
Preferred label (en)	**mass_concentration_of_pm10_ambient_aerosol_in_air**
Alternative label ()	
Definition (en)	Mass concentration means mass per unit volume and is used in the construction mass_concentration_of_X_in_Y, where X is a material constituent of Y. A chemical species denoted by X may be described by a single term such as 'nitrogen' or a phrase such as 'nox_expressed_as_nitrogen'. "Aerosol" means the suspended liquid or solid particles in air (except cloud droplets). "Ambient aerosol" is aerosol that has taken up ambient water through hygroscopic growth. The extent of hygroscopic growth depends on the relative humidity and the composition of the aerosol. "Pm10 aerosol" is an air pollutant with an aerodynamic diameter of less than or equal to 10 micrometers. To specify the relative humidity and temperature at which the particle size applies, provide scalar coordinate variables with the standard names of, respectively, "relative_humidity" and "air_temperature."
Version Info ()	2
Has Current Version	http://vocab.nerc.ac.uk/collection/P07/current/DIEJIDDA/2/
Has Version	http://vocab.nerc.ac.uk/collection/P07/current/DIEJIDDA/1/
PAV Version ()	2
PAV Authored On ()	2015-01-07 15:23:09.0
Deprecated()	true
ReplacedBy	http://vocab.nerc.ac.uk/collection/P07/current/6O34XRFW/
Same as	http://mmisw.org/ont/cf/parameter/mass_concentration_of_pm10_ambient_aerosol_in_air
Related	http://vocab.nerc.ac.uk/collection/P06/current/UKMC/
Date ()	2015-01-07 15:23:09.0

- Further software upgrades may be necessary now these rules have been established.

Versioning of concepts

- Enhancement allowing user access to previous versions of individual NVS concepts
- Agreed URI model:
- http://vocab.nerc.ac.uk/collection/P07/current/CF12N86/
  - Brings the current concept version
- http://vocab.nerc.ac.uk/collection/P07/current/CF12N86/1/
  - Brings concept version 1
- http://vocab.nerc.ac.uk/collection/P07/current/CF12N86/2/
  - Brings concept version 2
- http://vocab.nerc.ac.uk/collection/P07/current/CF12N86/3/
  - Brings concept version 3 etc.

