

Further exploiting our vocabularies

ALEXANDRA KOKKINAKI

Vocabulary Management Group at BODC
National Oceanography Centre, Liverpool and Southampton

Presented at BODC 50th, 8 May 2019, Liverpool



**National
Oceanography Centre**
NATURAL ENVIRONMENT RESEARCH COUNCIL

noc.ac.uk

NERC SCIENCE OF THE
ENVIRONMENT

FAIR Data

In general terms, your research data should be 'FAIR', that is

Findable,
Accessible,
Interoperable and
Re-usable.

- “H2020 Programme Guidelines on FAIR Data Management” Wilkinson, Mark D., Michel Dumontier, IJsbrand Jan Aalbersberg, Gabrielle Appleton, Myles Axton, Arie Baak, Niklas Blomberg et al. "The FAIR Guiding Principles for scientific data management and stewardship." *Scientific Data* 3 (2016).

The FAIR Guiding Principles for scientific data management and stewardship

Mark D. Wilkinson, Michel Dumontier [...] Barend Mons [✉](#)

Scientific Data 3, Article number: 160018 (2016) | [Download Citation](#) ↓

i An Addendum to this article was published on 19 March 2019

Abstract

There is an urgent need to improve the infrastructure supporting the reuse of scholarly data. A diverse set of stakeholders—representing academia, industry, funding agencies, and scholarly publishers—have come together to design and jointly endorse a concise and measurable

- **Findable**

- Persistent ID
- Metadata online

- **Accessible**

- Data online
- Restrictions where needed

- **Interoperable**

- Use standards, controlled vocabs
- Common (open) formats

- **Reusable**

- Rich documentation
- Clear usage licence

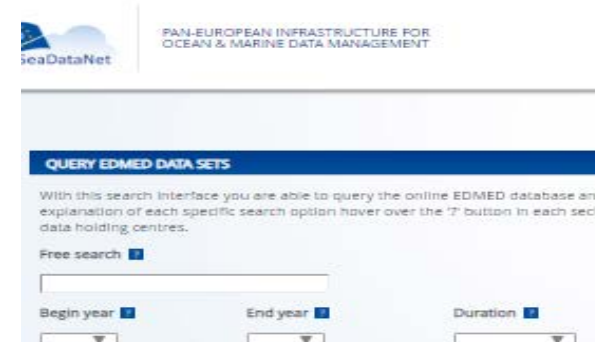
FAIR data checklist



European Directory of Marine Environmental Data (EDMED)

“Is a comprehensive reference to the marine data sets and collections held within **European research laboratories**, so as to provide marine **scientists, engineers** and **policy makers** with a simple mechanism for their identification.”

- Over 4200 data sets
- Over 700 organisations
- Annotated NERC Vocabulary Server (NVS)
- Submitted to BODC
- As XML records



