27-29 May 2024

International conference on Marine Data and Information Systems

SeaDataNet AISBL

NAVFORSKNINGS INSTITUTTET

Ifremer

MARIS

MEDIN

National Oceanography Centre

INGV

iMagine

ICDE

eosc Blue-Cloud2026

National Oceanography Centre

British Oceanographic Data Centre

THE SEMANTIC ANALYSER

Alexandra Kokkinaki, BODC (UK), alexk@noc.ac.uk, Gwenaëlle Moncoiffé, BODC (UK), gmon@noc.ac.uk , David Habgood KurrawongAI(AU), david@kurrawong.ai, Enrico Boldrini CNR(IT), enrico.boldrini@cnr.it, Nicholas Car, KurrawongAI (AU), nick@kurrawong.ai

INTRODUCTION

National Oceanography Centre

- Conception
- FAIR-EASE multidisciplinary project
  - Tools, services
    - Discover, access and analyse multidisciplinary data
    - Collaboratively
  - Earth & Environmental Dynamics
    - Coastal Water Dynamics Pilot
    - Earth Critical Zone planning and management Pilot
    - Volcano Space Observatory Pilot
  - Environmental Bio-geochemistry
  - Biodiversity Observations

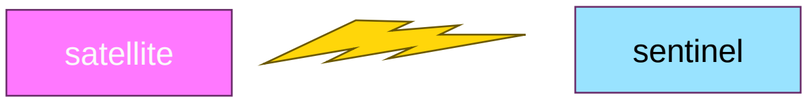
There are currently many datasets for the coastal domain whose formats and aggregation status vary widely. Present analysis and visualization systems deal with **single datasets** at a time, making **data connections and correlations between datasets difficult.**

No platform currently allows to seamlessly and jointly explore **satellite data** from both the **Solid Earth** and **Atmospheric** communities. Lack of a supporting platform allowing for the joint analysis of crucial near-real-time indicators.

SEMANTIC BROKERAGE

- Semantic brokers are intermediaries that provide semantic translation services across **diverse** sources

Argo
Wekeo
EurOBIS
SeaDataNet
ELIXIR-ENA
ICOS SOCAT
Vito/Copernicus
ICOS Data Portal
EMODNet Chemistry
US NODC Collections



- names for a type of observation platform
- ‘Sentinel’ is a particular type of earth observation satellite;
- make the relationship available in a machine-readable way.

Complements Syntactic brokerage

DIVERSITY OF SOURCES

Syntax:. Syntax refers to the structure or format of data. It involves the rules that define the arrangement of symbols, words, or elements in a system.

e.g. XML, RDF, CSW, openAPI, netCDF, ISO19115, ISO19139, json

Semantics: refer to the meaning of data. They involve the interpretation of the symbols, words, or elements based on their context.

e.g. sentinel # satellite

Sources

- Argo
- Wekeo
- EurOBIS
- SeaDataNet
- ELIXIR-ENA
- ICOS SOCAT
- Vito/Copernicus
- ICOS Data Portal
- EMODNet Chemistry
- US NODC Collections

We cannot change either the syntax or the semantics of the sources

SYNTACTIC HARMONISATION: DISCOVERY AND ACCESS (DAB) BROKER

National Oceanography Centre

19115/19139

Open Archives Initiative Protocol for Metadata Harvesting

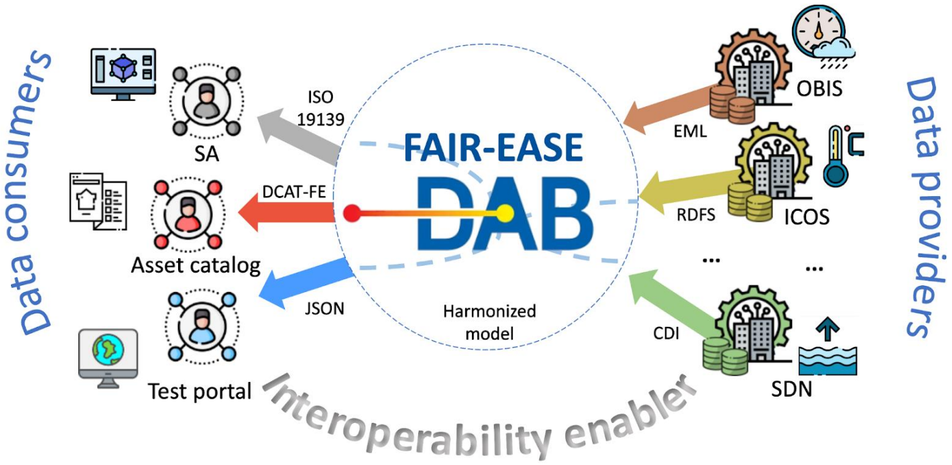
Open Geospatial Consortium

EML ERDDAP

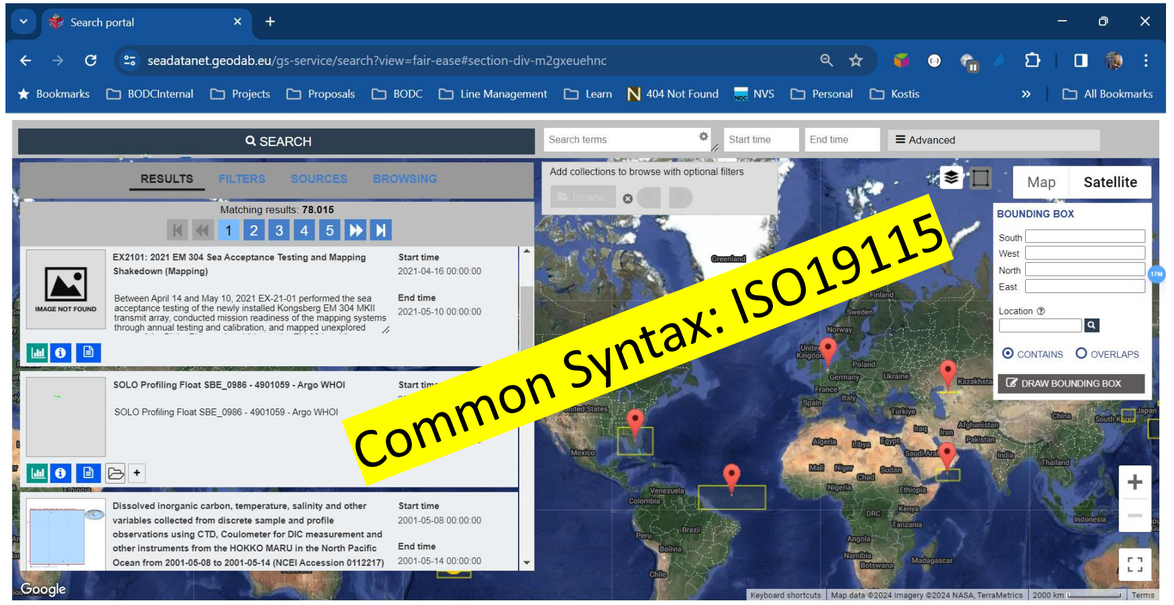
DCAT

openSearch

JSON-LD RDF JSON XML SPARQL



DAB is a brokering framework that interconnects hundreds of heterogeneous and autonomous supply systems by providing mediation, harmonization, transformation, and QoS capabilities