Provenance of mapping

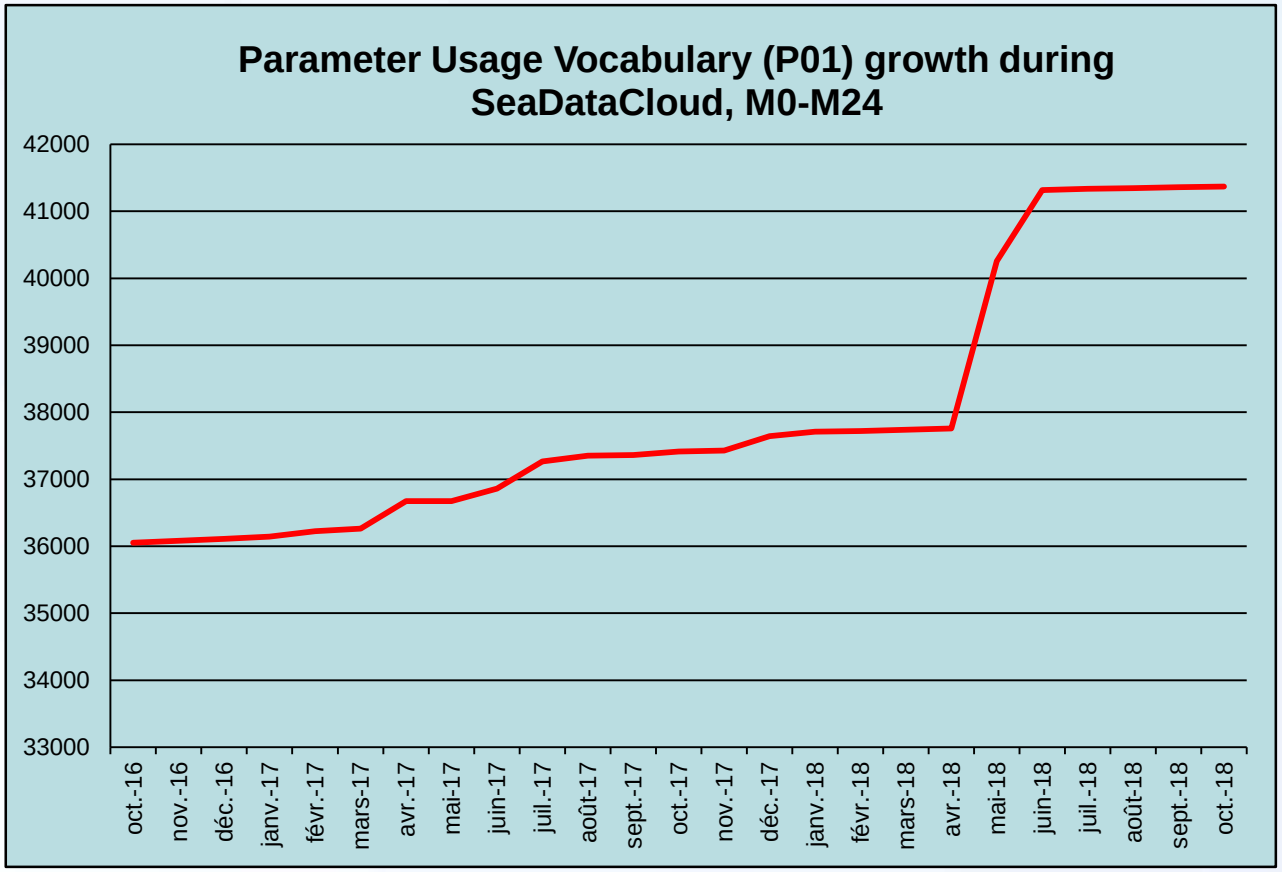
- To **build confidence** in NVS mappings requirement to store **who** produced a mapping and **when**
- Information was previously only stored internally
- Significant discussion to agree correct semantic approach, aligned with principles of Linked Data. Mappings delivered via URI:

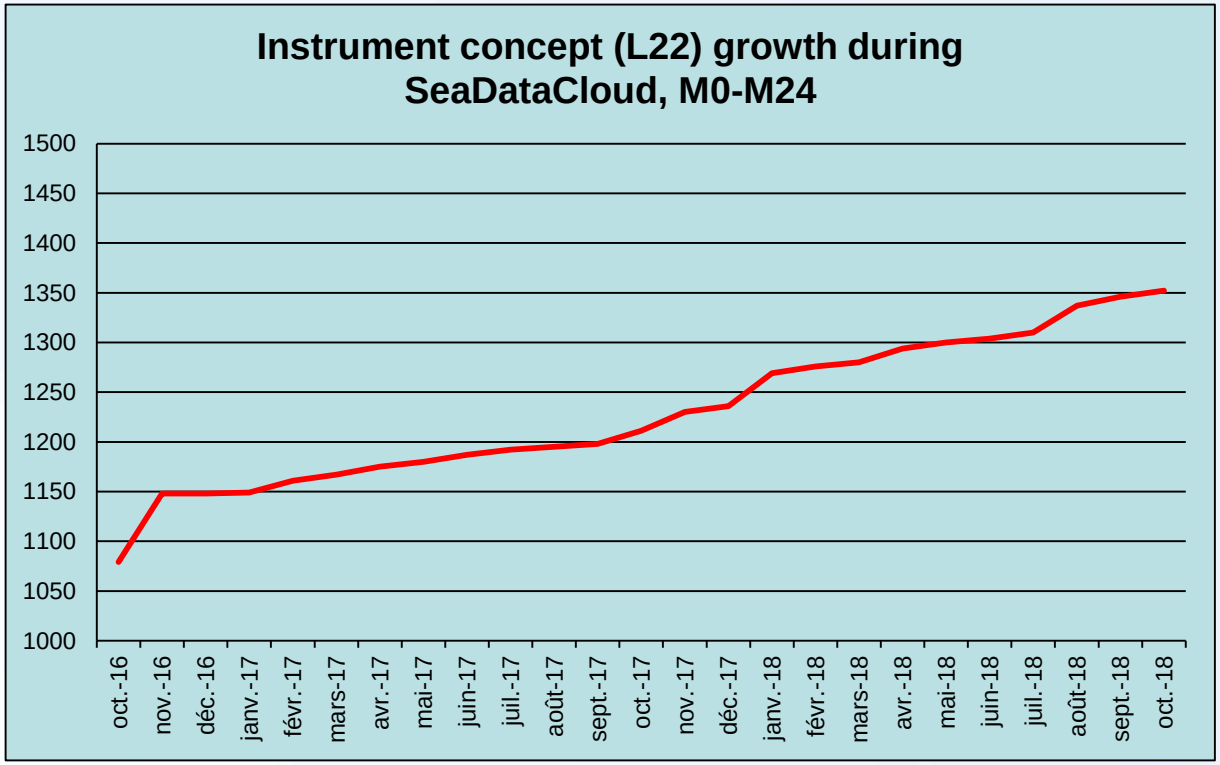
Vocabulary Content

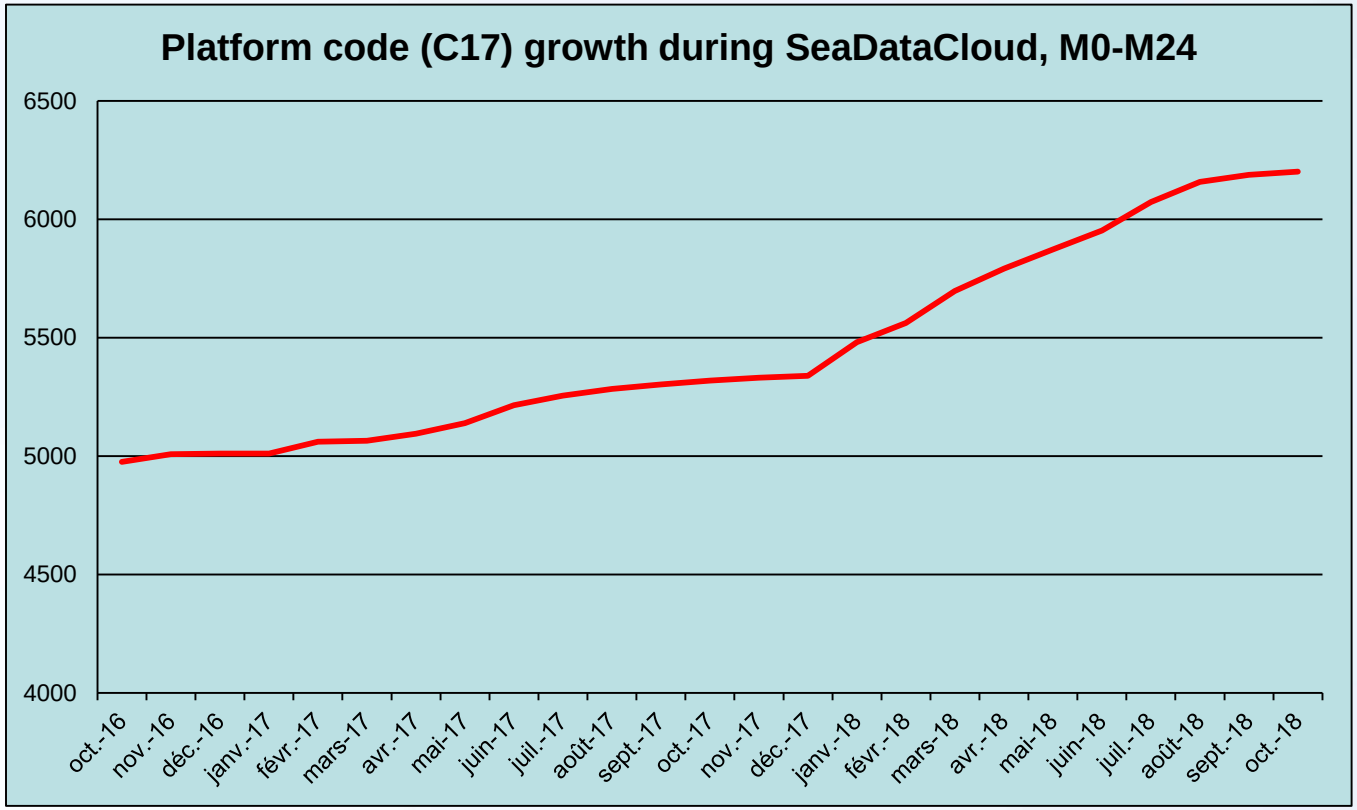
- New NVS-hosted vocabularies to support SWE Marine Profiles Group