EDMED metadata format, based on ISO 19115

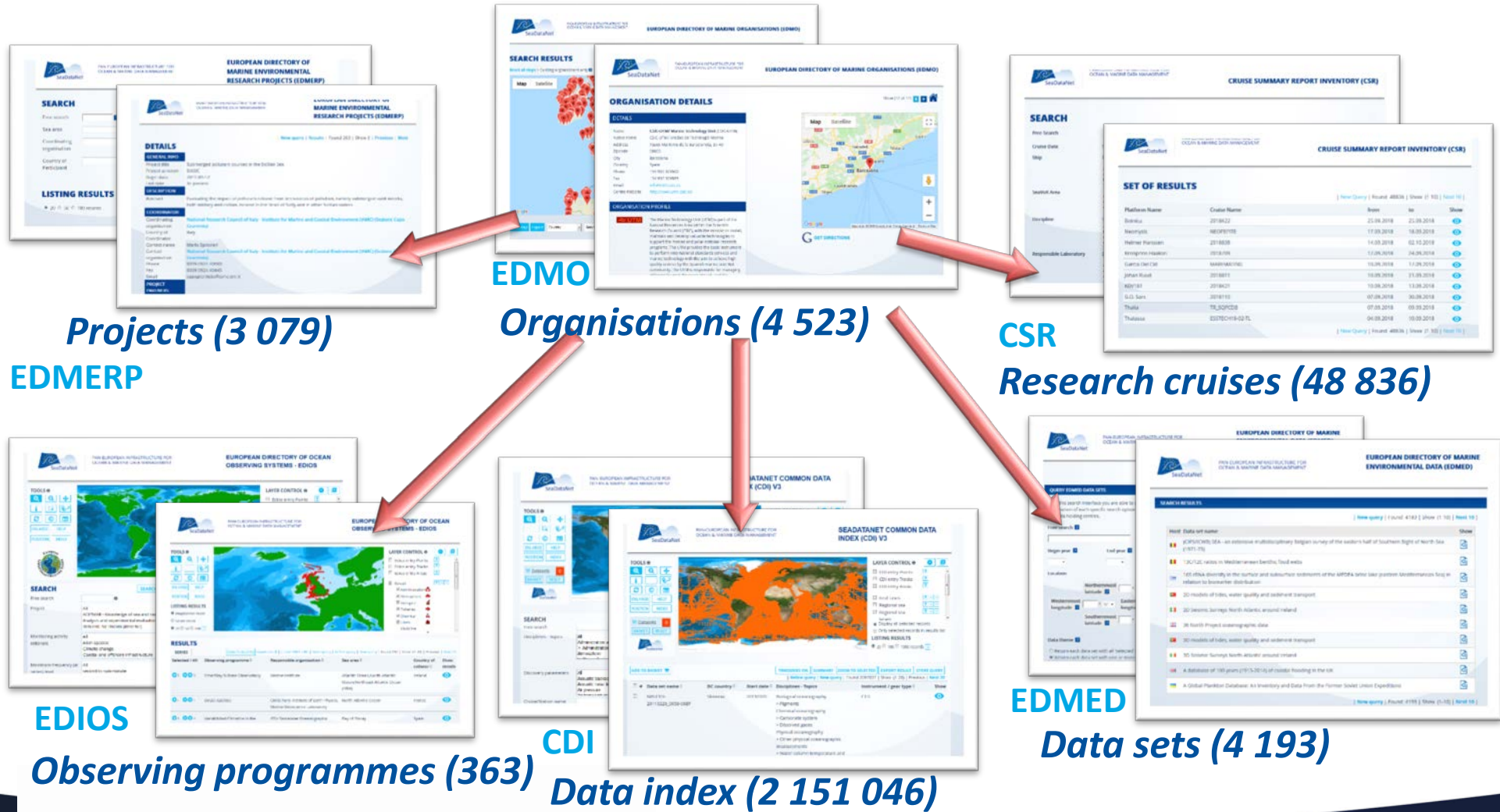
SEARCH RESULTS

| New query | Found: 4203 | Show (1-10) | Next 10 |

Host	Data set name	Show
	(CIPS/ICWB) SEA - an extensive multidisciplinary Belgian survey of the eastern half of Southern Bight of North Sea (1971-75)	
	¹³ C/ ¹² C ratios in Mediterranean benthic food webs	
	16S rRNA diversity in the surface and subsurface sediments of the MEDEA brine lake (eastern Mediterranean Sea) in relation to biomarker distribution	
	2015 Atlantic Margin Oil and Gas Exploration Licensing Round Awards	
	2D models of tides, water quality and sediment transport	
	2D Seismic Surveys Irish Offshore 1965-2013	
	30 North Project: oceanographic data	
	3D models of tides, water quality and sediment transport	
	3D Seismic Surveys Irish Offshore 1982-2014	
	A database of 100 years (1915-2014) of coastal flooding in the UK	

| New query | Found: 4203 | Show (1-10) | Next 10 |

SEADATANET METADATA SERVICES



QUERY EDMED DATA SETS

With this search interface you are able to search the online EDMED database and display the result of your query on screen. For a detailed explanation of each specific search criteria, please refer to the EDMED metadata format, based on ISO 19115.

Free search [?](#)

Begin year [?](#)

End year [?](#)

Location [?](#)

Northernmost latitude [?](#)

Westernmost longitude [?](#)

W

Southernmost latitude [?](#)

Data theme [?](#)

- Return each data set with all 'selected' data sets
- Return each data set with one or more 'selected' data sets

Administration and dimensions

- Administration and dimensions - D
- Administration and dimensions - E
- Administration and dimensions - H
- Administration and dimensions - H
- Administration and dimensions - M
- Administration and dimensions - M
- Administration and dimensions - P
- Administration and dimensions - Q
- Administration and dimensions - R
- Administration and dimensions - U

Instrument theme [?](#)

GENERAL

Data set name

Atlantic meridional overturning circulation observed by the RAPID-MOCHA-WBTS array at 26°N from 2004 to 2017

Data holding centre

British Oceanographic Data Centre

Country

United Kingdom

Project

Rapid Climate Change

Time period

01 April 2004 to 28 February 2017

Ongoing

No

Geographical area

North Atlantic across 26.5N from

OBSERVATIONS

Parameters

Vertical spatial coordinates; Depth of the water column; Temperature

Instruments

Current profilers; submarine cables; CTD; current meters

DESCRIPTION

Summary

The RAPID-MOCHA-WBTS dataset comprises measurements of current velocity, temperature, salinity and pressure. Oceanic volume transports are calculated from these variables ranging in continuous measurements of the Atlantic Meridional Overturning Circulation (AMOC). Data collection is obtained from a mooring array across 26.5N in the Atlantic Ocean and cable measurements across the Florida Straits. The measurement array extends from the Bahamas to the African coast. The data have been measured continuously between April 2004 and February 2017. The data are collected periodically (currently every 18 months) from various UK and USA research cruises. Measurements between the Bahamas and Africa were made using a variety of MicroCat CTD sensors, various current meters and ADCP. All instruments are located on 18 moorings in various locations at 26.5N. An undersea cable makes current velocity measurements across the Florida Straits. The RAPID-MOCHA-WBTS programme aims to deliver a multi-decadal time series of observations of AMOC. The observations will be used with data from other sources to determine and interpret recent changes in the AMOC, to assess the risk of rapid climate change due to changes in the MOC, and to investigate the potential for predicting the MOC and its impacts on climate. The RAPID-MOCHA-WBTS programme is a joint effort between NERC in the UK (the UK Principal Investigator is David Smeed), NOAA (Chris Meinen) and RSMAS (Prof. Bill Johns) in the USA. The Atlantic MOC transport (and its components), calculated from the above data, and