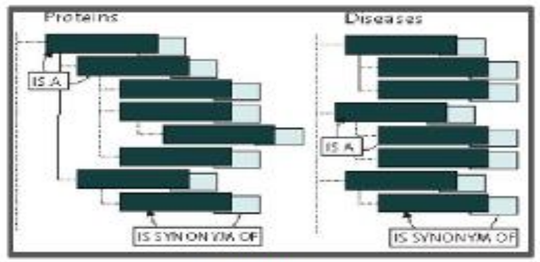


SEMANTIC BROKERAGE

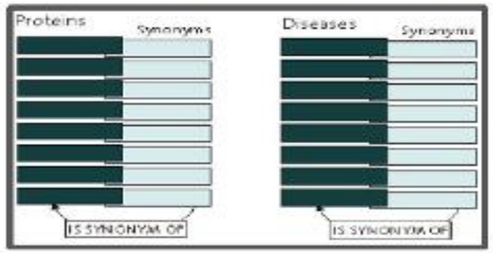
- Comprehensive understanding of the datasets
  - Metadata elements
  - Semantic artefacts

Controlled Vocabulary (CV):
An authoritative list of terms

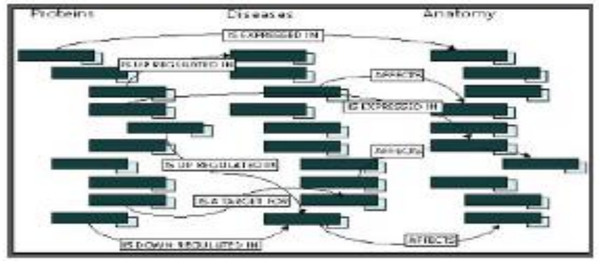
Taxonomy:
A CV with a tree-hierarchical (parent/child term) structure



Thesaurus:
A kind of taxonomy with structure and specific types of relationships between terms



Ontology:
is a kind of taxonomy, but the types of relationships are greater in number and more specific in their function



Argo
Wekeo
EurOBIS
SeaDataNet
ELIXIR-ENA
ICOS SOCAT
Vito/Copernicus
ICOS Data Portal
EMODNet Chemistry
US NODC Collections

From: "Terminology for FAIR data competency – Considerations: before we get our hands dirty!" (2019), S.A. Sansone and P. McQuilton – Terminology Workshop – May 20-21 2019 - Paris

SEMANTIC BROKER: METADATA ELEMENTS

The common FAIR-EASE metadata elements are:

- IDENTIFIER: Blue-Cloud unique and persistent code for the metadata record
- TITLE: a characteristic, and often unique, name by which the collection is known
- ABSTRACT: a short description of the collection
- KEYWORD: a commonly used word, formalised word or phrase used to describe the subject
- BOUNDING_BOX: exte e geographic space given as a bounding box
- TEMPORAL_EXTENT: time period covered by the content of the collection
- PARAMETER: name of the attribute urement value
- INSTRUMENT: measuring instrumer data
- PLATFORM: platform from which the data were taken
- ORGANIZATION: organiz he collection
- DATESTAMP: the latest u adata description
- REVISION_DATE: the late data
- RESOURCE_LINKS: download links where available and useful

Variable measured

Sensing or sampling instrument

Observation platform

National Oceanography Centre

THE ANALYSIS

```
▼<gmd:keyword>
    <gco:CharacterString>ammonia</gco:CharacterString>
  </gmd:keyword>
▼<gmd:keyword>
    <gco:CharacterString>nutrient</gco:CharacterString>
  </gmd:keyword>
▼<gmd:keyword>
    <gco:CharacterString>concentration (value)</gco:CharacterString>
  </gmd:keyword>
▼<gmd:keyword>
    <gco:CharacterString>soil pollution</gco:CharacterString>
  </gmd:keyword>
▼<gmd:keyword>
    <gco:CharacterString>phosphorus</gco:CharacterString>
  </gmd:keyword>
▼<gmd:keyword>
    <gco:CharacterString>ecosystem degradation</gco:CharacterString>
  </gmd:keyword>
▼<gmd:keyword>
    <gco:CharacterString>land use</gco:CharacterString>
  </gmd:keyword>
▼<gmd:keyword>
    <gco:CharacterString>agricultural land</gco:CharacterString>
  </gmd:keyword>
▼<gmd:keyword>
    <gco:CharacterString>soil degradation</gco:CharacterString>
  </gmd:keyword>
▼<gmd:keyword>
    <gco:CharacterString>agriculture</gco:CharacterString>
  </gmd:keyword>
▼<gmd:keyword>
    <gco:CharacterString>environmental pressure</gco:CharacterString>
  </gmd:keyword>
▼<gmd:keyword>
    <gco:CharacterString>soil</gco:CharacterString>
  </gmd:keyword>
```

Search results for: soil degradation

Previous 1

soil degradation ENVO:01000705

http://purl.obolibrary.org/obo/ENVO_01000705

Soil degradation is a process which results in a change in soil
in which the soil occurs to provide goods and services for its b

Ontology: ENVO

Strings

Search results for: phosphorus

Previou

phosphorus atom CHEBI:28659 IMPORTED

http://purl.obolibrary.org/obo/CHEBI_28659

Ontology: ENVO

Also appears in: RXNO ZP AGRO AISM EFO

```
▼<gmd:MD_Keywords>
  ▼<gmd:keyword>
     <gco:CharacterString>saline water (ENVO:00002010), including
   </gmd:keyword>
  ▼<gmd:keyword>
     <gco:CharacterString>marine epipelagic mixed layer (ENVO:xx
   </gmd:keyword>
  ▼<gmd:keyword>
     <gco:CharacterString>"saline water (ENVO:00002010), including plankton (ENVO:xxxxxxxx)"</gco:CharacterString>
   </gmd:keyword>
  ▼<gmd:keyword>
     <gco:CharacterString>surface water (ENVO:00002042) layer</gco:CharacterString>
   </gmd:keyword>
  ▼<gmd:keyword>
     <gco:CharacterString>mesopelagic zone (ENVO:00000213)</gco:CharacterString>
   </gmd:keyword>
  ▼<gmd:keyword>
     <gco:CharacterString>mesopelagic zone (ENVO:00000213) & marine oxygen minimum zone (ENVO:01000065)</gco:CharacterString>
   </gmd:keyword>
  ▼<gmd:keyword>
```

Environment Ontology (ENVO)

Keywords: marine epipelagic mixed layer | Search terms | Batch Search

No terms returned, please try different keywords.

Terms with 'marine epipelagic mixed layer' included in their label:

Strings: Combination label and code

saline water IMPORTED

http://purl.obolibrary.org/obo/ENVO_00002010 Copy

Water which contains a significant concentration of dissolved salts.

The threshold salt concentration for classifying water as saline varies, but typically begins at about 1,000 to 3,000 parts salt per million parts water or 0.1–0.3% salt by weight.

Defined by ENVO

Also appears in ECTO MICRO AGRO RBO MONDO ENM IDOMAL GAZ FOODON

```
    <gmi:identifier>
      <gmd:MD_Identifier id="11620516">
        <gmd:code>
          <gco:CharacterString>CTD_PRES</gco:CharacterString>
        </gmd:code>
      </gmd:MD_Identifier>
    </gmi:identifier>
    <gmi:description>
      <gco:CharacterString>Sensor Model: DRUCK_2900PSIA. Maker: DRUCK</gco:CharacterString>
    </gmi:description>
  </gmi:MI_Instrument>
</gmi:instrument>
<gmi:instrument>
  <gmi:MI_Instrument>
    <gmi:citation>
      <gmd:CI_Citation>
        <gmd:title>
          <gco:CharacterString>SBE41CP_V7.2.5</gco:CharacterString>
        </gmd:title>
      </gmd:CI_Citation>
    </gmi:citation>
    <gmi:identifier>
      <gmd:MD_Identifier id="14356">
        <gmd:code>
          <gco:CharacterString>CTD_TEMP</gco:CharacterString>
        </gmd:code>
      </gmd:MD_Identifier>
    </gmi:identifier>
    <gmi:description>
      <gco:CharacterString>Sensor Model: SBE41CP_V7.2.5. Maker: SBE</gco:CharacterString>
    </gmi:description>
  </gmi:MI_Instrument>
</gmi:instrument>
<gmi:instrument>
  <gmi:MI_Instrument>
    <gmi:citation>
      <gmd:CI_Citation>
        <gmd:title>
          <gco:CharacterString>SBE41CP_V7.2.5</gco:CharacterString>
        </gmd:title>
      </gmd:CI_Citation>
    </gmi:citation>
    <gmi:identifier>
      <gmd:MD_Identifier id="14356">
        <gmd:code>
          <gco:CharacterString>CTD_CNDC</gco:CharacterString>
        </gmd:code>
      </gmd:MD_Identifier>
    </gmi:identifier>
    <gmi:description>
      <gco:CharacterString>Sensor Model: SBE41CP_V7.2.5. Maker: SBE</gco:CharacterString>
    </gmi:description>
  </gmi:MI_Instrument>
</gmi:instrument>
```

