



SeaDataNet

History of the building and upgrade of an European oceanographic data infrastructure

Download the presentation:

<https://cloud.ifremer.fr/index.php/s/6BgEpDOBJGebGG5>

Michèle Fichaut, IFREMER, SISMER (NODC)

Open Research Data Week, Malta, 25 January 2022

sdn-userdesk@seadatanet.org – www.seadatanet.org

Overview

1. The context: Marine Data Management
2. The SeaDataNet infrastructure building
3. Results of SeaDataNet, SeaDataNet2 and SeaDataCloud projects
4. FAIR principles and data FAIRness

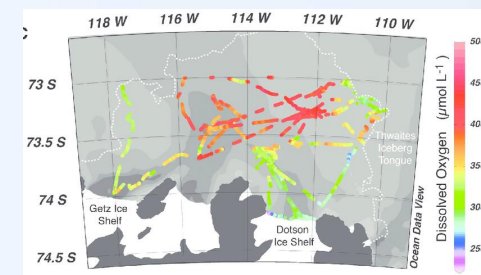
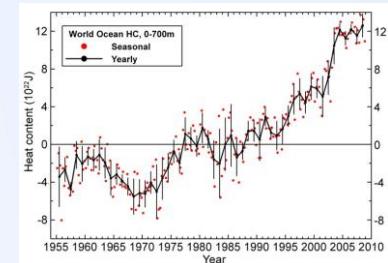
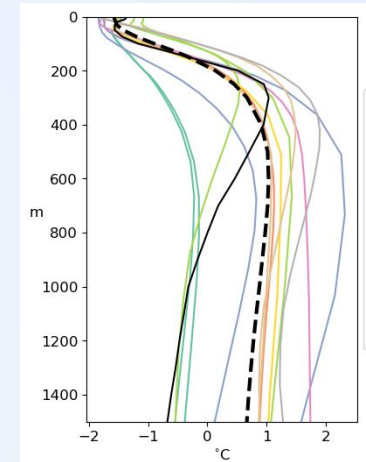
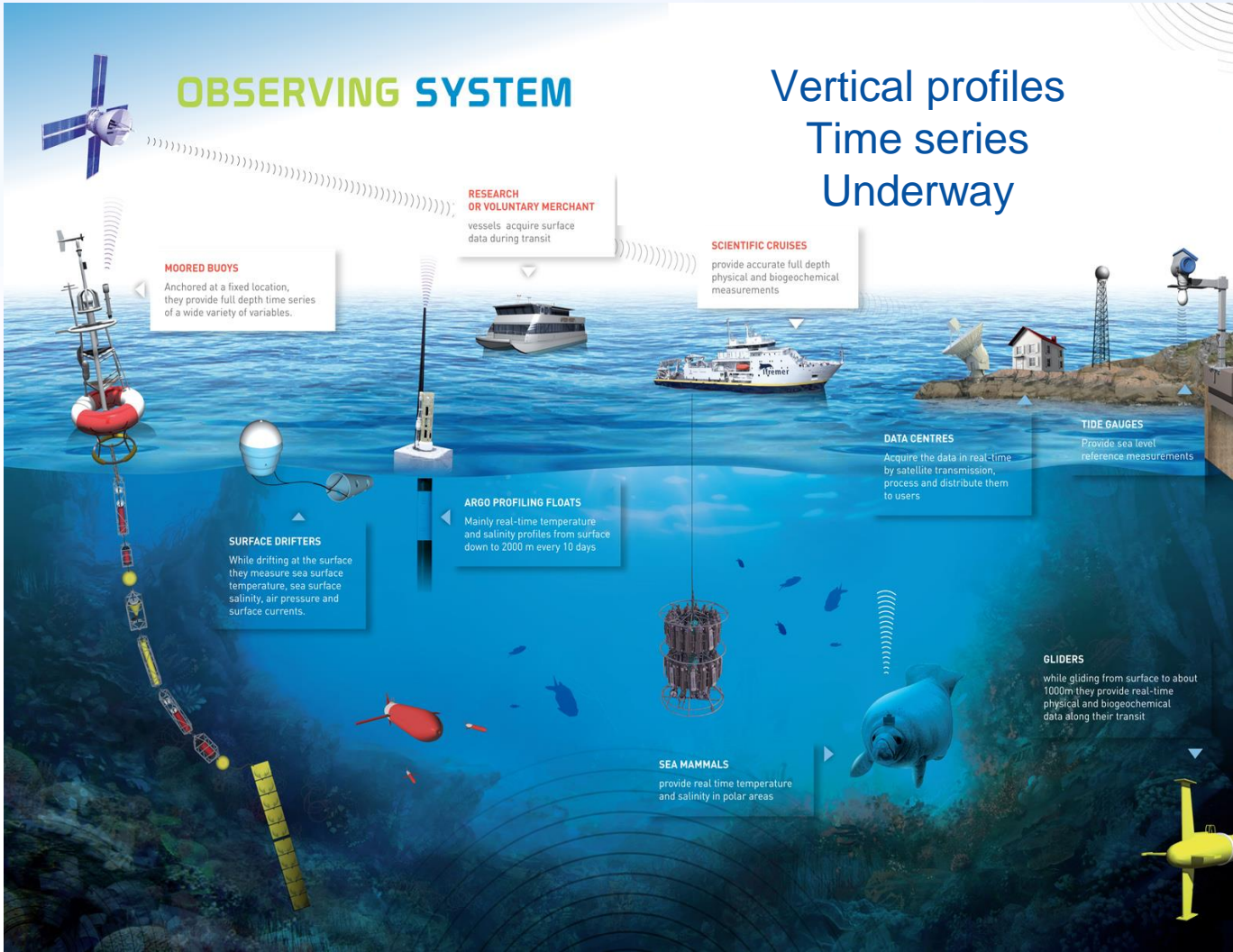
Part 1

- The context: marine data management
 - Data collection at sea
 - Data workflow: from collection to diffusion
 - Contribution of the data centres
 - Quality checks of the data
 - IODE and the NODC network
 - European Landscape

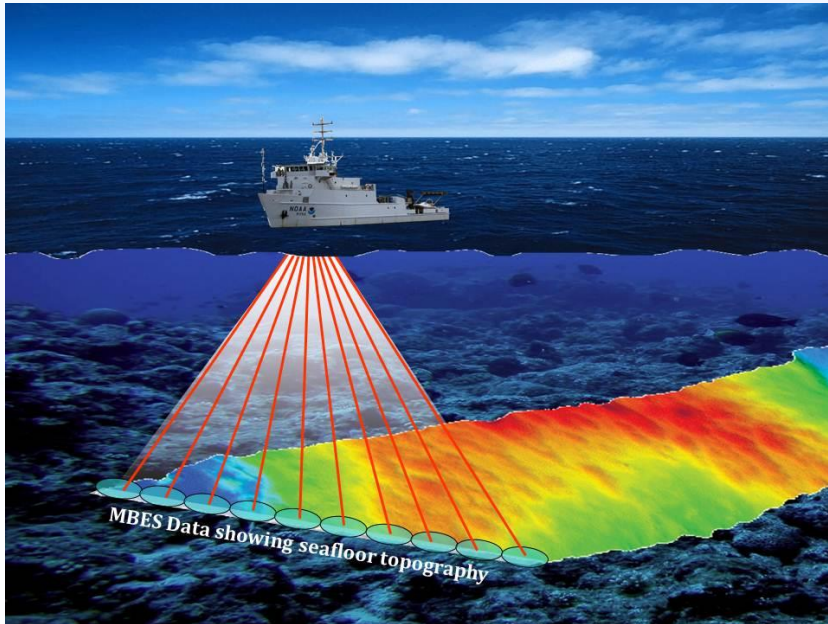
Marine Data management

- Data measured in Seas and Oceans
 - Physical
 - Temperature, salinity, current, waves...
 - Chemical
 - Nutrients(nitrate, phosphate, silicate), dissolved oxygen, PH, ...
 - Biological
 - Zooplankton and phytoplankton, benthos, Chlorophyll...
 - Geosciences
 - Bathymetry, magnetism, gravimetry...
 - Meteorology at the interface Ocean/Atmosphere