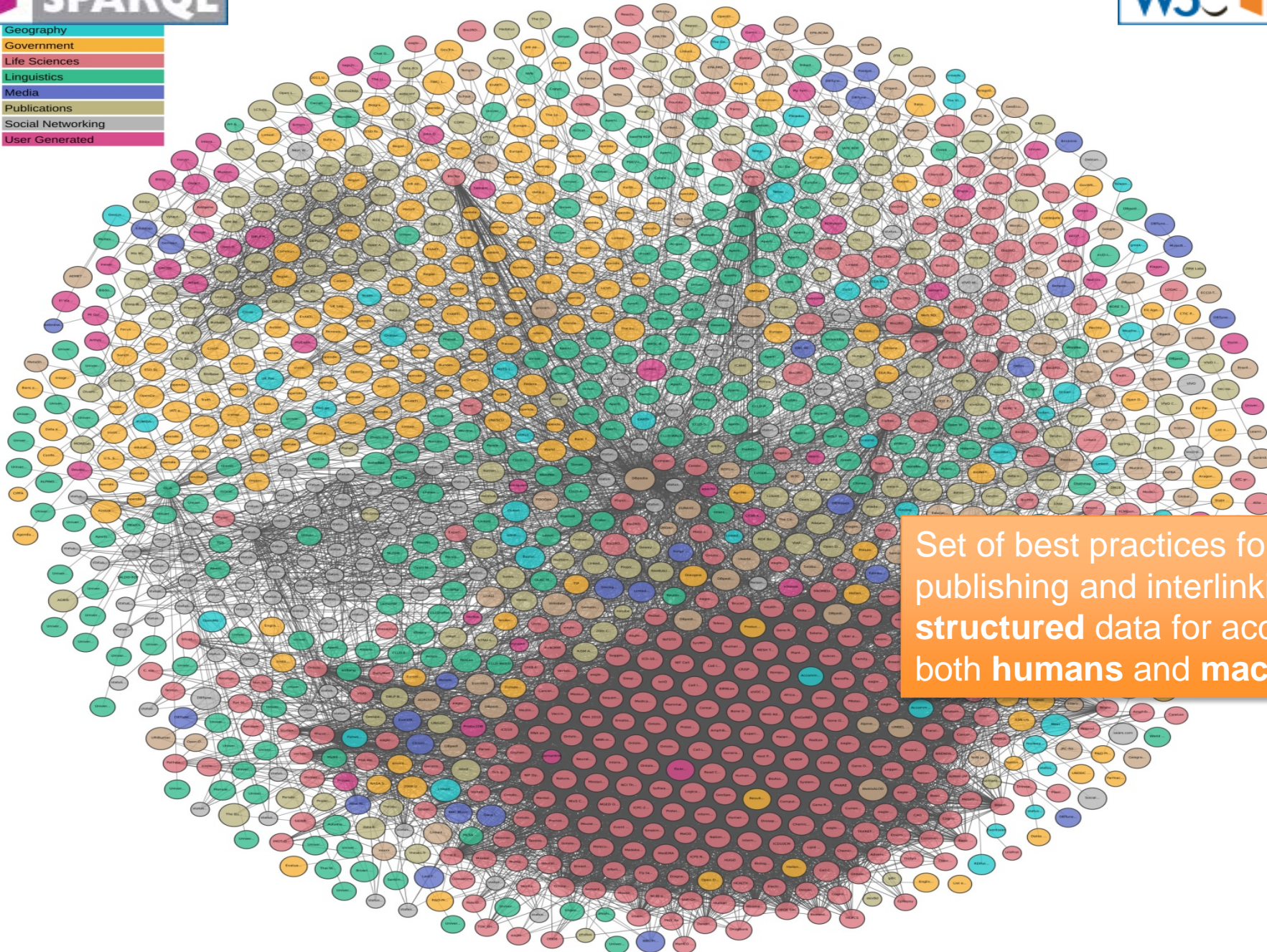
EDMED metadata format, based on ISO 19115

Annotated NERC Vocabulary Server (NVS)

MACHINE READABLE

HUMAN

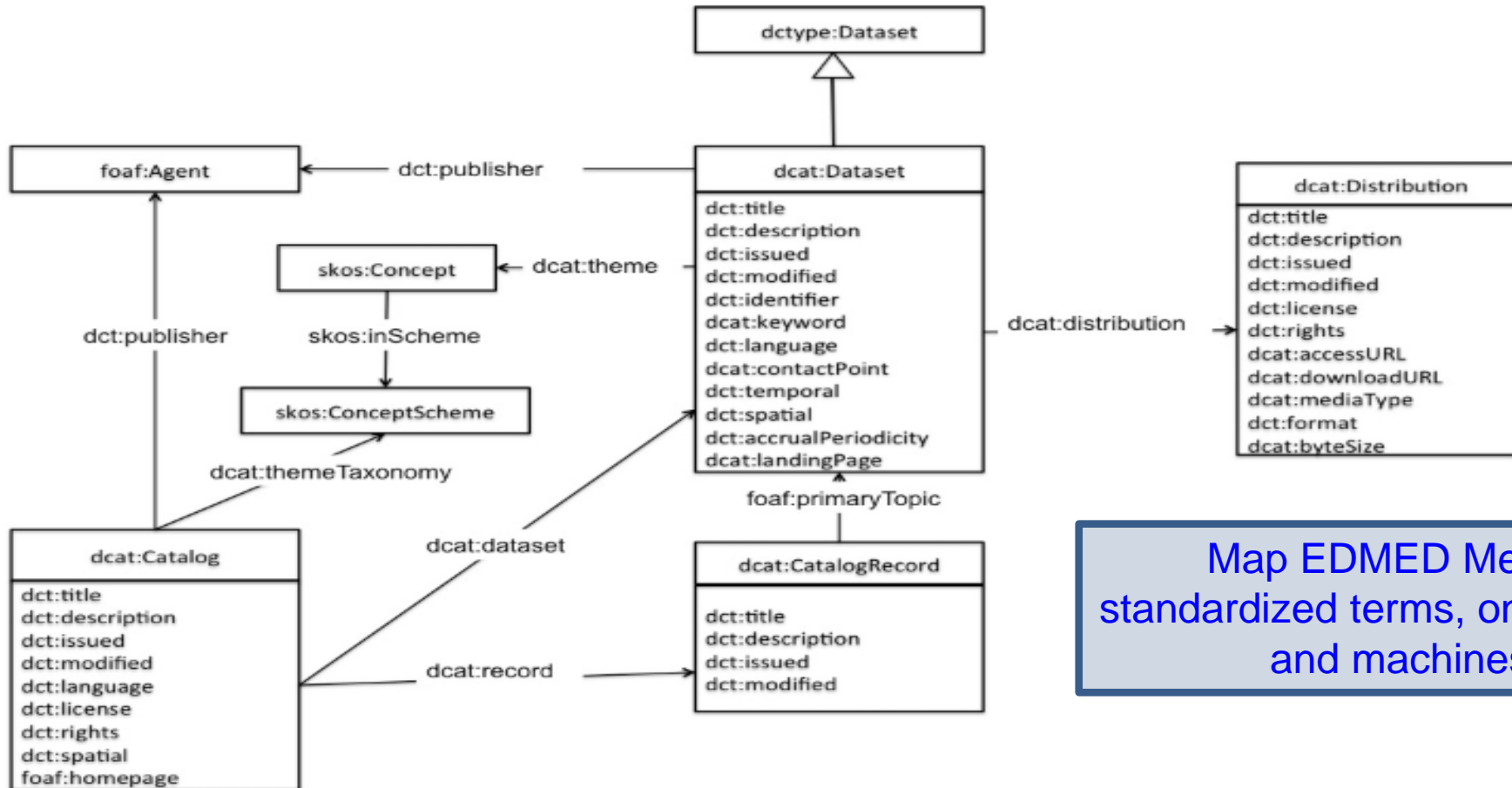
- Geography
- Government
- Life Sciences
- Linguistics
- Media
- Publications
- Social Networking
- User Generated