Concept

Conductivity Temperature Depth (CTD) sensors package measuring pressure

URI	http://vocab.nerc.ac.uk/collection/R25/current/CTD_PRES/
Within Vocab	Argo sensor types
Alternative Labels	CTD_PRES
Definition	Conductivity Temperature Depth (CTD) sensors package measuring pressure
Date	2019-10-11T15:06:36
Identifier	SDN:R25::CTD_PRES
Note	accepted
Has Current Version	1
version	1

Strings: Combination of Metadata fields+ label

```
<gmd:keyword>
  <gmx:Anchor xlink:href="http://vocab.nerc.ac.uk/collection/P02/current/ALKY/">Alkalinity, acidity and pH of the water column</gmx:Anchor>
</gmd:keyword>
<gmd:keyword>
  <gmx:Anchor xlink:href="http://vocab.nerc.ac.uk/collection/P02/current/CPWC/">Chlorophyll pigment concentrations in water bodies</gmx:Anchor>
</gmd:keyword>
<gmd:keyword>
  <gmx:Anchor xlink:href="http://vocab.nerc.ac.uk/collection/P02/current/PSAL/">Salinity of the water column</gmx:Anchor>
</gmd:keyword>
<gmd:keyword>
  <gmx:Anchor xlink:href="http://vocab.nerc.ac.uk/collection/P02/current/TEMP/">Temperature of the water column</gmx:Anchor>
</gmd:keyword>
<gmd:keyword>
  <gmx:Anchor xlink:href="http://vocab.nerc.ac.uk/collection/P02/current/NTRA/">Nitrate concentration parameters in the water column</gmx:Anchor>
</gmd:keyword>
<gmd:keyword>
  <gmx:Anchor xlink:href="http://vocab.nerc.ac.uk/collection/P02/current/NTRI/">Nitrite concentration parameters in the water column</gmx:Anchor>
</gmd:keyword>
<gmd:keyword>
  <gmx:Anchor xlink:href="http://vocab.nerc.ac.uk/collection/P02/current/PHOS/">Phosphate concentration parameters in the water column</gmx:Anchor>
</gmd:keyword>
<gmd:keyword>
  <gmx:Anchor xlink:href="http://vocab.nerc.ac.uk/collection/P02/current/TDPX/">Dissolved total or organic phosphorus concentration in the water column</gmx:Anchor>
</gmd:keyword>
<gmd:keyword>
  <gmx:Anchor xlink:href="http://vocab.nerc.ac.uk/collection/P02/current/AMON/">Ammonium and ammonia concentration parameters in water bodies</gmx:Anchor>
</gmd:keyword>
<gmd:keyword>
  <gmx:Anchor xlink:href="http://vocab.nerc.ac.uk/collection/P02/current/SLCA/">Silicate concentration parameters in the water column</gmx:Anchor>
</gmd:keyword>
<gmd:keyword>
  <gmx:Anchor xlink:href="http://vocab.nerc.ac.uk/collection/P02/current/TDNT/">Dissolved total and organic nitrogen concentrations in the water column</gmx:Anchor>
</gmd:keyword>
<gmd:keyword>
  <gmx:Anchor xlink:href="http://vocab.nerc.ac.uk/collection/P02/current/DOXY/">Dissolved oxygen parameters in the water column</gmx:Anchor>
</gmd:keyword>
<gmd:keyword>
  <gmx:Anchor xlink:href="http://vocab.nerc.ac.uk/collection/P02/current/SECC/">Secchi disk depth</gmx:Anchor>
</gmd:keyword>
<gmd:keyword>
  <gmx:Anchor xlink:href="http://vocab.nerc.ac.uk/collection/P02/current/NTOT/">Particulate total and organic nitrogen concentrations in the water column</gmx:Anchor>
</gmd:keyword>
<gmd:keyword>
  <gmx:Anchor xlink:href="http://vocab.nerc.ac.uk/collection/P02/current/MTWD/">Dissolved metal concentrations in the water column</gmx:Anchor>
</gmd:keyword>
```