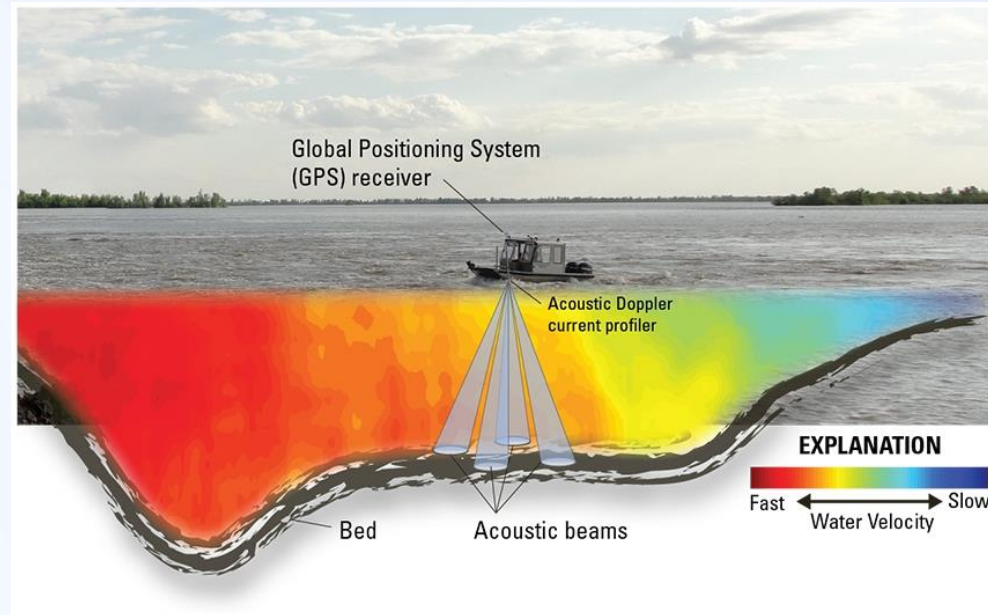
Different observing systems measuring data



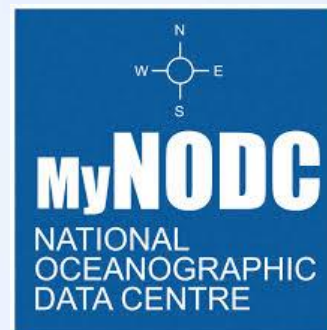
Underway measurements



Bathymetry



Currents measurements with ADCPs
(Acoustic Doppler Current Profilers)



In-situ data workflows

1. Transmitted to a data centre

- **In Near Real Time**(between 2 hours and one week after the data collection)
 - Automatic sensors like tide Gauge, Argo floats, sensors on sea mammals...
- **In Delayed mode**
 - Systematic measurements on ships: transmitted at the end of the research cruise
 - by the ship owner or the chief scientist (generally 2 months max after the end of a cruise)
 - Thermosalinometer, meteorological data, bathymetry, seismic, gravimetry, magnetism, ADCP
 - Targeted measurements by the scientific team on board, analysed and calibrated in the laboratories (several months to several years after the end of the end of the cruise)
 - Chemistry measurements that require equipment not available on board
 - Measurements requiring post-calibration

In-situ data workflows

2. Archived by the projects that funded their collection
3. Kept in the scientific laboratories
sometimes on non sustainable supports



Less and less true because of

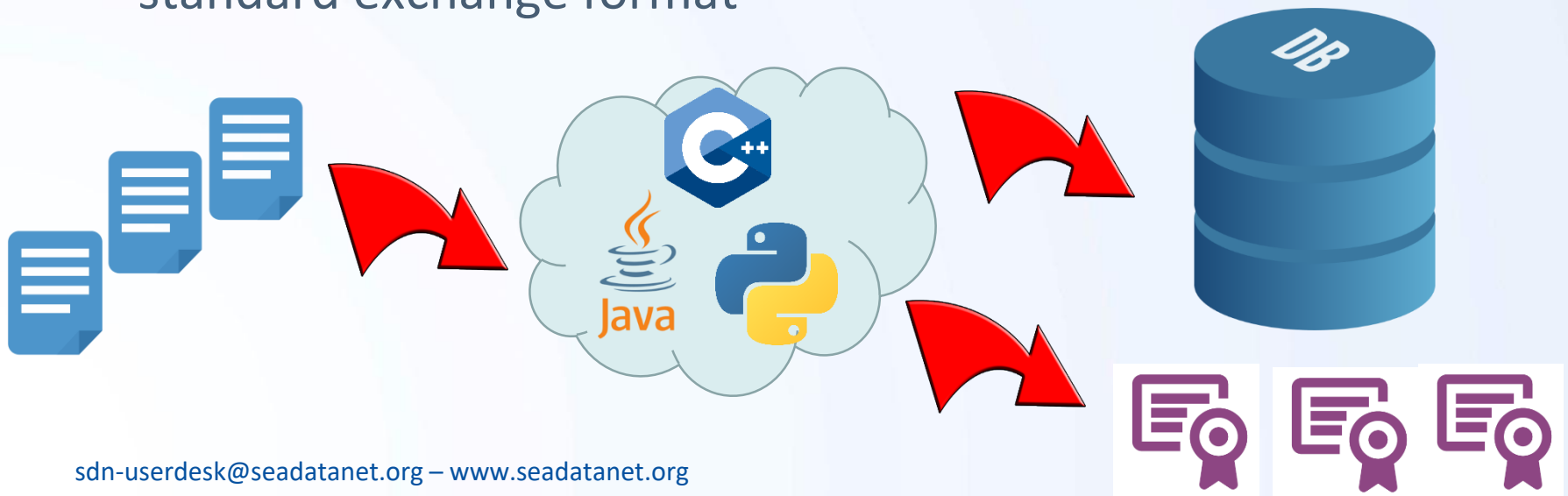


Format of the data sent to NODCs

- For Near Real Time data
 - Manufacturer's format of the sensors
- For Delayed mode data
 - Manufacturer's format of the sensors
 - Format of the on-board data acquisition centre
 - Any format when the data have been transmitted through the scientific laboratories
 - ASCII (text, CSV,...), Binary (netCDF), Excel sheets...