Set of best practices for publishing and interlinking structured data for access by both humans and machines

Data Catalog Vocabulary (DCAT)

W3C Recommendation



Map EDMED Metadata format to standardized terms, ontologies, that humans and machines understand

Thomas and Leadbetter, 2018, Exposing the SDN metadata catalogues as Linked Data: from metadata to data on the web, EGU, Vienna

[Query EDMED data sets](#) | [Query EDMED organisations](#) | [SPARQL endpoint](#)

SPARQL ENDPOINT

SPARQL query

```
select ?EDMEDRecord ?Title where {  
  ?EDMEDRecord a  
  <http://www.w3.org/ns/dcat#Dataset> ; <http://purl.org/dc/terms/title> ?  
  Title .}
```

Output:

If XML output, add XSLT style sheet (blank for none):

Force the accept header to text/plain regardless.

SEARCH

RESET

- machine-actionable
 - machine readable
 - machine comprehensible
- Interoperable
- Data mining/AI
- Knowledge



Wood et al, 2018, Exposing the SeaDataNet metadata catalogues via SPARQL endpoints, IMDIS Conf, Barcelona.

How will you discover EDMED?

Search engine for datasets **Over metadata** provided by data providers

Search engine to tell the user **where** the data is

Google Dataset Search Beta

Search for Data Sets

Try [boston education data](#) or [weather site:noaa.gov](#)

[Find out more](#) about including your datasets in Dataset Search.

Leadbetter et al, 2018, The place of schema.org in Linked Ocean Data, IMDIS Conf, Barcelona.

Yahoo, Microsoft, Google, Yandex

Welcome to Schema.org

Schema.org is a collaborative, community activity with a mission to create, maintain, and promote schemas for structured data on the Internet, on web pages, in email messages, and beyond.

Schema.org vocabulary can be used with many different encodings, including RDFa, Microdata and JSON-LD. These vocabularies cover entities, relationships between entities and actions, and can easily be extended through a well-documented extension model. Over 10 million sites use Schema.org to markup their web pages and email messages. Many applications from Google, Microsoft, Pinterest, Yandex and others already use these vocabularies to power rich, extensible experiences.

Founded by Google, Microsoft, Yahoo and Yandex, Schema.org vocabularies are developed by an open [community](#) process, using the public-schemaorg@w3.org mailing list and through [GitHub](#).

A shared vocabulary makes it easier for webmasters and developers to decide on a schema and get the maximum benefit for their efforts. It is in this spirit that the founders, together with the larger community have come together – to provide a shared collection of schemas.

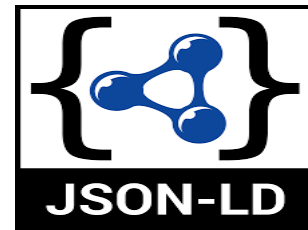
We invite you to [get started!](#)

View our blog at blog.schema.org or see [release history](#) for version 3.6.


```

"@context": "http://schema.org",
"@type": "Dataset",
"name": "ARABESQUE Project Data Set - upper ocean biogeochemistry data collected in the Arabian Sea and the Gulf of Oman in 1994",
"description": "The data set comprises hydrographic, biogeochemical and biological data, including measurements of temperature, salinity and attenuation, plus co-
hydrocarbons, sedimentation flux, sulphur and dissolved carbon. Analyses of bacterial, zooplankton and phytoplankton communities were also undertaken. The ocean
meteorological parameters. Data were collected across three repeated sections: one along the Gulf of Oman; a section at 67deg East from 8 to 14.5deg North; and
Other one-off sections were also traversed in the Arabian Sea and Gulf of Oman areas. Measurements were collected during two cruises: one between 27 August and
December 1994. Sections were covered by underway surface ocean measurements (one minute sampling of multiple parameters providing some 5 million measurements) c
were repeated. ARABESQUE was organised by the Plymouth Marine Laboratory of NERC's Centre for Coastal and Marine Sciences and involved the University of Wales,
University of Edinburgh; University of Newcastle; the Bedford Institute of Oceanography, Canada; the Max Planck Institute for Limnology, Germany and the Sultan
provided by the British Oceanographic Data Centre. All data collected as part of the project were lodged with BODC who had responsibility for assembling, calib
checked for instrument spikes or malfunction, values beyond the calibration range, unreasonable ratios of chemical constituents and unreasonable deviations from
database, complete with supporting documentation and a user manual. The full data set has been published by BODC on CD-ROM complete with user interface.",
"temporalCoverage": "1994/1994",
"version": "20170523",
"url": "https://www.bodc.ac.uk/resources/inventories/edmed/report/192/",
"keywords": [
  "Ammonium and ammonia concentration parameters in water bodies",
  "Transmittance and attenuation of the water column",
  "Bacteria generic abundance in water bodies",
  "Carotenoid and flavenoid pigment concentrations in water bodies",
  "Phytoplankton taxonomic biomass in water bodies",
  "Chlorophyll pigment concentrations in water bodies",
  "Organosulphur and organoselenium species concentration parameters in water bodies",
  "Dissolved organic carbon concentration in the water column",
  "Dissolved oxygen parameters in the water column",
  "Microzooplankton taxonomic abundance in water bodies",
  "Sedimentation flux quantification in the water column",
  "Microzooplankton non taxonomy-related biomass expressed as carbon per unit volume of the water column",
  "Microzooplankton generic abundance in water bodies",
  "Microzooplankton grazing",
  "Particulate total and organic nitrogen concentrations in the water column",
  "Nitrate concentration parameters in the water column",
  "Dissolved concentration parameters for other gases in sediment pore waters",
  "Concentration of other hydrocarbons in the water column",
  "Phytoplankton taxonomic abundance in water bodies",
  "Partial pressure (pCO2) and fugacity (fCO2) of carbon dioxide in the water column",
  "Phosphate concentration parameters in the water column",
  "Phaeopigment concentrations in the water column",
  "Phycobolin pigment concentrations in the water column",
  "Primary production in the water column",
  "Salinity of the water column",
  "Plankton abundance per unit volume of the water column",
  "Molecular biology parameters",
  "Silicate concentration parameters in the water column",

```



71 results found


 Liverpool Bay/Irish Sea Coastal Observatory Data Set
www.bodc.ac.uk

 Coupled carbon-climate Earth System Model GFDL ESM2M...
www.bodc.ac.uk

 World Ocean Isopycnal-Level Velocity Inverted from GDE...
www.bodc.ac.uk

 Seabed profile, CTD and sediment data collected fro...
www.bodc.ac.uk
edmed.seadatanet.org

 Time series dataset from the Plankton Monitoring...
www.bodc.ac.uk

 Netherlands National Monitoring of bathymetric...
www.bodc.ac.uk



Liverpool Bay/Irish Sea Coastal Observatory Data Set

Dataset name

 European Directory of Marine Environmental Dataset...