Combination of URI+label

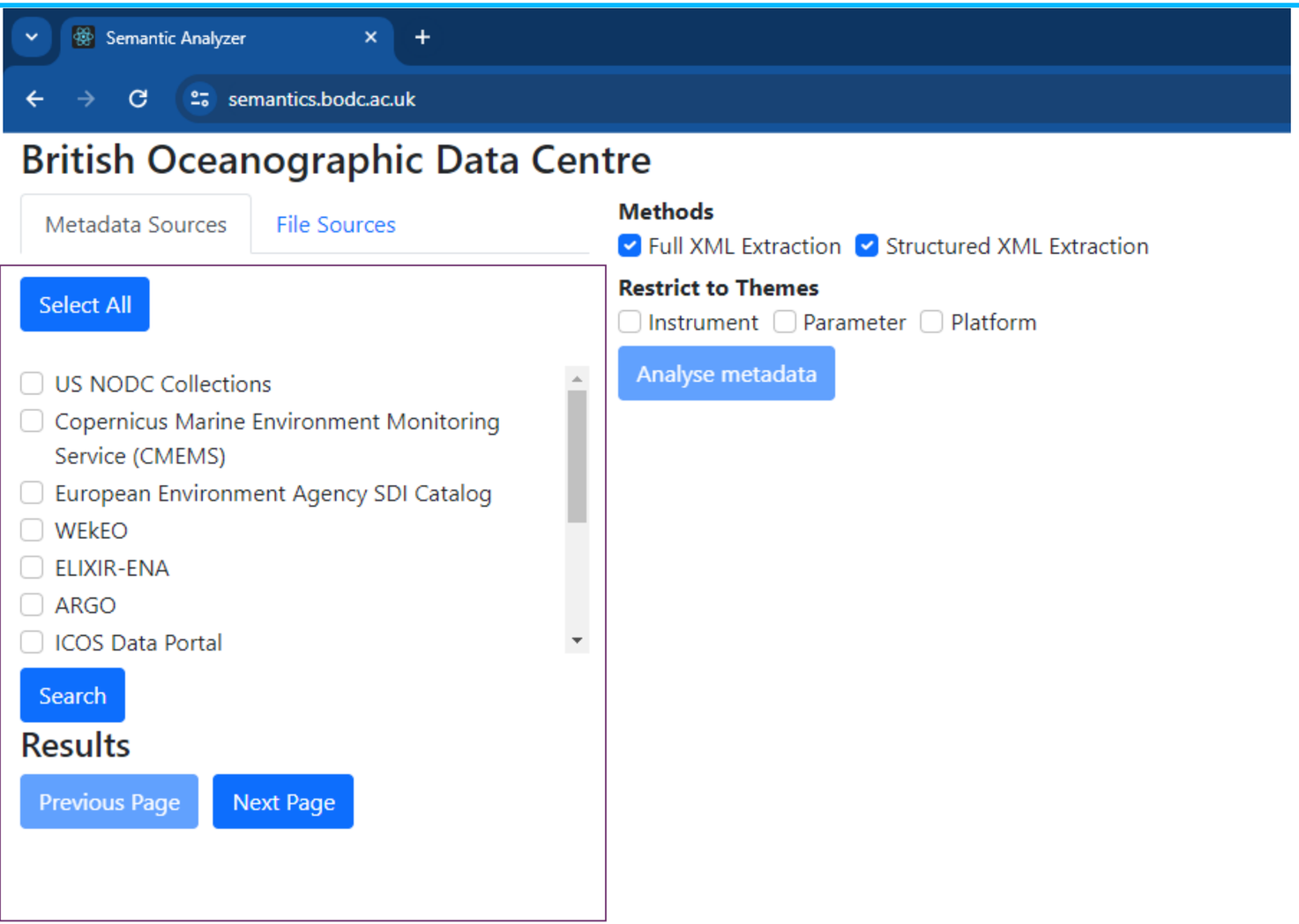
MACHINE READABLE

~~MACHINE ACTIONABLE~~

National Oceanography Centre

THE SEMANTIC ANALYSER

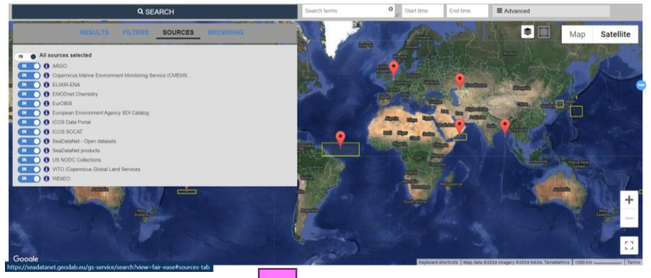
THE SEMANTIC ANALYSER



- Tool to analyse the semantics of metadata and data files
- Across multiple domains
- For internal use – but can be used by external users too

https://semantics.bodc.ac.uk/

THE ARCHITECTURE



ISO19115
netCDF

Metadata records and data files

Extract Terms from Structure of XML

Extract all terms from XML

Match Terms in KB

URI + Exact Match + Proximity Match

Results Presentation - all options / output shown to user as table

Output estimates

Select Terms with a Poor Match

Federated Search across other search portals

Results Page

EarthPortal

BioPortal

Query Knowledge Base

NVS

GCMD

EIONET

Vocabulary Collections

Knowledge Base

Metadata Models

Source institution information

Apache Jena Fuseki

THE KNOWLEDGE BASE (KB)

National Oceanography Centre

Apache Jena Fuseki datasets manage help server status

/fair-ease-with-rdfs

query add data edit info

SPARQL Query

To try out some SPARQL queries against the selected dataset, enter your query here.

Example Queries

Selection of triples Selection of classes

Prefixes

rdf rdfs owl xsd

SPARQL Endpoint

/fair-ease-with-rdfs/query

Content Type (SELECT)