http://vocab.nerc.ac.uk/collection/W03/current/ --> SensorML Event Types
http://vocab.nerc.ac.uk/collection/W04/current/--> SensorML Capabilities
http://vocab.nerc.ac.uk/collection/W05/current/--> SensorML Characteristics
http://vocab.nerc.ac.uk/collection/W06/current/--> SensorML Classification
http://vocab.nerc.ac.uk/collection/W07/current/--> SensorML Identification
http://vocab.nerc.ac.uk/collection/W08/current/--> SensorML Contacts

- Notable support to marine micro-litter and flow cytometry


Parameter Usage Vocabulary (P01) growth during SeaDataCloud, M0-M24
42000
41000
40000
39000
38000
37000
36000
35000
34000
33000
oct.-16
nov.-16
déc.-16
janv.-17
févr.-17
mars-17
avr.-17
mai-17
juin-17
juil.-17
août-17
sept.-17
oct.-17
nov.-17
déc.-17
janv.-18
févr.-18
mars-18
avr.-18
mai-18
juin-18
juil.-18
août-18
sept.-18
oct.-18


Instrument concept (L22) growth during SeaDataCloud, M0-M24
1500
1450
1400
1350
1300
1250
1200
1150
1100
1050
1000
oct.-16
nov.-16
déc.-16
janv.-17
févr.-17
mars-17
avr.-17
mai-17
juin-17
juil.-17
août-17
sept.-17
oct.-17
nov.-17
déc.-17
janv.-18
févr.-18
mars-18
avr.-18
mai-18
juin-18
juil.-18
août-18
sept.-18
oct.-18