Data set provided by

- [Met Office](#)
- [National Oceanography Centre, Southampton](#)
- [Plymouth Marine Laboratory](#)
- [Proudman Oceanographic Laboratory](#)
- [National Oceanography Centre](#)
- University of Liverpool, Port Erin Marine Laboratory
- Centre for Environment, Fisheries and Aquaculture Science, Lowestoft Laboratory
- Environment Agency Head Office
- University of Liverpool Department of Earth and Ocean Sciences
- University of Wales, School of Ocean Sciences
- Bangor University School of Ocean Sciences
- Department of Agriculture and Rural Development for Northern Ireland

Provider

Licence

[SeaDataNet licence](#)

License <http://vocab.nerc.ac.uk/collection/L08/current/LS/>

Time period covered

Jan 1, 2001 - Sep 20, 2011

Temporal coverage

Area covered

Irish Sea

Spatial coverage

Description

The data set comprises measurement of physical and biogeochemical oceanographic parameters and complementary meteorological parameters collected during the Liverpool Bay/Irish Sea Coastal Observatory initiative. It includes measurements from across the Liverpool Bay and Irish Sea area with data collection spanning a decade from 2001 to 2011. It incorporates, for example, hydrographic survey cruises (typically 8 - 10 per year) undertaken by the RV Prince Madog, data collected via instrumented ferries, time series data from oceanographic moorings and at two meteorological stations, namely Proudman Oceanographic Observatory (up to 2004) and Hilbre Island (2004 - 2011), and a shore-based high-frequency (HF) radar measuring waves and surface currents out to a range of 50km. The hydrographic surveys include conductivity-temperature-depth (CTD) casts with attached auxiliary sensors and data collected via the ships' underway monitoring system. Oceanographic parameters include temperature, salinity, dissolved oxygen, attenuation, turbidity, fluorescence, chlorophyll, nutrients, irradiance, waves and currents and meteorological parameters include air temperature, air pressure, wind velocities, humidity, precipitation and atmospheric irradiance. The instrumented ferries also incorporated an underway monitoring system for sea surface properties. The Observatory integrated (near) real-time measurements with the POLCOMS (Proudman Oceanographic Laboratory Coastal Ocean Modelling System) models. The objective was to understand a coastal sea's response both to natural forcing and to the consequences of human activity. The foci were the impact of storms, variations in

Dataset description

Where next

- Structured metadata contain free text
- Ideal feed NVS vocabularies in schema.org
- Adding a layer of semantics in google dataset search



ROSETTA STONE SERVICE



**National
Oceanography Centre**
NATURAL ENVIRONMENT RESEARCH COUNCIL

noc.ac.uk

NERC SCIENCE OF THE
ENVIRONMENT

Ocean Data Interoperability Platform (ODIP II)

- Successor to ODIP (October 2012 – September 2015)
 - The fundamental objective of the Ocean Data Interoperability Platform project is to support development of a common global framework for marine data management to bridge the gaps between regional data infrastructures in Europe, USA and Australia.
- Grant No: 654310
- Call: FP7-INFRA-SUPP-6-2014 – International co-operation for research infrastructures
- Addressing specific challenge: to support multilateral cooperation on research infrastructures in marine science
- ODIP II: April 2015 – March 2018

Administration and dimensions
 > Administration and dimensions
 Atmosphere
 > Atmospheric chemistry

Discovery parameters

All
 Acoustic backscatter in the
 Acoustic noise in the
 Air pressure
 Air temperature