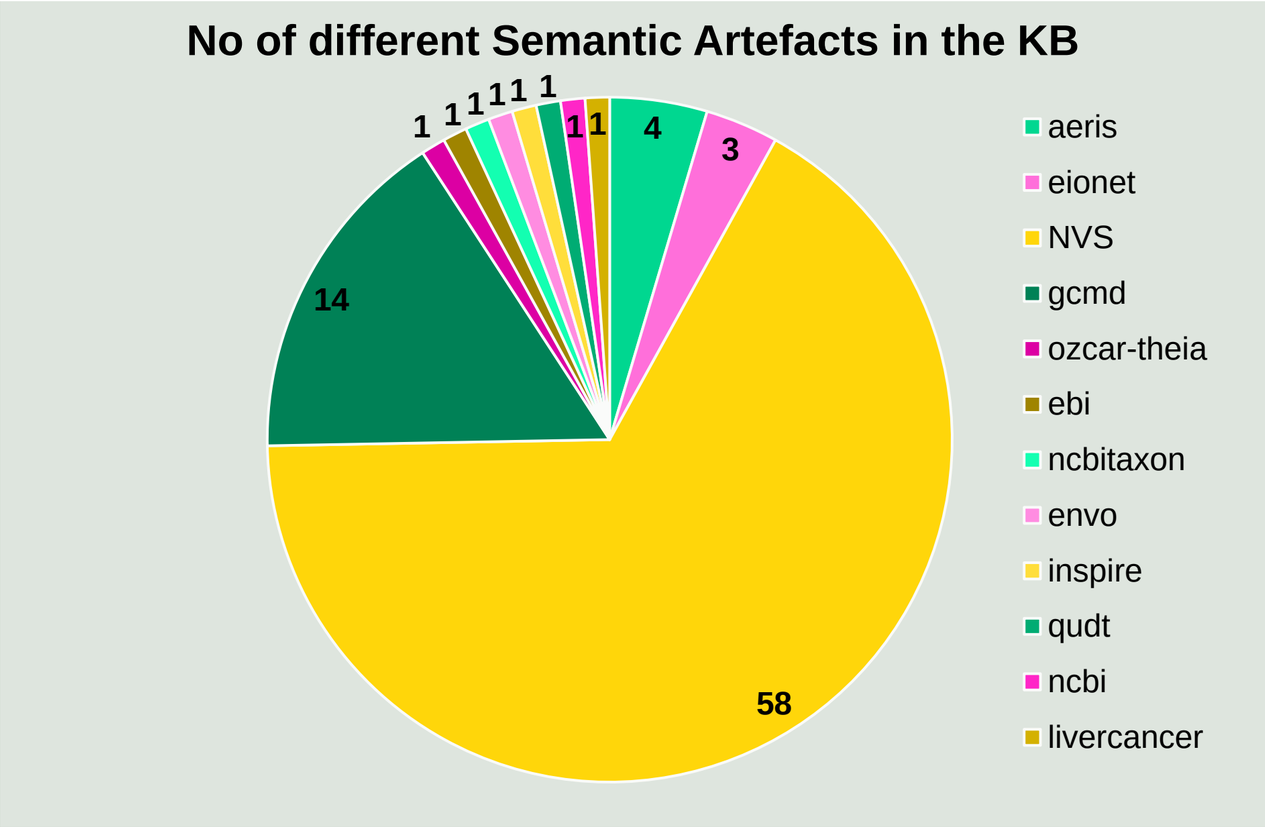
JSON

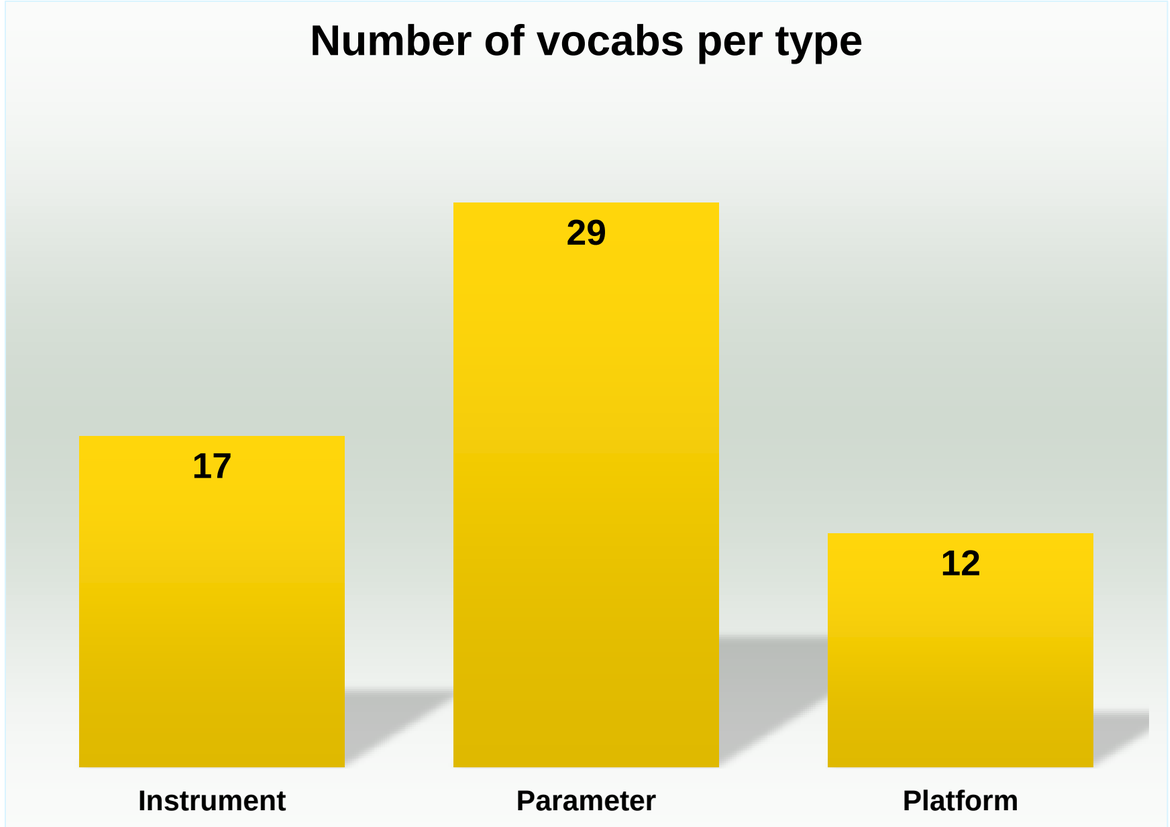
Content Type (GRAPH)

Turtle

```
1  SELECT  DISTINCT ?g
2    WHERE  { GRAPH ?g {} }
3  ORDER BY  ?g
```

THE KB


No of different Semantic Artefacts in the KB
4
3
58
14
1
1
1
1
1
1
1
1
aeris
eionet
NVS
gcmd
ozcar-theia
ebi
ncbitaxon
envo
inspire
qudt
ncbi
livercancer


Number of vocabs per type
17
29
12
Instrument
Parameter
Platform

UPDATES OF SEMANTIC ARTEFACTS

- **Dynamic Nature of Semantic Artefacts**
  - as new use cases emerge
  - as the underlying content and comprehension undergo changes
- **Update Frequency**:
  - The frequency of updates varies significantly across different artefacts.
- **Lack of Consensus**: There is no agreement within the semantic community on identifying the latest version of a semantic resource for automated updates.

Daily updates: for controlled vocabularies from NVS

Twice a year: Remaining KB

Note: Optimal frequency is still a subject of debate

Semantic Analyzer

semantics.bodc.ac.uk

British Oceanographic Data Centre

Metadata Sources | File Sources

Select All

- VITO /Copernicus Global Land Services
- SeaDataNet - Open datasets
- SeaDataNet products
- EMODnet Chemistry
- ELIXIR-ENA
- [x] ARGO
- ICOS Data Portal
- ICOS SOCAT

1

Search

Results

Previous Page | Next Page

C4FCB80C939B2B9DC3E9686BCDB67507780799C3
PROVOR-V JUMBO Profiling Float - 2903783 - Argo LOV
PROVOR-V JUMBO Profiling Float - 2903783 - Argo LOV

B5BB4DD923E61118CC6D968CBF386EAC292EF68F
SOLO Profiling Float SBE_0986 - 4901059 - Argo WHOI
SOLO Profiling Float SBE_0986 - 4901059 - Argo WHOI

2

550D4232CFACDB562DC3663B33A58C1A4CBF4908
APEX Profiling Float - 2900141 - US Argo Project (Peter Hacker)
APEX Profiling Float - 2900141 - US Argo Project (Peter Hacker)

7100A10CA839765C29AC690EE6F315435E3813EC
APEX Profiling Float - 2900604 - Argo KORDI
APEX Profiling Float - 2900604 - Argo KORDI

FCA2805FE6D2E666E1062B4DE85D337F1B7762475

1 files selected:
- PROVOR-V JUMBO Profiling Float - 2903783 - Argo LOV View XML Remove

Methods
[x] Full XML Extraction [x] Structured XML Extraction

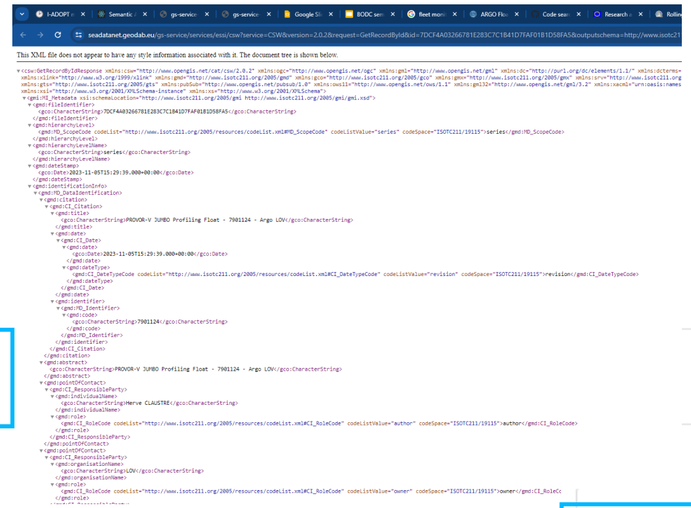
Restrict to Themes
[ ] Instrument [ ] Parameter [ ] Platform

Analyse metadata

3

Terms not found - search in BioPortal and EarthPortal
PROVOR-V JUMBO Profiling Float - 2903783 - Argo LOV

EarthPortal BioPortal



Export
- Terms Not Found CSV
- Results CSV
- JSON

Document: **PROVOR-V JUMBO Profiling Float - 2903783 - Argo LOV**

Analyser Method: **Full XML Extraction**
No results found for method:
Analyser Method: **Structured XML Extraction**

SearchTerm	Match Concept	MatchProperty	Vocab Categories
Instrument (137 items)			
URI Match (0 items)			
Exact Match (15 items)			
Proximity Match (19 items)			
Wildcard Match (103 items)			
Variable (104 items)			
URI Match (0 items)			
Exact Match (38 items)			
Proximity Match (10 items)			
Wildcard Match (56 items)			
Platform (10 items)			
URI Match (0 items)			
Exact Match (0 items)			

URI Match
- http, https

Exact Match
- prefLabel, altLabel, label

Proximity Match
- Fuzzy search

Wildcard Match
- ?

EXAMPLE: INSTRUMENTS

semantics.bodc.ac.uk

British Oceanographic Data Centre

Metadata Sources | File Sources

Select All

- VITO /Copernicus Global Land Services
- ICOS Data Portal
- SeaDataNet - Open datasets
- SeaDataNet products
- ARGO
- [x] ELIXIR-ENA
- ICOS SOCAT
- EMODnet Chemistry

Search

Results

Previous Page | Next Page

DA9362C17114D21008C3ABAE3D29DE30E0A636F6
Amplicon sequencing of Tara Oceans RNA samples corresponding to size fractions for protists.
Analysis of RNA tags in Tara Oceans Protists size fractions through amplicon sequencing: Seawater wa...