Processing of the data file received in NODCs

- When formats are defined, known and stable
 - One reading script per format
- When formats are arbitrary
 - one generic program able to read several formats
- To ingest the data in a database or to convert them in a standard exchange format

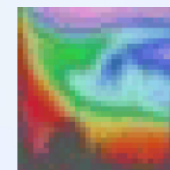


Archiving data in NODCs

- Conversion to standards
- Quality check of the data
- Long-term archiving
- Data distribution by mean of different supports and protocols
 - Web site (Discovery through metadata, Access to the data), FTP, OpenDap, ...
- Following a data policy: public data, confidential data, data with a moratorium

Standard formats for oceanographic data

- OceanData View (ODV)  ASCII CSV format



Ocean Data View

- netCDF  binary format



- MedAtlas  ASCII



```
001 OVIDE 06 ..... Maria S. MERIAN .. NB .....  
*No station,nom campagne,nom navire,type sonde  
*(i3,2(lx,a16),lx,a6)  
24052006 1428 N .. 38 26.00 W .. 10 42.00 .. 4 4582 4915 ..... 1 4582  
*Date,heure,lat.,lon.,nb parametres,nb mesures,fond,pmin,pmax  
*(a8,lx,i2,i2,lx,all,lx,all,lx,i2,4(lx,i4))  
PRESSION ..... dbar .....  
TEMPERATURE ..... deg.cels. ....  
SALINITE ..... p.s.u. ....  
OXYGENE DISS. .... um/kg .....  
... 1.0 16.512 36.198 233.5  
... 2.0 16.512 36.198 233.5  
... 3.0 16.512 36.198 233.5  
... 4.0 16.512 36.198 233.5  
... 5.0 16.512 36.198 233.5  
... 6.0 16.512 36.198 233.5  
... 7.0 16.512 36.198 233.5  
... 8.0 16.512 36.198 233.5  
... 9.0 16.512 36.198 233.5  
... 10.0 16.512 36.198 233.5  
... 11.0 16.512 36.198 233.5  
... 12.0 16.512 36.198 233.5  
... 13.0 16.512 36.198 233.5  
... 14.0 16.512 36.198 233.5  
... 15.0 16.512 36.198 233.5  
... 16.0 16.512 36.198 233.5  
... 17.0 16.512 36.198 233.5  
... 18.0 16.512 36.198 233.5  
... 19.0 16.512 36.198 233.5  
... 20.0 16.496 36.197 236.3  
... 21.0 16.492 36.197 236.5  
... 22.0 16.464 36.197 236.3  
... 23.0 16.417 36.207 236.9  
... 24.0 16.380 36.219 235.8  
... 25.0 16.351 36.225 235.0  
... 26.0 16.333 36.228 236.8  
... 27.0 16.353 36.223 237.6  
... 28.0 16.346 36.225 237.8  
... 29.0 16.336 36.226 238.5  
... 30.0 16.328 36.226 238.8  
... 31.0 16.300 36.226 238.6  
... 32.0 16.259 36.224 237.6  
... 33.0 16.234 36.222 238.8  
... 34.0 16.191 36.217 239.9  
... 35.0 16.166 36.211 240.7
```

Open Research Data Week, Malta, 25 January 2022

Example of received file

- Text file with minimum metadata
- And 4 measured parameters

```

*FI35200653001 OVIDE.3
21/05/2006-28/06/2006 North Atlantic Ocean
35_Ifremer_CNRS_Universit  de Bretagne Occidentale/IUEM_Instituto Investi
LHERMINIER_Pascale Project=OVIDE
Regional Archiving= FI Availability=PC
Data_Type=H10.n=120.QC=N
Data_Type=H21.n=120.QC=N
COMMENT
SeaDataNet_MEDATLAS_PHYSICOCHEMICAL_profile--Generated_by_Nemo.1.7.0--2020-03-18_16:25:37+0100
*FI3520065300100001.Data_Type=H10
*DATE=24052006.TIME=1428.LAT=N38.26.00.LON=W010.42.00.DEPTH=4915.QC=1111
*NB_PARAMETERS=04.RECORD.LINES=04582
*PRES_SEA.PRESSURE_sea_surface=0.(decibar=10000.pascals).def.=999.99
*TEMP_SEA.TEMPERATURE.(Celsius.degree).def.=99.999
*PSAL.PRACTICAL.SALINITY.(P.S.U.).def.=99.999
*DOX2.DISSOLVED.OXYGEN.(micromole/kg).def.=999.999
*GLOBAL_PROFILE_QUALITY_FLAG=1.GLOBAL.PARAMETERS_QC_FLAGS=1111
*DC.HISTORY=
*DM.HISTORY=
*COMMENT
*SDN_parameter_mapping
<subject>SDN:LOCAL:PRES</subject><object>SDN:P01: PRESPR01</object><units>SDN:P06::UPDB</units>
<subject>SDN:LOCAL:TEMP</subject><object>SDN:P01: TEMPPR01</object><units>SDN:P06::UPAA</units>
<subject>SDN:LOCAL:PSAL</subject><object>SDN:P01: PSLTZ201</object><units>SDN:P06::UUUU</units>
<subject>SDN:LOCAL:DOX2</subject><object>SDN:P01: DOXMZXX</object><units>SDN:P06::KGUM</units>
*EDMO_CODE=486
*CDI_ID=FI35200653001_00001_H10
<sdn_reference xlink:href="https://seadatanet.org/urn:sdn:0:0:2020:03:18:16:25:37+0100:00001_H10" xlink:role="isObservedBy" xlink:type="SDN:L23::CSR"/>
<sdn_reference xlink:href="https://www.seadatanet.org/urn:sdn:0:0:2020:03:18:16:25:37+0100:00001_H10" xlink:role="isObservedBy" xlink:type="SDN:L23::NVS2CON"/>
<sdn_reference xlink:href="https://doi.org/10.26434/chemrxiv-2020-03-18-16:25:37+0100:00001_H10" xlink:role="isDescribedBy" xlink:type="SDN:L23::CDI"/>
*SURFACE_SAMPLES=
*PRES...TEMP...PSAL...DOX2...

```

Same file converted at MEDATLAS format

```

...1.0.16.512.36.198.233.500 1111
...2.0.16.512.36.198.233.500 1111
...3.0.16.512.36.198.233.500 1111
...4.0.16.512.36.198.233.500 1111
...5.0.16.512.36.198.233.500 1111
...6.0.16.512.36.198.233.500 1111
...7.0.16.512.36.198.233.500 1111
...8.0.16.512.36.198.233.500 1111
...9.0.16.512.36.198.233.500 1111
...10.0.16.512.36.198.233.500 1111
...11.0.16.512.36.198.233.500 1111
...12.0.16.512.36.198.233.500 1111
...13.0.16.512.36.198.233.500 1111
...14.0.16.512.36.198.233.500 1111
...15.0.16.512.36.198.233.500 1111
...16.0.16.512.36.198.233.500 1111
...17.0.16.512.36.198.233.500 1111
...18.0.16.512.36.198.233.500 1111
...19.0.16.512.36.198.233.500 1111
...20.0.16.496.36.197.236.300 1111
...21.0.16.492.36.197.236.500 1111
...22.0.16.464.36.197.236.300 1111
...23.0.16.417.36.207.236.900 1111
...24.0.16.380.36.219.235.800 1111
...25.0.16.351.36.225.235.000 1111
...26.0.16.333.36.228.236.800 1111
...27.0.16.353.36.223.237.600 1111