Cruise/Station name

Project

Datas

Sea re

Waterdepth (m) from

Originator

CDI partner

Country

Access restricti

commercial charge

Instrument type

Instrument depth (m) from

Platform type

Measuring area type

Temporal resolution

Date (yyyymmdd) from



Duration

Unit

NERC Vocabulary Server

AODN vocabularies

NODC vocabularies



ODIP INTEROPERABILITY SCOPE



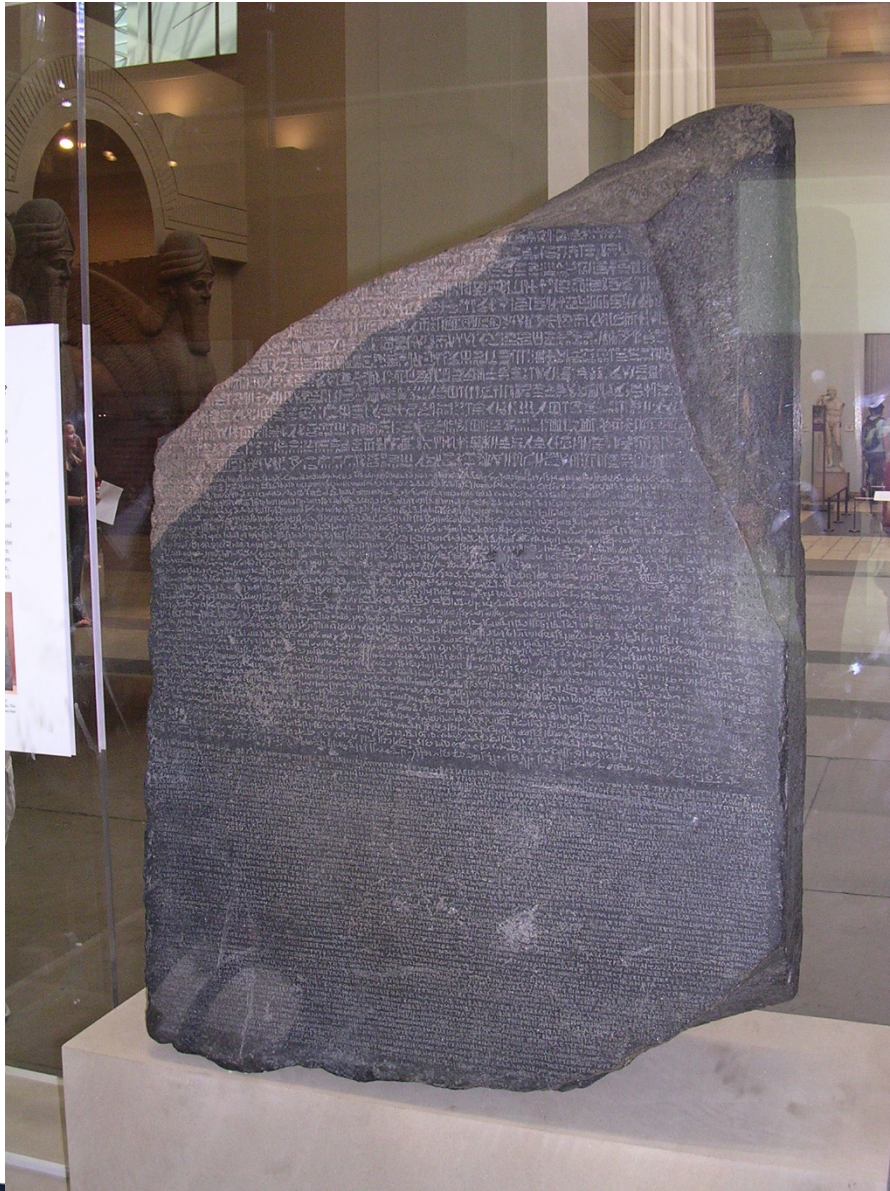
Scope of ODIP is interconnect marine communities worldwide.

Global users accessing data from regional data providers in EU, US and Australia



Global Users





Hieroglyphics

Used for important and religious documents



Egyptian writing (demotic)

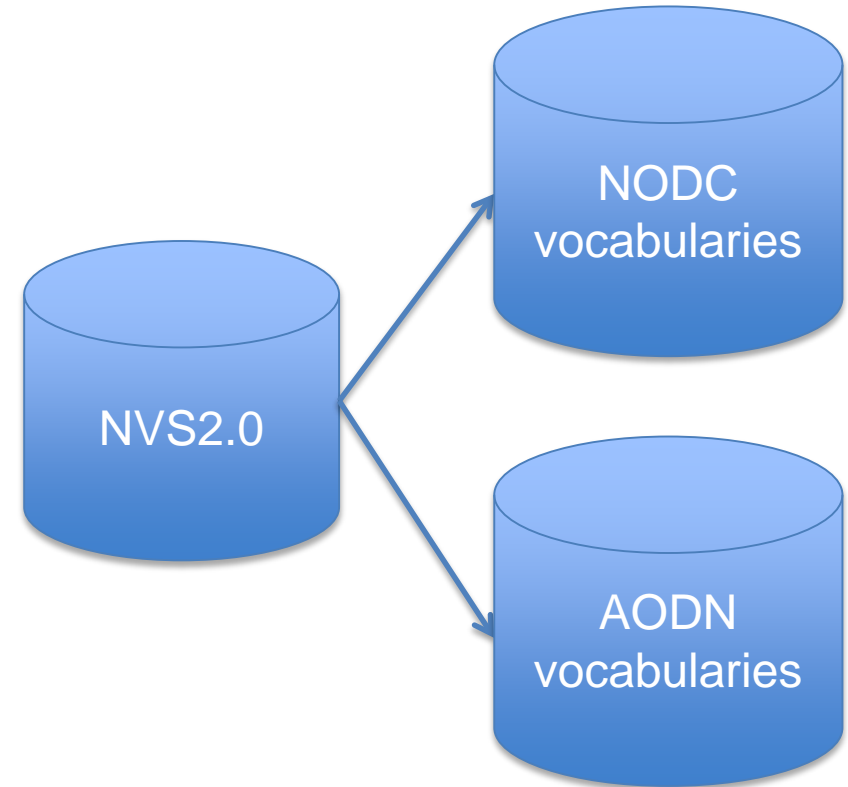
Used by most people in Egypt at that time.



Greek writing

Used by the rulers

What?



How



Bring all the terms that are the same as NODC:373 (**Sea surface height above geoid**)

prefix skos:<<http://www.w3.org/2004/02/skos/core#>>

prefix rdf:<<http://www.w3.org/1999/02/22-rdf-syntax-ns#>>

prefix owl:<<http://www.w3.org/2002/07/owl#>>

prefix dc:<<http://purl.org/dc/terms/>>

select * where

{?NVSTerm owl:sameAs <http://vocab.aodn.org.au/def/discovery_parameter/entity/643>.