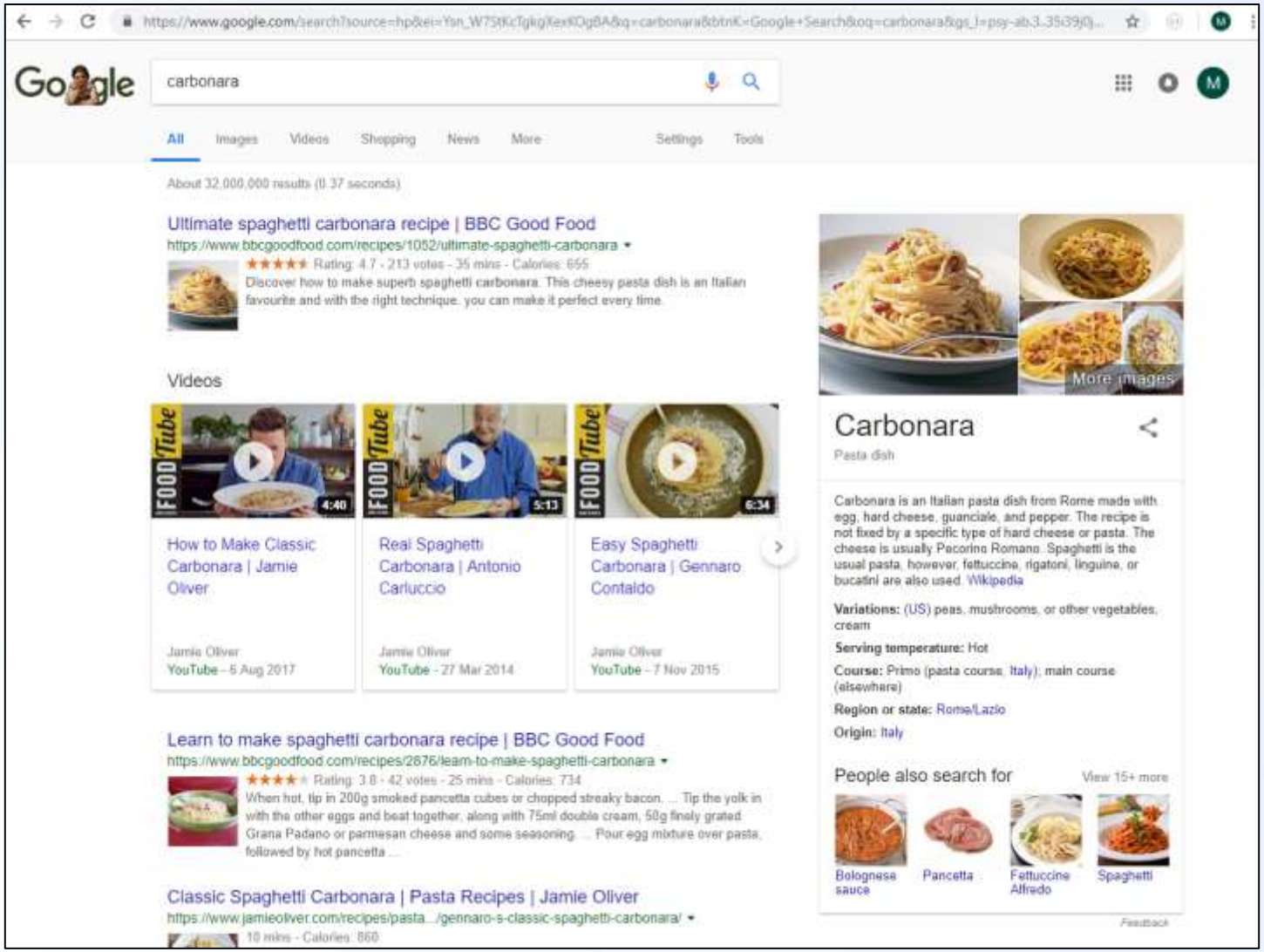


- Ongoing collaboration between ICES and BODC to enhance C17 content
- Management of underwater glider platform class now with BODC