```

- Additional metadata
 - Cruise name, ship, chief scientist, laboratories involved, project, instrument used...
 - Links to the Cruise report, to the full metadata file
- Standard name for the parameters
- Quality flags on all measurements and metadata

```

<sdn_reference xlink:href="http://vocab.nere.ac.uk/collection/U1/current/U6M2" xlink:role="isObservedBy" xlink:type="SDN:L23::CSR"/>
<sdn_reference xlink:href="http://seadatanet.maris2.nl/v_cdi_v3/print_xml.asp?edmo=486&identifier=FI35200653001_00001_H10" xlink:role="isDescribedBy"
xlink:type="SDN:L23::CDI" sdn:scope="486:FI35200653001_00001_H10"/>
//SDN_parameter_mapping
</subject>SDN:LOCAL:DEPHPR01</subject><object>SDN:P01::DEPHPR01</object><units>SDN:P06::ULAA</units>
</subject>SDN:LOCAL:Pressure</subject><object>SDN:P01::PRESPR01</object><units>SDN:P06::UPDB</units>
</subject>SDN:LOCAL:Temperature</subject><object>SDN:P01::TEMPPR01</object><units>SDN:P06::UPAA</units>
</subject>SDN:LOCAL:Salinity</subject><object>SDN:P01::PSLTZZ01</object><units>SDN:P06::UUUU</units>
</subject>SDN:LOCAL:Oxygen</subject><object>SDN:P01::DOXMZZXX</object><units>SDN:P06::KGUM</units>

```

Cruise Station Type	yyyy-mm-ddThh:mm:ss.sss	Longitude [degrees_east]	Latitude [degrees_north]	LOCAL_CDI_ID	EDMO_code	Bot. Depth [m]	DEPHPR01 [m]
QV:SEADATANET Pressure [Decibars]	QV:SEADATANET	Temperature [Degrees Celsius]	QV:SEADATANET	Salinity [Dimensionless]	QV:SEADATANET	Oxygen [Micromoles per kilogram]	QV:SEADATANET
001 C	2006-05-24T14:28:00.000	-10.700000	+38.333	FI35200653001_00001_H10	486	15.0	1.0 1 1.0 1 16.512 1 36.198 1 233.5 1
2.0	1	2.0	1	16.512	1	36.198	1 233.5 1
3.0	1	3.0	1	16.512	1	36.198	1 233.5 1
4.0	1	4.0	1	16.512	1	36.198	1 233.5 1
5.0	1	5.0	1	16.512	1	36.198	1 233.5 1
6.0	1	6.0	1	16.512	1	36.198	1 233.5 1
6.9	1	7.0	1	16.512	1	36.198	1 233.5 1
7.9	1	8.0	1	16.512	1	36.198	1 233.5 1
8.9	1	9.0	1	16.512	1	36.198	1 233.5 1
9.9	1	10.0	1	16.512	1	36.198	1 233.5 1
10.9	1	11.0	1	16.512	1	36.198	1 233.5 1
11.9	1	12.0	1	16.512	1	36.198	1 233.5 1
12.9	1	13.0	1	16.512	1	36.198	1 233.5 1
13.9	1	14.0	1	16.512	1	36.198	1 233.5 1
14.9	1	15.0	1	16.512	1	36.198	1 233.5 1
15.9	1	16.0	1	16.512	1	36.198	1 233.5 1
16.9	1	17.0	1	16.512	1	36.198	1 233.5 1
17.9	1	18.0	1	16.512	1	36.198	1 233.5 1
18.9	1	19.0	1	16.512	1	36.198	1 233.5 1
19.8	1	20.0	1	16.496	1	36.197	1 236.3 1
20.8	1	21.0	1	16.492	1	36.197	1 236.5 1
21.8	1	22.0	1	16.464	1	36.197	1 236.3 1
22.8	1	23.0	1	16.417	1	36.207	1 236.9 1
23.8	1	24.0	1	16.380	1	36.219	1 235.8 1
24.8	1	25.0	1	16.351	1	36.225	1 235.0 1
25.8	1	26.0	1	16.333	1	36.228	1 236.8 1
26.8	1	27.0	1	16.353	1	36.223	1 237.6 1

Same file converted at ODV format

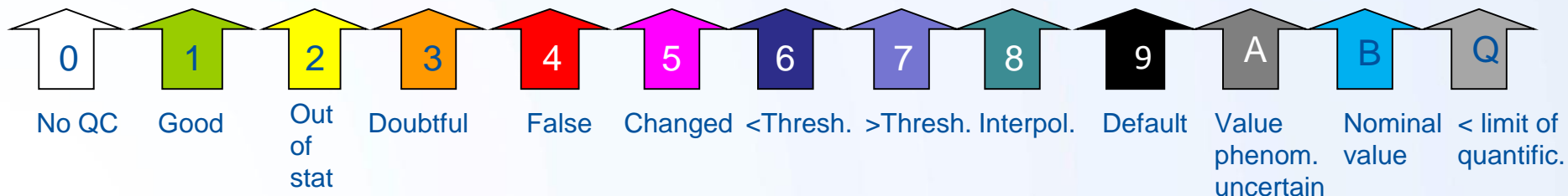
- Additional metadata
 - Cruise name, link to the Cruise report, link to the full metadata file, ship, instrument used
- Standard name for the parameters
- Quality flags on all measurements

Quality checks (QC)

- QC of all data received at the Data centre
- 3 steps
 - Automatic checks of the format
 - Automatic and visual checks of the metadata
 - Automatic and visual checks of the measured data

– Results

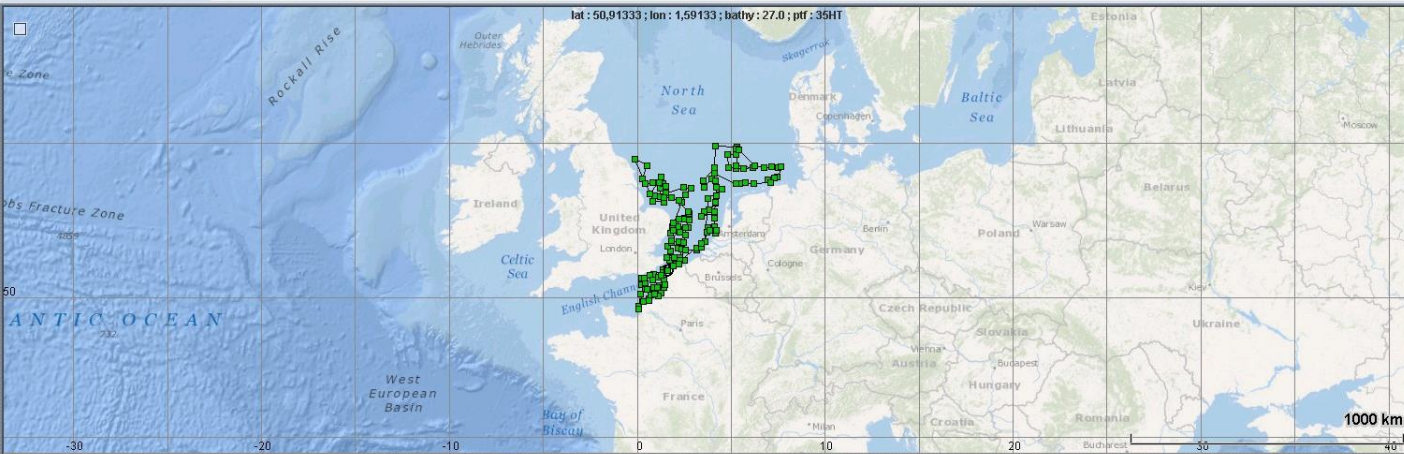
- A quality flag on all numerical values