?NVSTerm owl:sameAs ?Translation

}

NVSTerm

| Translation

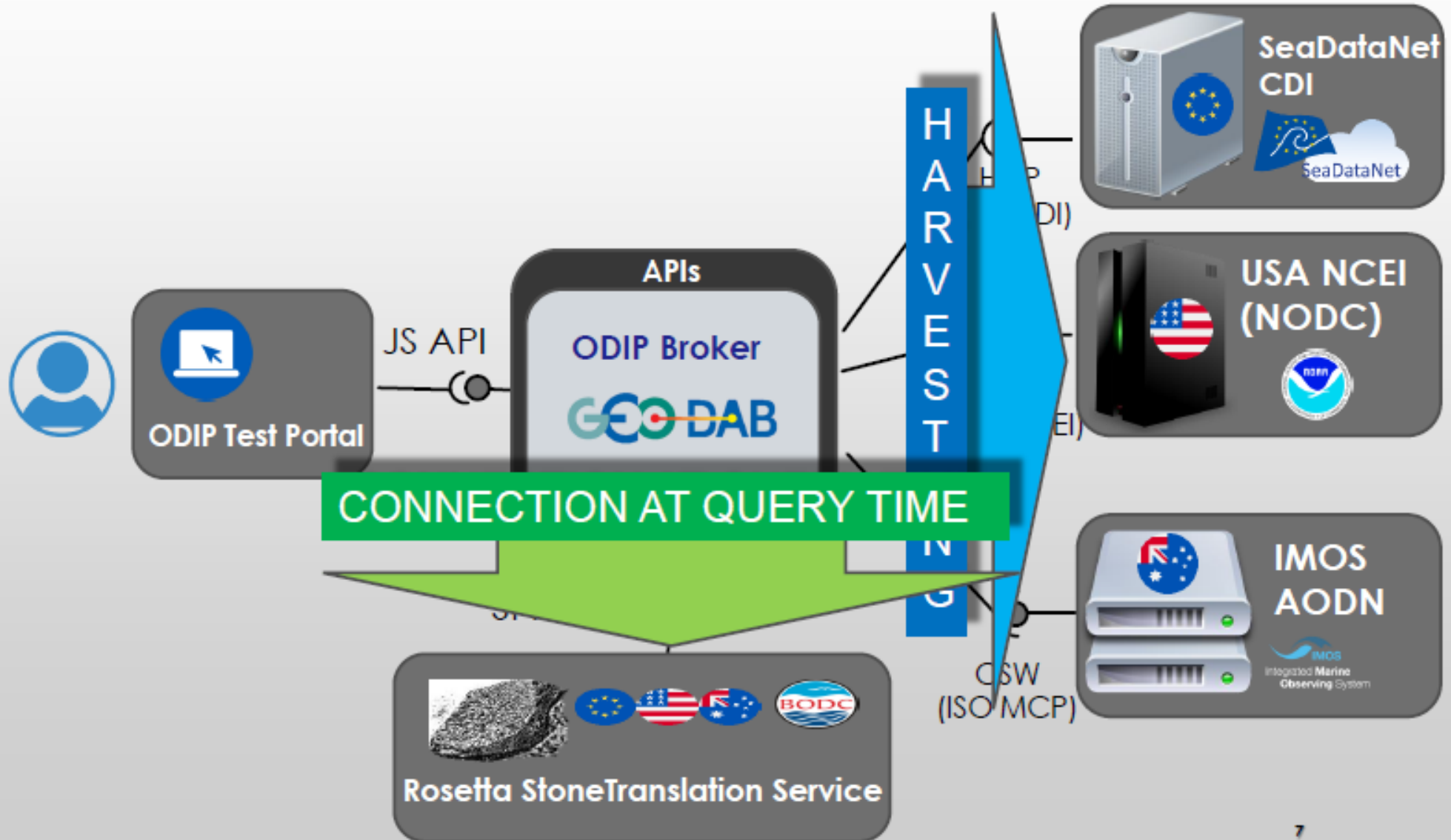
|

=====|
<<http://vocab.nerc.ac.uk/collection/P02/current/ASLV/>> | <<https://www.nodc.noaa.gov/cgi-bin/OAS/prd/datatype/details/312>>

| <<http://vocab.nerc.ac.uk/collection/P02/current/ASLV/>> | <http://vocab.aodn.org.au/def/discovery_parameter/entity/643>



BROKERING PROTOTYPE – SEMANTIC DISCOVERY



SEARCH

salinity

Start time

End time

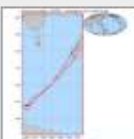
Advanced



RESULTS FILTERS SOURCES BROWSING

Matching results: 8.698

1 2 3 4 5



Partial pressure (or fugacity) of carbon dioxide, temperature, salinity and other variables collected from Surface underway observations using Carbon dioxide (CO₂) gas analyzer and other instruments from TANGAROA in the Indian Ocean, South Pacific Ocean and others from 1999-02-02 to 1999-02-28 (NCEI Accession 0155958)

NCEI Accession 0155958 includes Surface underway, chemical, meteorological and physical data collected from TANGAROA in the Indian Ocean, South Pacific Ocean, Southern Oceans (> 60 degrees South) and Tasman Sea from 1999-02-02 to 1999-02-28. These data include AIR-SEA DIFFERENCE - PARTIAL PRESSURE (OR FUGACITY) OF CARBON DIOXIDE, BAROMETRIC //



Real-time current, wave, temperature, salinity, and meteorological data from Gulf of Maine Ocean Observing System (GoMOOS) buoys, 5/22/2005 - 5/28/2005 (NODC Accession 0002212)

No description available



Dissolved Inorganic carbon, pH, alkalinity, temperature, salinity and other variables collected from discrete sample and profile observations using CTD, Coulometer for DIC measurement and other instruments from the SHUMPU MARU in the Philippine Sea from 1998-08-19 to 1998-08-27 (NODC Accession 0112304)

NODC Accession 0112304 includes biological, chemical, discrete sample, physical and profile data collected from SHUMPU MARU in the Philippine Sea from 1998-08-19 to 1998-08-27 and retrieved during cruise PACIFICA_49SU19980819 and SH98-08. These data include ALKALINITY, AMMONIUM, CHLOROFLUOROCARBON-11 (CFC-11), CHLOROFLUOROCARBON-12 (CFC- //