CF9EC0557FCE1447B26D9186BF46345067C955C6
Metatranscriptome sequencing from Tara Oceans Polar Circle samples corresponding to size fractions for protists.
Analysis of the genes present in Tara Oceans Polar Circle protist size fractions through RNA sequenc...

BC136CDF96B6714C6CC31A61829E5AF7047463CC
Shotgun Sequencing of Tara Oceans DNA samples corresponding to size fractions for protist
Seawater was filtered from different depths to retain small and large cell sizes. The DNA was extrac...

1 files selected:

- Amplicon sequencing of Tara Oceans RNA samples cor... View XML Remove

Methods

[x] Full XML Extraction [x] Structured XML Extraction

Restrict to Themes

[ ] Instrument [ ] Parameter [ ] Platform

Analyse metadata

Terms not found - search in BioPortal and EarthPortal
Amplicon sequencing of Tara Oceans RNA samples corresponding to size fractions for protists.

Document: **Amplicon sequencing of Tara Oceans RNA samples corresponding to size fractions for p**

Analyser Method: **Full XML Extraction**

No results found for method:

Analyser Method: **Structured XML Extraction**

SearchTerm	Match Concept			
Instrument (24 items)				
URI Match (0 items)				
Exact Match (1 items)				
Collection: Unknown (1 items)				
Illumina Genome Analyzer IIx	Illumina Genome Analyzer IIx	label	Instrument	Text Match
Wildcard Match (22 items)				
Variable (11 items)				
URI Match (0 items)				
Exact Match (1 items)				
Collection: Unknown (1 items)				

Search term | Matched term URL | Matched Property | Vocabulary Category | Method used

OLS ONTOLOGY SEARCH

Home | Ontologies | Help | About | Downloads

Ontologies ▸ EFO ▸ Classes ▸ EFO:0004202 Copy

Illumina Genome Analyzer IIx

http://www.ebi.ac.uk/efo/EFO_0004202 Copy

A DNA sequencer developed by Illumina.

Search EFO...

[ ] Exact match [ ] Include obsolete terms [x] Include imported terms

Tree | Graph

- experimental factor (48,619)
  - material entity BFO (13,236)
    - instrument (81)
      - sequencer (42)
        - DNA sequencer OBI (3)
          - Illumina Genome Analyzer IIx
        - high throughput sequencer (39)
          - Illumina Genome Analyzer IIx

EXAMPLE: URI MATCHES

semantics.bodc.ac.uk

Search

Results

Previous Page | Next Page

452e6b4d-9667-4902-b4dd-e2b98f04a312
EX2101: 2021 EM 304 Sea Acceptance Testing and Mapping Shakedown (Mapping)
Between April 14 and May 10, 2021 EX-21-01 performed the sea acceptance testing of the newly install...

681aec46-fcd1-47ff-acef-ca7f80886b9d
Dissolved inorganic carbon, temperature, salinity and other variables collected from discrete sample and profile observations using CTD, Coulometer for DIC measurement and other instruments from the HOKKO MARU in the North Pacific Ocean from 2001-05-08 to 2001-05-14 (NCEI Accession 0112217)
This dataset includes chemical, discrete sample, physical and profile data collected from HOKKO MARU...

49dd9c0d-f6af-4167-88f3-5ef8702ba6a5
High-resolution ocean and atmosphere pCO2 time-series measurements from mooring Mooring BOBOA_90E_15N in the Indian Ocean (NCEI Accession 0162473)
This dataset includes chemical, meteorological, physical time series data collected from BOBOA_90E_1...

46cef878-9bdb-4f10-bbe7-c07224608281
NOAA/WDS Paleoclimatology - Lake Status Records, Caspian and Aral Seas
This archived Paleoclimatology Study is available from the NOAA National Centers for Environmental I...

66711400-4ef8-4071-9b31-ae9342a44b89
Temperature, salinity, pressure, and other data

https://gcmd.earthdata.nasa.gov/kms/concept/4fde380a-38c5-4d46-bc80-4f2515a43983

Document: **Dissolved inorganic carbon, temperature, salinity and other variables collected from discrete sample and profile observations using CTD, Coulometer for DIC measurement and other instruments from the HOKKO MARU in the North Pacific Ocean from 2001-05-08 to 2001-05-14 (NCEI Accession 0112217)**

Export

Analyser Method: **Full XML Extraction**

SearchTerm	Match Concept	MatchProperty	Vocab Categories	Method
All (13 items)				
URI Match (13 items)				
Collection: Unknown (13 items)				
https://gcmd.earthdata.nasa.gov/kms/concept/4fde380a-3	NITRATE	N/A	Parameter, Instrument, ...	URI Match
https://gcmd.earthdata.nasa.gov/kms/concept/0b6020a0-2	COULOMETERS	N/A	Instrument	URI Match
https://gcmd.earthdata.nasa.gov/kms/concept/c91c8879-1	SILICATE	N/A	Parameter, Instrument, ...	URI Match
https://gcmd.earthdata.nasa.gov/kms/concept/46206e8c-8	WATER TEMPERATURE	N/A	Parameter, Instrument, ...	URI Match
https://gcmd.earthdata.nasa.gov/kms/concept/1bb21d0f-b	Ships	N/A	Platform	URI Match
https://gcmd.earthdata.nasa.gov/kms/concept/941410da-0	NITRITE	N/A	Parameter, Instrument, ...	URI Match
https://gcmd.earthdata.nasa.gov/kms/concept/7e95b5fc-1	SALINITY	N/A	Parameter, Instrument, ...	URI Match
https://gcmd. /concept/772f5ac7-2	DOE/ORNL/ESD/CDIAC	N/A		URI Match
https://gcmd.earthdata.nasa.gov/kms/concept/01cc0beb-7	CTD	N/A	In	URI Match
https://gcmd.earthdata.nasa.gov/kms/concept/d9b4f30d-b	INORGANIC CARBON	N/A	Parameter, Instrument, ...	URI Match
https://gcmd.earthdata.nasa.gov/kms/concept/5488b1cc-0	NORTH PACIFIC OCEAN	N/A		URI Match
https://gcmd.earthdata.nasa.gov/kms/concept/0b513d8c-b	PHOSPHATE	N/A	Parameter, Instrument, ...	URI Match
https://gcmd.earthdata.nasa.gov/kms/concept/dd025312-0	WATER PRESSURE	N/A	Parameter, Instrument, ...	URI Match
Exact Match (0 items)				
Proximity Match (0 items)				
Wildcard Match (0 items)				

Search term

Matched term URL

Vocabulary Category

Method used

Analyser Method: **Structured XML Extraction**

SearchTerm	Match Concept	MatchProperty	Vocab Categories	Method
Keywords (294 items)				
URI Match (0 items)				

EXAMPLE: UNMATCHED TERMS

semantics.bodc.ac.uk

British Oceanographic Data Centre

Metadata Sources | File Sources

Select All

- VITO /Copernicus Global Land Services
- ICOS Data Portal
- SeaDataNet - Open datasets
- SeaDataNet products
- ARGO
- ELIXIR-ENA
- ICOS SOCAT
- EMODnet Chemistry

Search

Results

Previous Page | Next Page

DA9362C17114D21008C3ABAE3D29DE30E0A636F6
Amplicon sequencing of Tara Oceans RNA samples corresponding to size fractions for protists.
Analysis of RNA tags in Tara Oceans Protists size fractions through amplicon sequencing: Seawater wa...

CF9EC0557FCE1447B26D9186BF46345067C955C6
Metatranscriptome sequencing from Tara Oceans Polar Circle samples corresponding to size fractions for protists.
Analysis of the genes present in Tara Oceans Polar Circle protist size fractions through RNA sequenc...