Quality Checks – Checks of the metadata

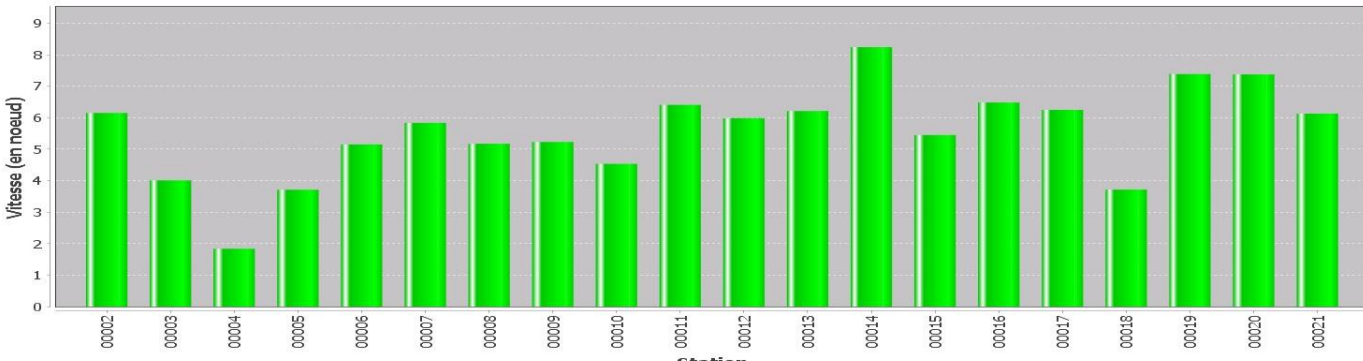
Description Campagne

Référence campagne	F135202112370
Nom campagne	IBTS 2021
Nom navire	Thalassa
Code navire	35HT
Date début	15/01/2021
Date fin	08/02/2021
Région	English Channel
Pays	35
Organisme	IFREMER Manche Mer du Nord, IFREMER Bretagne, IFREMER Méditerranée
Responsable	AUBER Arnaud
Projet	
Centre d'archivage	FI
Type de données 0	Data Type=B02 n= 148 QC=Y
Type de données 1	Data Type=H10 n= 148 QC=Y
Type de données 2	Data Type=H16 n= 148 QC=Y
Type de données 3	Data Type=H17 n= 148 QC=Y
Type de données 4	Data Type=H21 n= 148 QC=Y
Type de données 5	Data Type=H28 n= 148 QC=Y
Commentaire	SeaDataNet/MEDATLAS PHYSICO-CHEMICAL PROFILE - Generated by SCOP 1.52 - 2022-01-21 09:23:39:0100 DOXY = (valeurs fichiers en mg/l * 0.6998) * 44.66 Les profils de salinité constants sont dus au nombre de 3 décimales sur le paramètre Sonde pH défaillante pendant la campagne (tous les profils flaggués à 4).



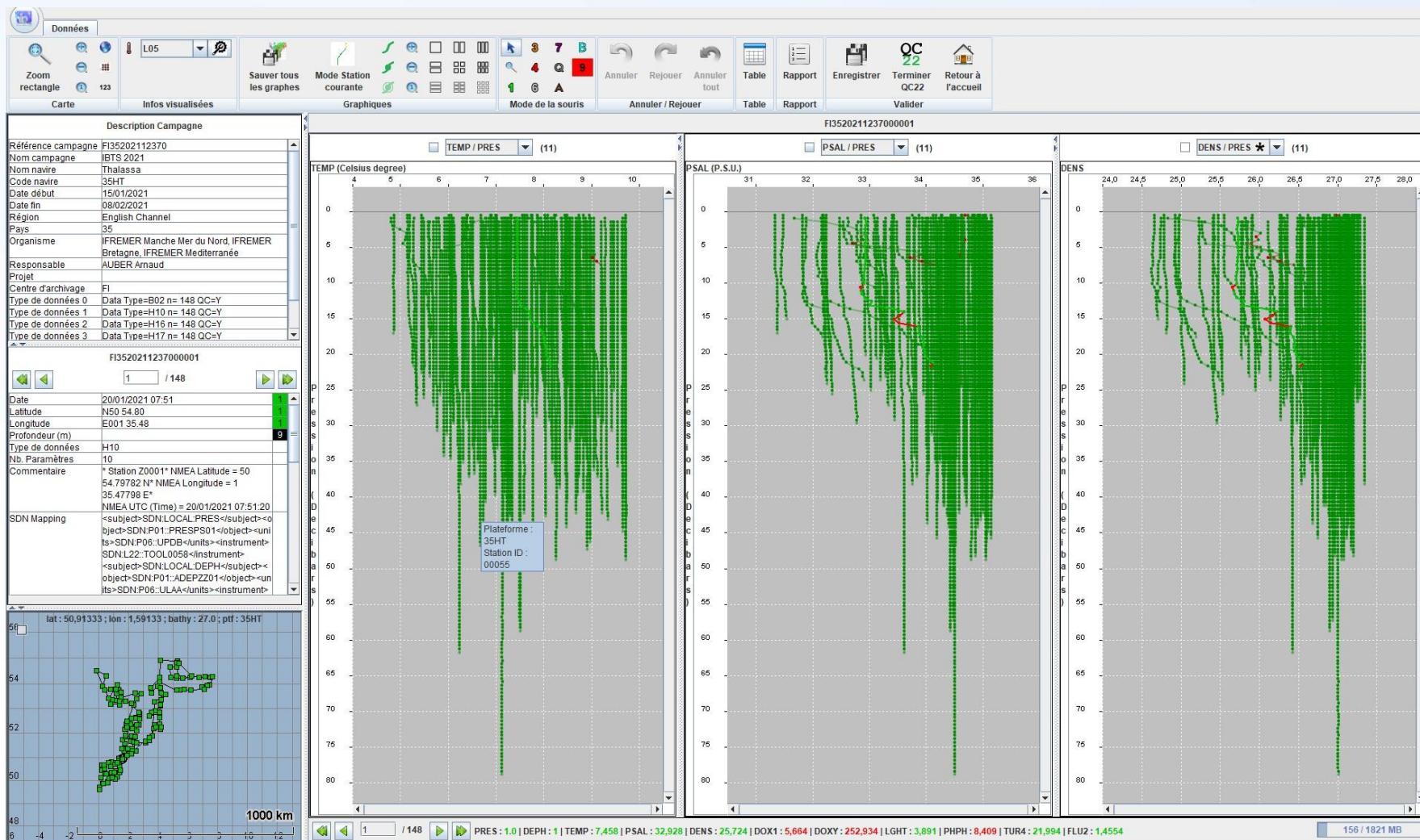
lat: 50,91333 ; lon : 1,59133 ; bathy : 27.0 ; pif : 35HT

Extrapolation des vitesses de la plateforme



Station	Vitesse (en noeud)
0002	6.2
0003	4.0
0004	1.8
0005	3.6
0006	5.2
0007	5.8
0008	5.2
0009	5.2
0010	4.5
0011	6.5
0012	6.0
0013	6.2
0014	8.2
0015	5.5
0016	6.5
0017	6.2
0018	3.5
0019	7.5
0020	7.5
0021	6.0