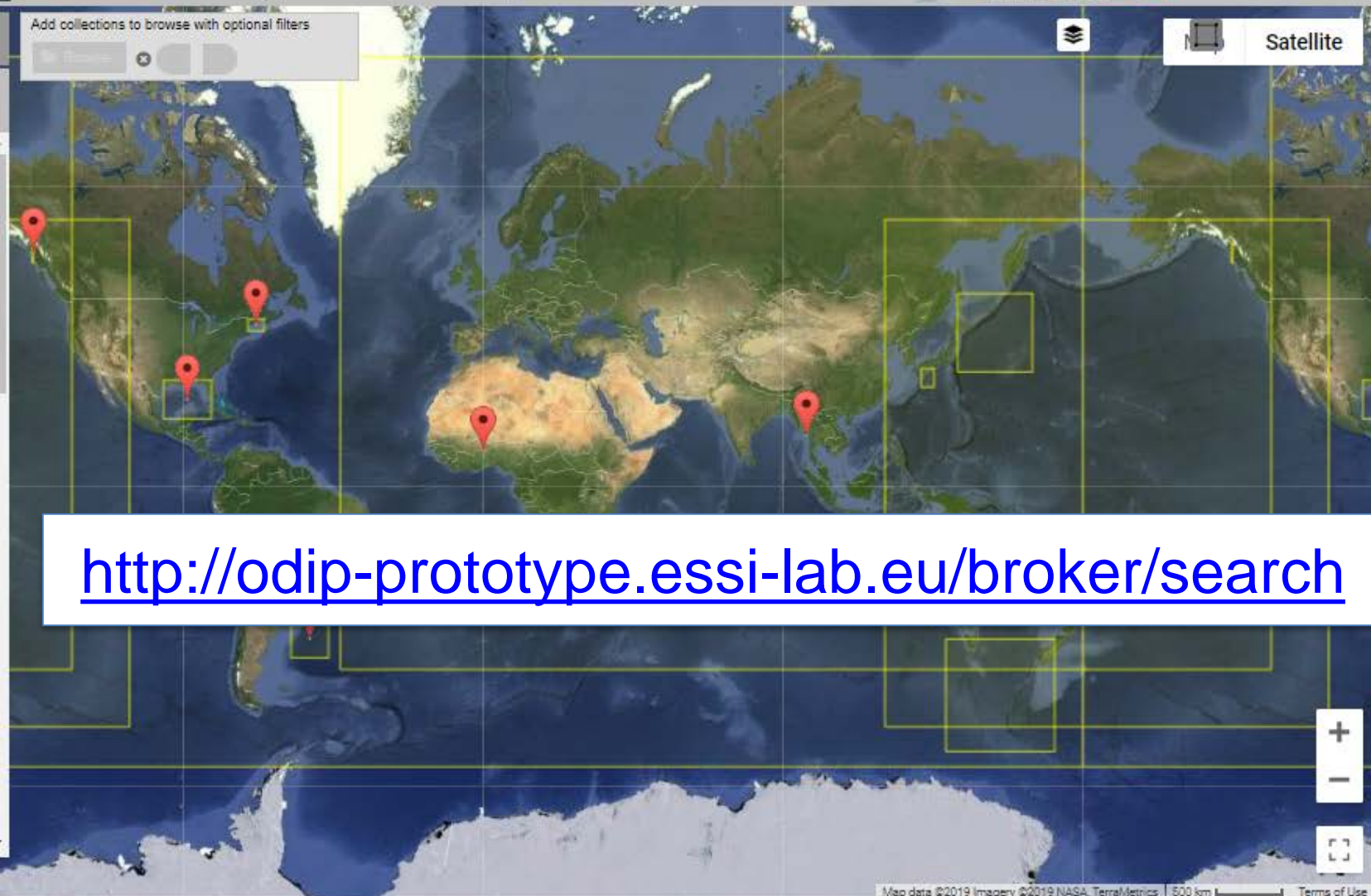


Temperature and salinity profile data from globally distributed Argo profiling floats for the week of 2006-10-08 for the Global Argo Data Repository, data ranged from 2005-04-19 to 2006-10-14 (NODC Accession 0010464)

IMAGE NOT FOUND

The U.S. National Oceanographic Data Center (NODC) operates the Global Argo Data Repository

Add collections to browse with optional filters



<http://odip-prototype.essi-lab.eu/broker/search>

Google

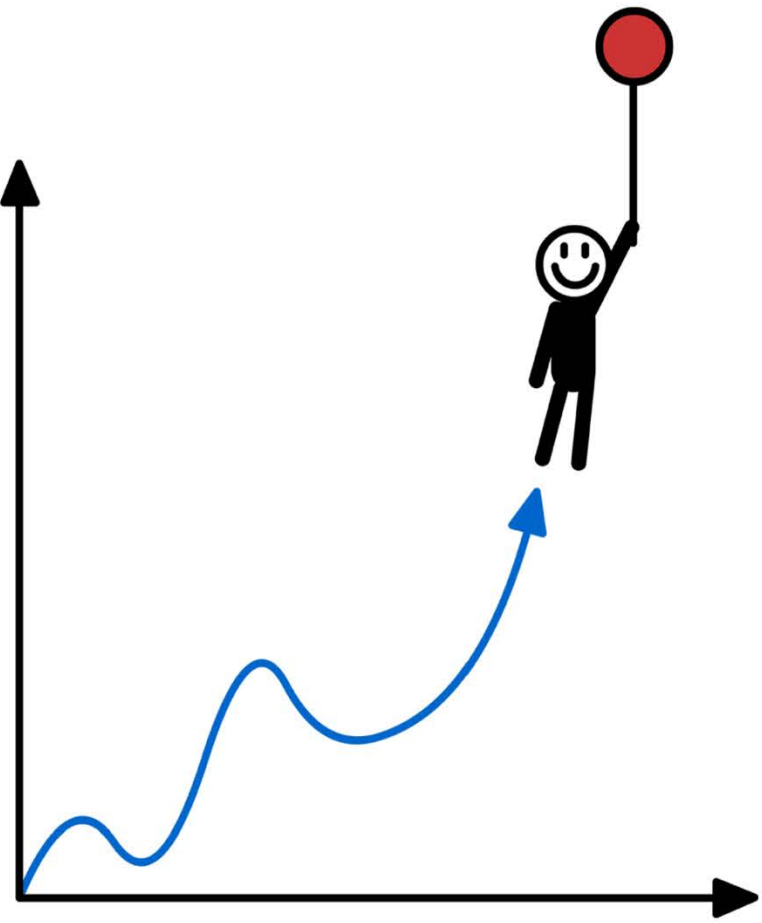
Map data ©2019 Imagery ©2019 NASA, TerraMetrics | 500 km Terms of Use



National Oceanography Centre
NATURAL ENVIRONMENT RESEARCH COUNCIL

noc.ac.uk

NERC SCIENCE OF THE ENVIRONMENT



Birthdays are good for us. Statistics show that the people who have the most live the longest.

I hope you have exponentially
more fun than last year!