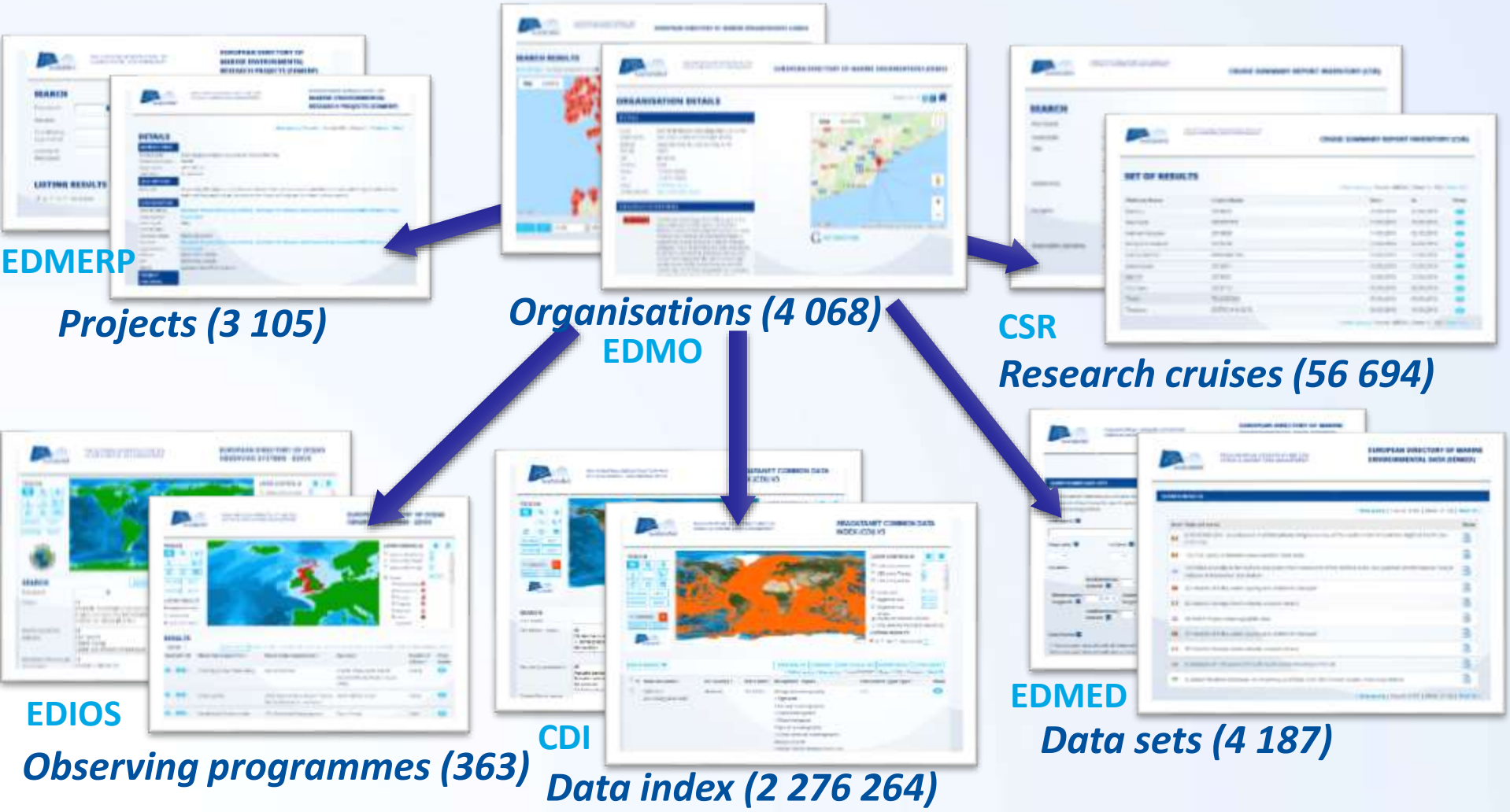
Linked Data

"The Semantic Web isn't just about putting data on the web. It is about making links, so that a person or machine can explore the web of data. With Linked Data, when you have some of it, you can find other, related, data." ***Sir Tim Berners-Lee, 2006***