Quality Checks – Checks of the data



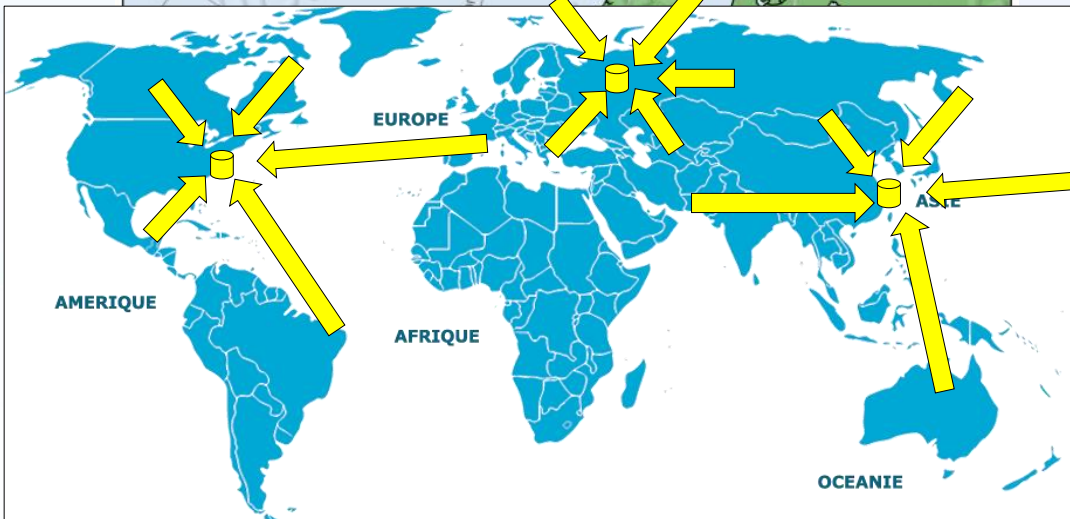
IOC/IODE NODC network

- The programme "International Oceanographic Data and Information Exchange" (**IODE**) of the "**Intergovernmental Oceanographic Commission**" (**IOC**) of **UNESCO** was established in 1961. Its purpose is to enhance marine research, exploitation and development, by **facilitating the exchange of oceanographic data and information between participating Member States**, and by meeting the needs of users for data and information products.

IOC/IODE objectives

- To **facilitate and promote the discovery, exchange of, and access to, marine data** and information
- To **encourage the long term archival**, preservation, documentation, management and services of all marine data, data products, and information
- To develop or use existing **best practices** for the discovery, management, exchange of, and access to marine data and information
- To assist Member States **to acquire the necessary capacity** to manage marine research and observation data and information and become partners in the IODE network (Capacity building)
- To **support** international scientific and operational marine programmes

NODC network in Europe



- Oceanographic data are managed by the NODCs (part of the IODE network)
- ICES in Copenhagen is a European data centre collecting data from EU NODCs
- 3 World Data centres collecting data from NODCs worldwide

NODCs network

- Through IOC/IODE programme, NODCs meet every 2 years during one week
- During these meetings the IODE projects are discussed, recommendations are given on best practices for data management, quality assurance, standards ...
- But no standards are imposed
 - No common vocabularies
 - No common formats



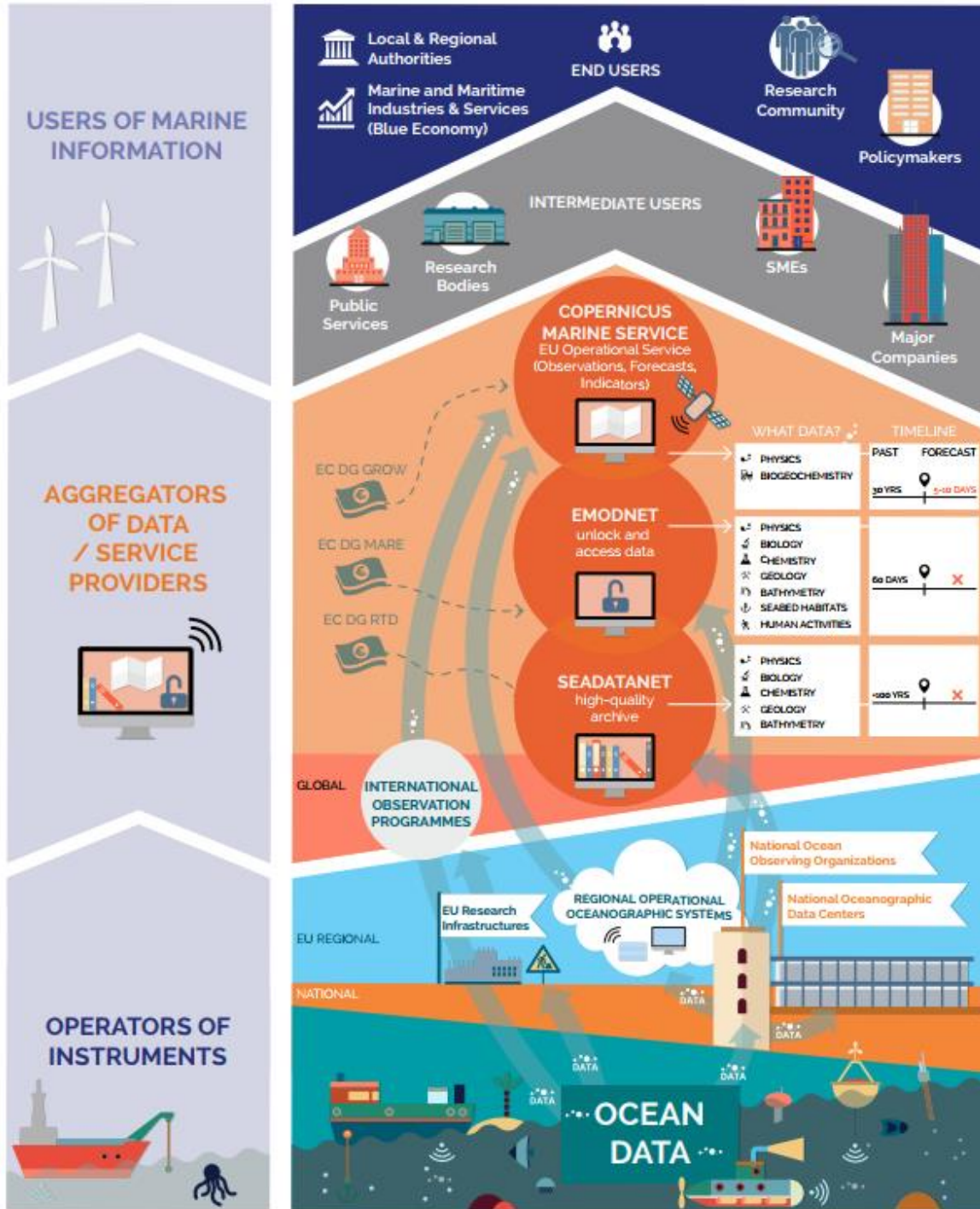
This is done in the frame of European projects