BC136CDF96B6714C6CC31A61829E5AF7047463CC
Shotgun Sequencing of Tara Oceans DNA samples corresponding to size fractions for protist
Seawater was filtered from different depths to retain small and large cell sizes. The DNA was extrac...

1 files selected:

- Amplicon sequencing of Tara Oceans RNA samples cor... View XML Remove

Methods

Full XML Extraction | Structured XML Extraction

Restrict to Themes

Instrument | Parameter | Platform

Analyse metadata

Terms not found - search in BioPortal and EarthPortal
Amplicon sequencing of Tara Oceans RNA samples corresponding to size fractions for protists.

particulate matter, including plankton (ENVO:xxxxxxxx)	IIx	http://www.isotc211.org/2005/resources/codeList.xml
2021-06-30T00:00:00Z	ARH	2014-10-13
Genome	Titanium	1.9378
Amplicon sequencing of Tara Oceans RNA samples corresponding to size fractions for protists.	PRJEB7315	Illumina
metagenome	http://www.isotc211.org/2005/resources/codeList.xml#CI_DateTypeCode	Analyzer
ISOTC211/19115	marine metagenome	marine

Document: **Amplicon sequencing of Tara Oceans RNA samples corresponding to size fractions for protists.** Export

Analyser Method: **Full XML Extraction**
No results found for method:
Analyser Method: **Structured XML Extraction**

SearchTerm	Match Concept	MatchPrope
Instrument (24 items)		

RESULTS OF THE FEDERATED SEARCH

Bioportal Search

Result	Property	Value
1	ID	http://www.projecthalo.com/aura#Titanium
	prefLabel	Titanium
	cui	N/A
	ontologyType	ONTOLOGY
2	ID	http://www.co-ode.org/ontologies/galen#Titanium
	prefLabel	Titanium
	cui	N/A
	ontologyType	ONTOLOGY
3	ID	http://purl.bioontology.org/ontology/LNC/LP16839-0
	prefLabel	Titanium
	cui	C0040302
	ontologyType	ONTOLOGY
4	ID	http://purl.bioontology.org/ontology/LNC/MTHU014132
	prefLabel	Titanium
	cui	C0040302
	ontologyType	ONTOLOGY
5	ID	http://purl.bioontology.org/ontology/RCD/X79AI
	prefLabel	Titanium
	cui	C0040302

Earthportal Search

Result	Property	Value
1	ID	http://www.eionet.europa.eu/gemet/concept/8495
	prefLabel	titanium
	cui	N/A
	ontologyType	ONTOLOGY
2	ID	https://vocabulary.actris.nilu.no/actris_vocab/titanium
	prefLabel	titanium
	cui	N/A
	ontologyType	ONTOLOGY
3	ID	http://sweetontology.net/matrMineral/Titanium
	prefLabel	Titanium
	cui	N/A
	ontologyType	ONTOLOGY
4	ID	https://w3id.org/ozcar-theia/c_ffd50577
	prefLabel	Titanium
	cui	N/A
	ontologyType	ONTOLOGY
5	ID	https://vocabulary.actris.nilu.no/actris_vocab/precipitationtitaniummassconcentration
	prefLabel	precipitation titanium mass concentration
	cui	N/A

HIGH LEVEL ANALYSIS OF RESULTS

PRELIMINARY RESULTS: INSTRUMENTS

National Oceanography Centre

Experimental Factor Ontology

Version 3.66.0

- braces
- centrifuge OBI
- computer OBI
- Fluidigm C1 microfluidics platform
- heating block OBI
- homogenizer OBI
- hybridization chamber OBI
- hybridization station OBI
- liquid handler OBI
- mass spectrometer OBI
- microarray wash station
- oligonucleotide synthesizer OBI
- probe design element
- sequencer (42)
  - DNA sequencer OBI (3)
  - high throughput sequencer (39)
  - long read sequencer
- sonicator OBI
- spectrophotometer OBI
- thermal cycler OBI
- vacuum dryer OBI
- vortexer OBI
- water bath OBI

instrument part (16)

The NERC Vocabulary Server (NVS)

R25	Argo sensor types	Terms describing sensor types mounted on Argo floats. Argo netCDF variable SENSOR is populated by R25 altLabel.
R26	Argo sensor manufacturers	Terms describing developers and manufacturers of sensors mounted on Argo floats. Argo netCDF variable SENSOR_MAKER is populated by R26 altLabel.
L05	SeaDataNet device categories	Terms used to classify groups of sensors, instruments, sources of algorithmically computed data (numerical models) or samplers (collectors of water, SPM, sediment, rock, air or biota samples).
L22	SeaVoX Device Catalogue	Terms for distinct sampling or measuring devices that may be identified in the real world in terms of manufacturer and model number.

- Not often populated
- Not many exact matches
- Many occurrences of

N/A

UNKNOWN

EARTHDATA
GCMD Keyword Viewer

Instruments

Search for...

- Instruments
  - Earth Remote Sensing Instruments
  - In Situ/Laboratory Instruments
  - Solar/Space Observing Instruments

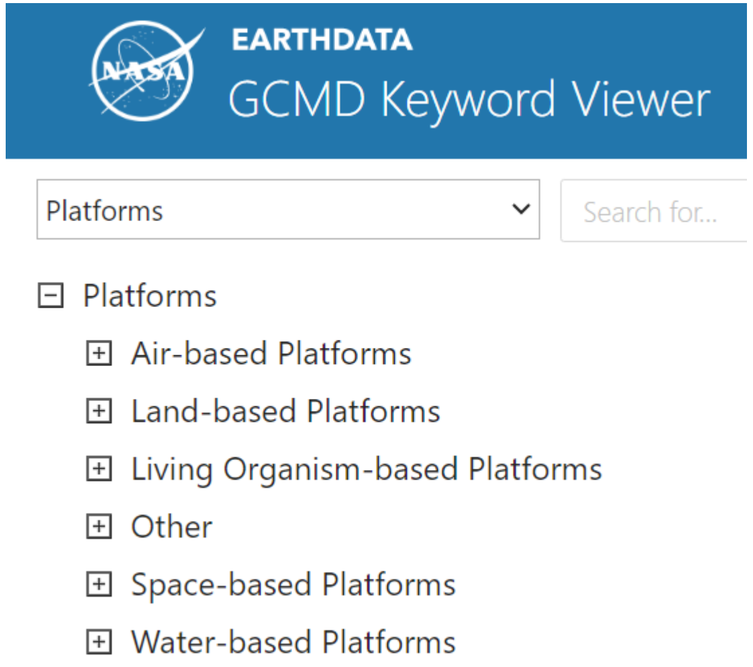
PLATFORM ANALYSIS

National Oceanography Centre

The NERC Vocabulary Server (NVS)

L06	SeaVoX Platform Categories	2-level grouping term hierarchy used for vehicles, structures or organisms capable of bearing instruments or tools for the collection of physical, chemical, geological or biological samples or data..
C17	ICES Platform Codes	Identifiers and metadata for platform instances (combinations of names and physical entities such as hulls or airframes).
R23	Argo platform type	List of Argo float types. Argo netCDF variable PLATFORM_TYPE is populated by R23 altLabel.