- Standards – Uniform Resource Identifiers (**URIs**), Hypertext Transfer Protocol (**HTTP**) and Resource Description Framework (**RDF**) enabling web interoperability


carbonara
All
Images
Videos
Shopping
News
More
Settings
Tools
About 32,000,000 results (0.37 seconds)
Ultimate spaghetti carbonara recipe | BBC Good Food
https://www.bbcgoodfood.com/recipes/1052/ultimate-spaghetti-carbonara
Rating: 4.7 - 213 votes - 35 mins - Calories: 655
Discover how to make superb spaghetti carbonara. This cheesy pasta dish is an Italian favourite and with the right technique, you can make it perfect every time.
Videos
4:40
How to Make Classic Carbonara | Jamie Oliver
Jamie Oliver
YouTube - 5 Aug 2017
5:13
Real Spaghetti Carbonara | Antonio Carluccio
Jamie Oliver
YouTube - 27 Mar 2014
6:34
Easy Spaghetti Carbonara | Gennaro Contaldo
Jamie Oliver
YouTube - 7 Nov 2015
Learn to make spaghetti carbonara recipe | BBC Good Food
https://www.bbcgoodfood.com/recipes/2876/learn-to-make-spaghetti-carbonara
Rating: 3.8 - 42 votes - 25 mins - Calories: 734
When hot, tip in 200g smoked pancetta cubes or chopped streaky bacon. ... Tip the yolk in with the other eggs and beat together, along with 75ml double cream, 50g finely grated Grana Padano or parmesan cheese and some seasoning. ... Pour egg mixture over pasta, followed by hot pancetta ...
Classic Spaghetti Carbonara | Pasta Recipes | Jamie Oliver
https://www.jamieoliver.com/recipes/pasta.../gennaro-s-classic-spaghetti-carbonara/
10 mins - Calories: 860
More images
Carbonara
Pasta dish
Carbonara is an Italian pasta dish from Rome made with egg, hard cheese, guanciale, and pepper. The recipe is not fixed by a specific type of hard cheese or pasta. The cheese is usually Pecorino Romano. Spaghetti is the usual pasta, however, fettuccine, rigatoni, linguine, or bucatini are also used. Wikipedia
Variations: (US) peas, mushrooms, or other vegetables, cream
Serving temperature: Hot
Course: Primo (pasta course, Italy); main course (elsewhere)
Region or state: Rome/Lazio
Origin: Italy
People also search for
View 15+ more
Bolognese sauce
Pancetta
Fettuccine Alfredo
Spaghetti