European landscape in terms of Marine Data management: 3 main components

sdn-userdesk@seadatanet.org – www.s



End of part 1



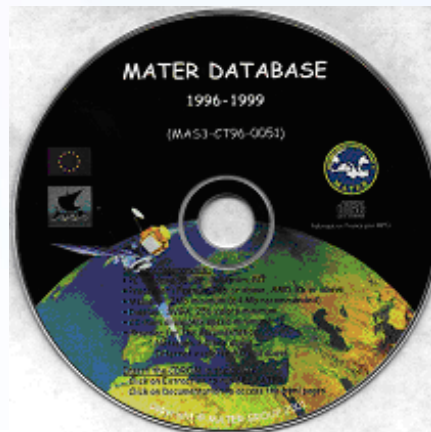
Part 2

The SeaDataNet infrastructure

- History
- Objective and principle
- Roadmap
- Method applied to reach interoperability of data and metadata

Need for standardisation

- Over the last 30 years, development of European projects
 - Bringing together NODCs from different countries in joint oceanographic data projects
 - Need for standardisation, to disseminate homogeneous and coherent data sets: CD-ROMs at the end of the 1990s, beginning of the 2000s



The SeaDataNet infrastructure

- Stop projects requiring the centralisation of homogeneous data
- Data remain in NODCs but are accessible from a single location: the SeaDataNet portal = virtual data centre, based on the European research infrastructure (RI) SeaDataNet





90s

EDMED

Euronodim

MEDATLAS

2002-2005

Sea-Search

2006-2011

SeaDataNet

2011-2015

SeaDataNet II

2016-2021

SeaDataCloud

EU – MAST

EU – MAST II

EU-FP5

EU-FP6

EU-FP7

H2020

Partners : 35 countries, bordering European seas
Coordinated by IFREMER

SeaDataNet - Objectives

- Federation of the the oceanographic data centres of 35 countries bordering the European seas
- Creation of a single virtual data centre, allowing a user searching for physics/chemistry data to connect to all 35 countries from a single user interface
- Distribute complete datasets in specific sea basins to privileged users (modellers)
- Creation and dissemination of products (climatologies and aggregated datasets) made from the data put into the infrastructure

SeaDataNet Infrastructure - Principle

- semi-distributed system that aggregates **NODCs** and enhances the existing NODC network.
- The technical developments implemented allow the NODCs connected to the SeaDataNet system to be seen as a **single virtual data centre** able to deliver **quality controlled data**, metadata and products through a **single web portal**

SeaDataNet – Road map

- SeaDataNet (2006-2011) : 10 M€
 - System implementation in 2 steps
 - Connection of the 10 most technologically advanced data centres to implement and test the system
 - Connecting the other 29 data centres in a progressive way with assistance from the 10 already connected
 - SeaDataNet 2 (2011-2015) : 6 M€
 - Make the system more reliable (monitoring, automated (machine-to-machine data exchange) and sustainable (infrastructure funded outside the European project)
 - Connect more data centres
 - Add more data, and more types of data (biology)
- ❑ Bigger, Better, Faster



SeaDataNet – Road map

- SeaDataCloud (2016-2021) : 10 M€
 - Improving access to data
 - Take into account the evolution of technologies => Cloud, HPC
 - More data processing capacity
 - Improved response times
 - Give a central role to the users
 - Provide the user with tools in a virtual environment (VRE, Virtual Research Environment) in which he/she will be able to work on his/her own data + data from SeaDataNet
 - To store his/her working environment: MySeaDataCloud




SeaDataNet : Method

- Development of standards
 - Common vocabulary for metadata
 - Common protocol for data and metadata control (Have comparable data)
 - Common file formats
- Definition of common catalogues
- Definition of rules for making data available
- Use of common software developed in the framework of SeaDataNet and made available to all partners (and more)

Speak the same language: Common vocabularies Interoperability



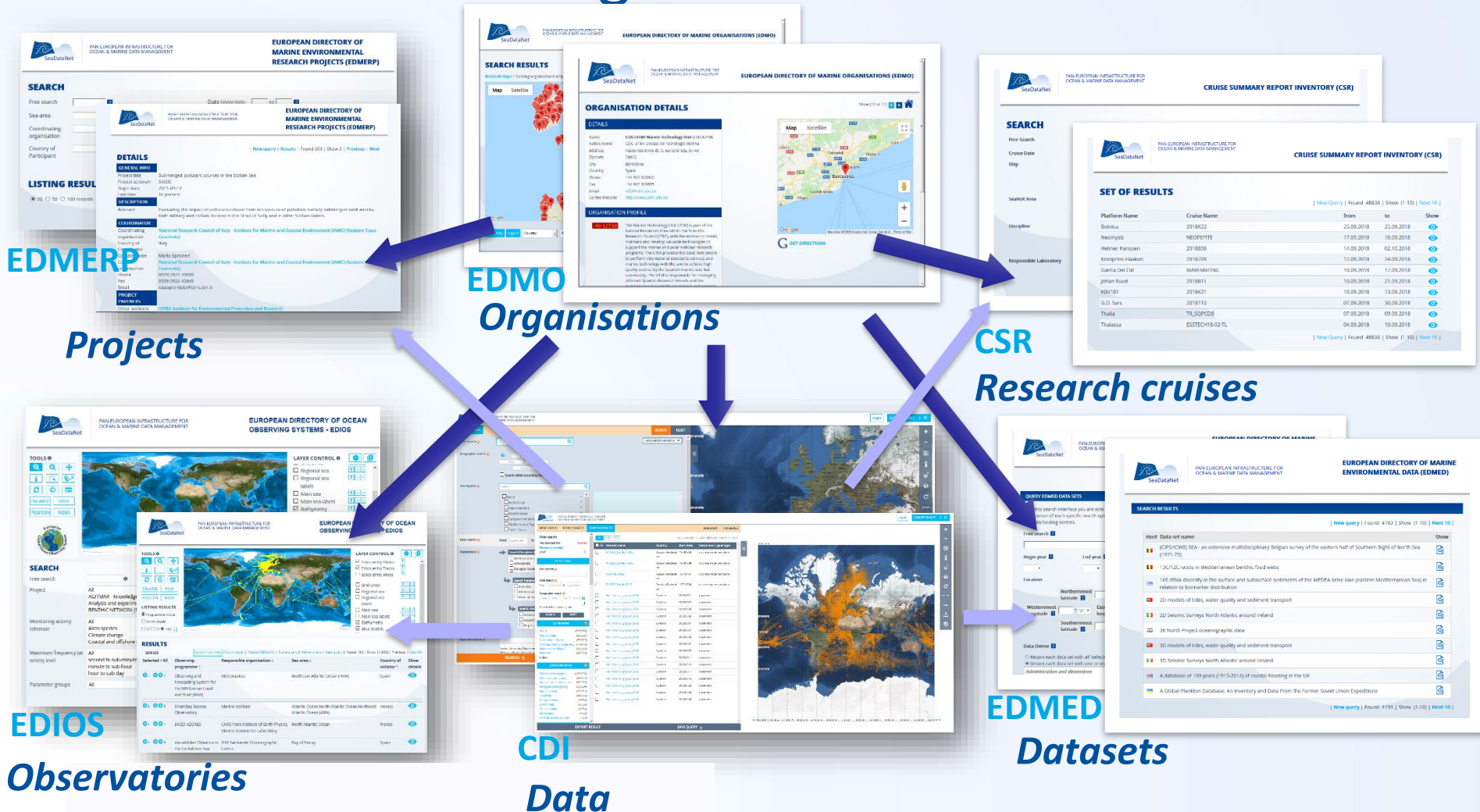
- Vocabulary lists maintained by the British Oceanographic Data Centre (NercVocabularyServer/BODC)
- 90,000 terms in over 110 vocabulary lists
 - Geographical area, ships, ports, scientific disciplines, data types, parameters, measurement units, instruments, positioning systems.....
- On-line through
 - Web site : <https://www.seadatanet.org>  Look-up vocabularies
 - Web services: http://www.bodc.ac.uk/products/web_services/vocab/