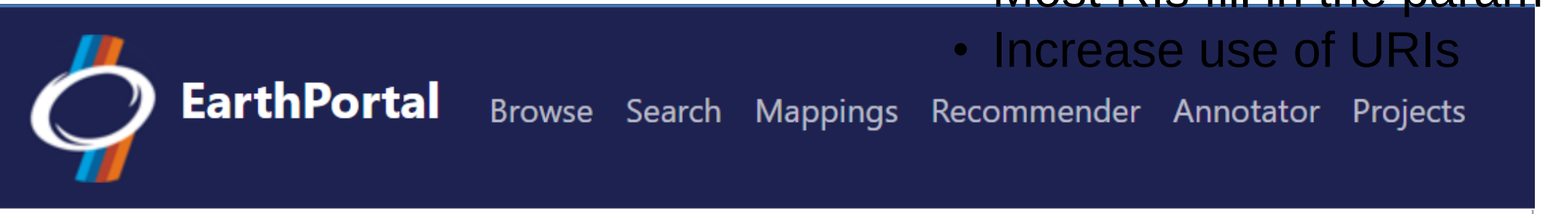
- Same as for instruments
- Few hits match “Satellite” “ships”
- Many occurrences of



The NERC Vocabulary Server (NVS)

P36	EMODnet Chemistry chemical groups
P35	EMODnet Chemistry aggregated parameter names
P01	BODC Parameter Usage Vocabulary
S26	BODC parameter semantic model matrices
P09	MEDATLAS Parameter Usage Vocabulary
R03	Argo parameter codes
L04	Matrix Categories
P07	Climate and Forecast Standard Names
P02	SeaDataNet Parameter Discovery Vocabulary

- Most RIs fill in the parameter category
- Increase use of URIs



Theia-OZCAR Thesaurus (TOZ)

concepts: 4,083 | FAIR score: 186

Thesaurus for in situ data from Environmental and Critical Zone Sciences

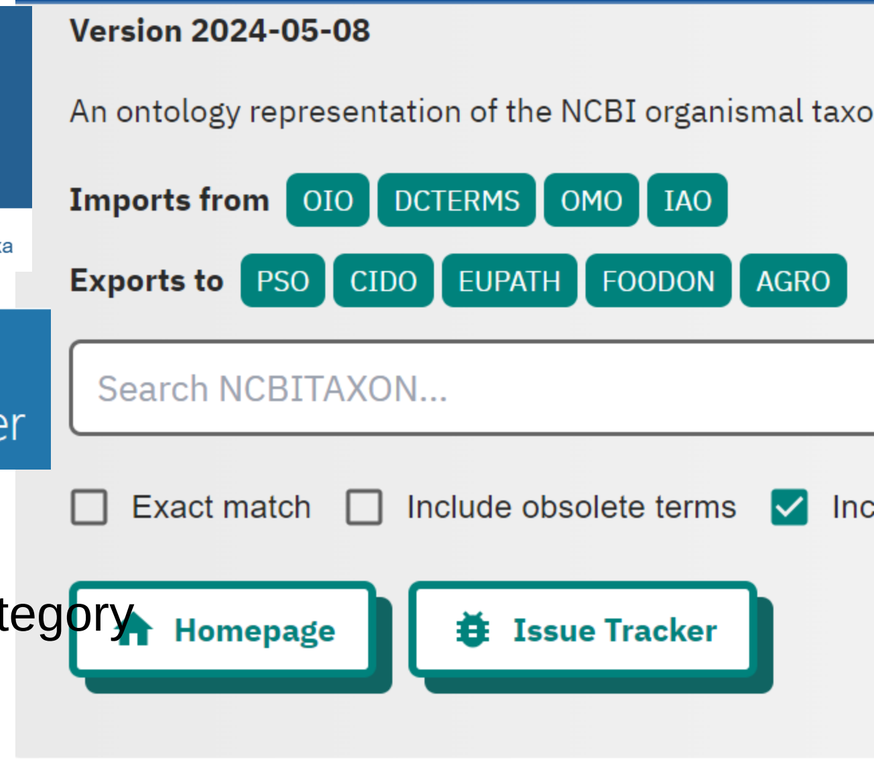
Uploaded: 5/8/24

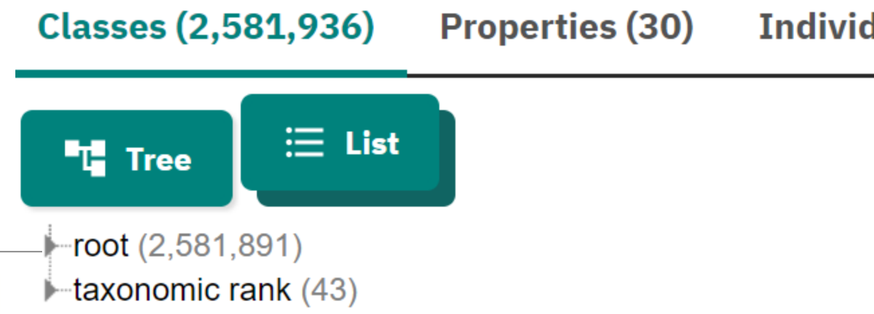
AERIS Parameters (AER_PRM)

concepts: 908 | FAIR score: 198 | projects: 1

Parameters thesaurus of the French cluser for atmospheric data and services AERIS.

Uploaded: 9/15/23





PARAMETERS

LESSONS LEARNT

National Oceanography Centre

- **Use URIs and prefLabels to reference semantic artefacts in annotations**
  - Ensure each term is uniquely identified
- **Versioning of Semantic Artefacts**
  - Versioning practices are inconsistent across different repositories, complicating tracking and updates.
  - Need for semantic artefact versioning consensus
  - Update frequency: Optimal when in sync with the update
- **SPARQL Endpoints Reliability**
  - Not all SPARQL endpoints are reliably operational, affecting data retrieval and query execution.
- **Interoperability of Widely Used Semantic Artefacts**
  - Many widely adopted artefacts ensure interoperability using standards like SKOS, OWL, and RDF.
  - Adhering to the standards does not necessarily achieve interoperability
- **Non-FAIR Artefacts**
  - Some artefacts are still provided as PDF lists, which hinders mapping and interoperability efforts.
- **From Machine readable to machine actionable**

HOW TO REFER TO SEMANTIC ARTEFACTS

```
<gmd:attributeDescription>

<gco:RecordType xlink:href="http://vocab.nerc.ac.uk/collection/P02/current/ALKY/" xlink:title="Alkalinity, acidity and pH of the water column">Alkalinity, acidity and pH of the water column</gco:RecordType>

</gmd:attributeDescription>

<gmd:contentType>

<gmd:MD_CoverageContentTypeCode codeList="http://www.isotc211.org/2005/resources/codeList.xml#MD_CoverageContentTypeCode" codeListValue="physicalMeasurement" codeSpace="ISOTC211/19115">physicalMeasurement</gmd:MD_CoverageContentTypeCode>

</gmd:contentType>
```

XML-ISO19115

```
{
 "@type": "DefinedTerm",
"name": " Alkalinity, acidity and pH of the water column",
"inDefinedTermSet":"http://vocab.nerc.ac.uk/collection/P02/current/",
"url": "http://vocab.nerc.ac.uk/collection/P02/current/ALKY/",
"identifier":
{ "@type": "PropertyValue",
 "propertyID": "http://vocab.nerc.ac.uk/",
"value": "ALKY",
"url": " http://vocab.nerc.ac.uk/collection/P02/current/“

 }
```

json-LD: schema.org

WHERE NEXT?

- Enhance the knowledge base with additional semantic resources utilizing SA feedback
- Improve search methods
- Enhance user experience (UX)
- Working with the BlueCloud project
  - Provide useful reports and recommendations
  - FAIR: I2. (Meta)data use vocabularies that follow FAIR principles and reference them uniquely

27-29 May 2024

International conference on Marine Data and Information Systems

SeaDataNet AISBL

NAVFORSKNINGS INSTITUTTET

Ifremer

MARIS

MEDIN

National Oceanography Centre

INGV

iMagine

ICDE

eosc Blue-Cloud2026