Linked Data and SeaDataNet

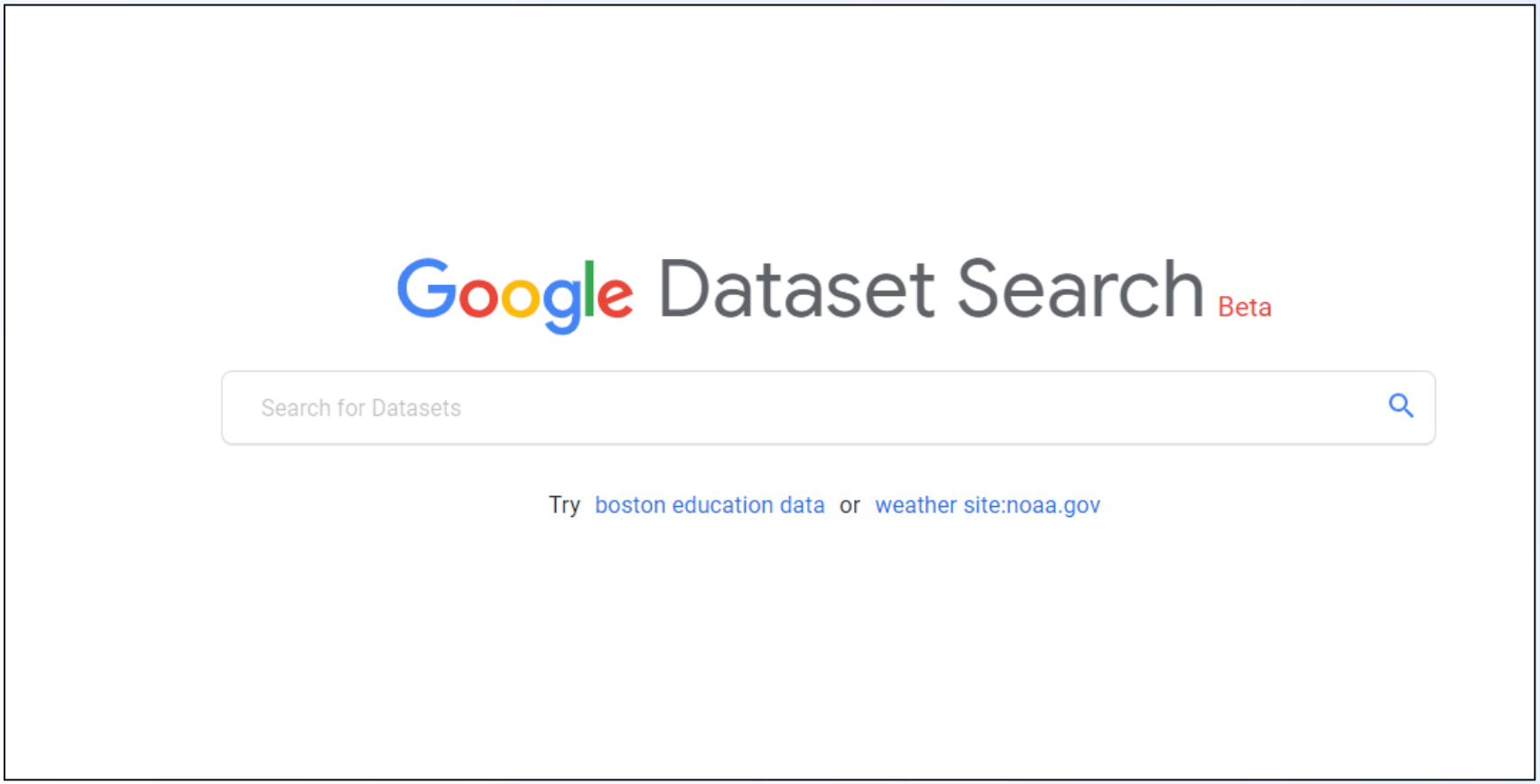


Metadata catalogues modelled using existing standards…

Catalogue	Entity	Exisiting Standard(s)	Host
EDMO	Organisation	W3C Organisation	Maris
EDMED	Data Set	W3C DCAT	BODC
EDMERP	Research Project	W3C Prov-O DBPedia Research Project	Maris
EDIOS	Observing System	INSPIRE Environmental Monitoring Facilities	BODC
CSR	Cruise	?... Publish SDC work	BSH
CDI	Data granule/series	W3C DCAT	Maris

SPARQL Endpoints

- A web service enabling users (human or machine) to query metadata databases as Linked Data
- SPARQL is a W3C standard and the language used to query RDF data
- Deploying SPARQL Endpoints enables interested parties to harvest, reuse and link to the SeaDataNet catalogues
- **EDMED, EDIOS, EDMO are now exposed as Linked Data resources, via SPARQL Endpoints. EDMERP and CSR are in progress**


Google Dataset Search Beta
Search for Datasets
Try boston education data or weather site:noaa.gov

Clean URIs

- Common URIs adopted for the SeaDataNet metadata catalogues to enable simpler linking for machine-to-machine services:

  http:///*[metadir]*.seadatanet.org/

- Appropriate redirects in place by individual catalogue providers

Other WP8 highlights

Review of data formats, also considering the INSPIRE data models

Including:

- Proof of concept mappings - extension of NetCDF and ODV formats to achieve **INSPIRE-compliance** within SeaDataNet. Subsequently modelled by OGS for nutrient data in the Mediterranean Sea
- Review of SeaDataNet migration from **NetCDF v3.6 to v4.0**
- Formalised SeaDataNet NetCDF (CF) format for gridded data

Infrastructural upgrades to assimilate EUDAT

- User authentication and authorisation
- Monitoring system for 'health' of SeaDataNet infrastructure