To speak the same language : Common metadata descriptions ☐ interoperability

- Following International/European standards
 - Metadata descriptions based upon ISO-19115 and ISO-19139 for the compliancy to INSPIRE
 - available on-line on SeaDataNet website (<https://www.seadatanet.org/Standards/Metadata-formats>) and on the Ocean best practices repository (<https://www.oceanbestpractices.org/>)



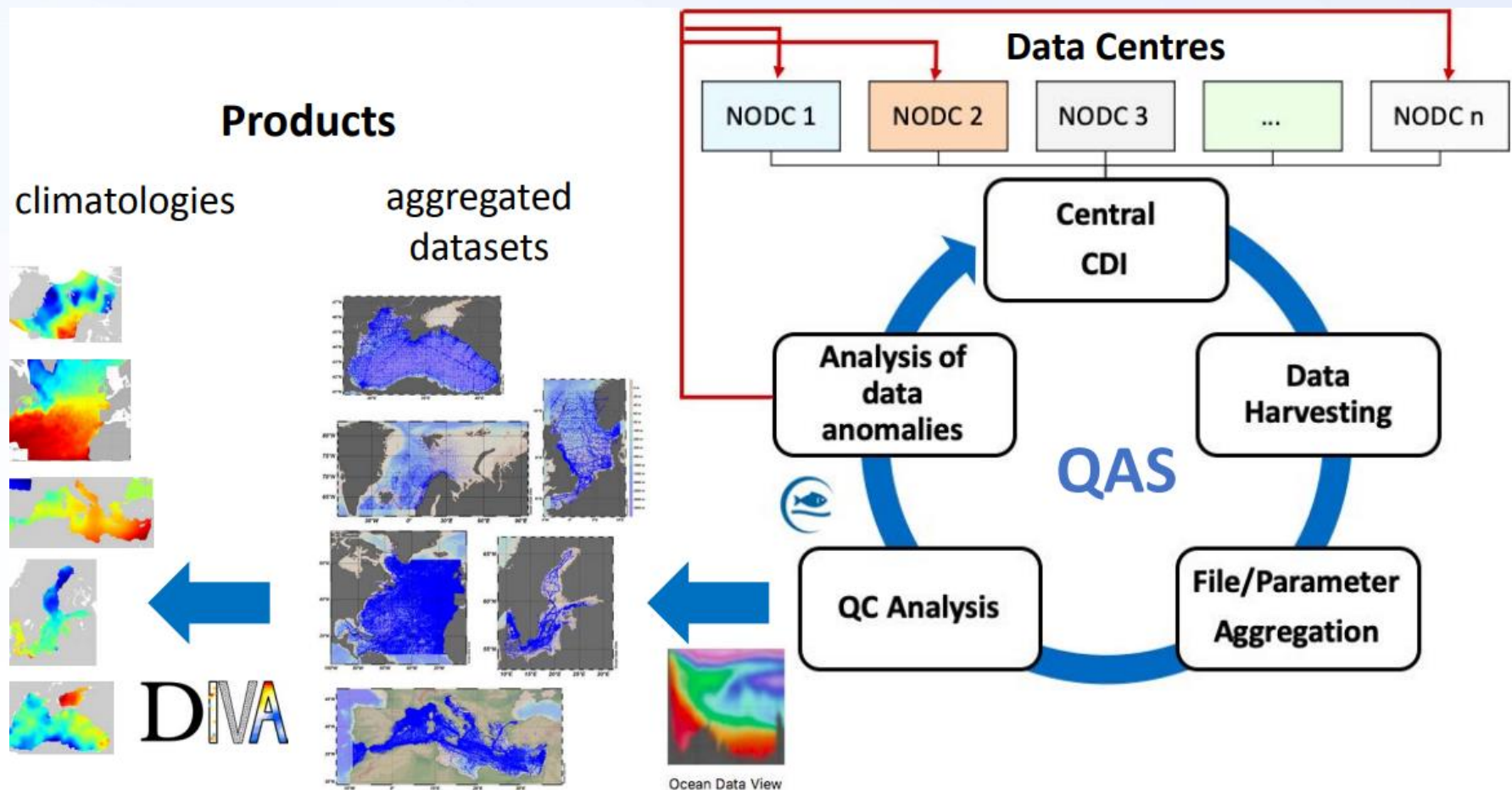
6 common catalogues of metadata



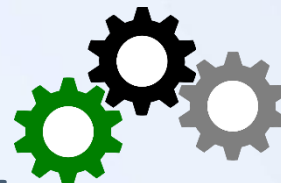
Have comparable data interoperability

- Same file formats relying on the common vocabularies
 - ODV - Ocean Data view (ASCII)
 - MEDATLAS (ASCII)
 - NETCDF – CF (Binaire)
- Quality check protocol based on international (IOC/IODE) recommendations applied by all data centres
 - With automatic and manual checks
- Same quality flags on all measurements (part of the common vocabularies)
- Quality Assurance strategy, implementing a QC -Loop

QC-Loop with feedback to data centres



Have common tools for data and metadata preparation



- **Tools are distributed to all SeaDataNet partners**



- **MIKADO** : To generate the metadata descriptions of the SDN catalogues



- **NEMO** To convert files to SeaDataNet formats

- **OCTOPUS** : To convert from One SDN format to another



- **Ocean Data View (ODV)** : To visualise and QC the data

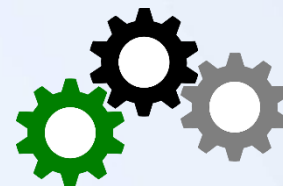


- **DIVA** : For adata analysis and product generations (climatology)



- **Download Manager (DM) and Replication Manager (RM)** : to send datafiles from one data centre to the users or to the cloud

Have common tools for data and metadata access

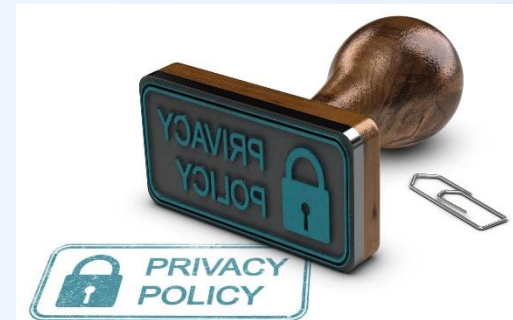


- **Installed on central servers:**
 - **Central catalogues and the corresponding CMS** = Content Management System for updating (for projects and cruises and datasets)
 - **Web interfaces** for querying the various catalogues and searching for data
 - **Request Status Manager (RSM):** so that a user can track the progress of his or her data request
 - **Marine-ID** for identification in case of data downloading

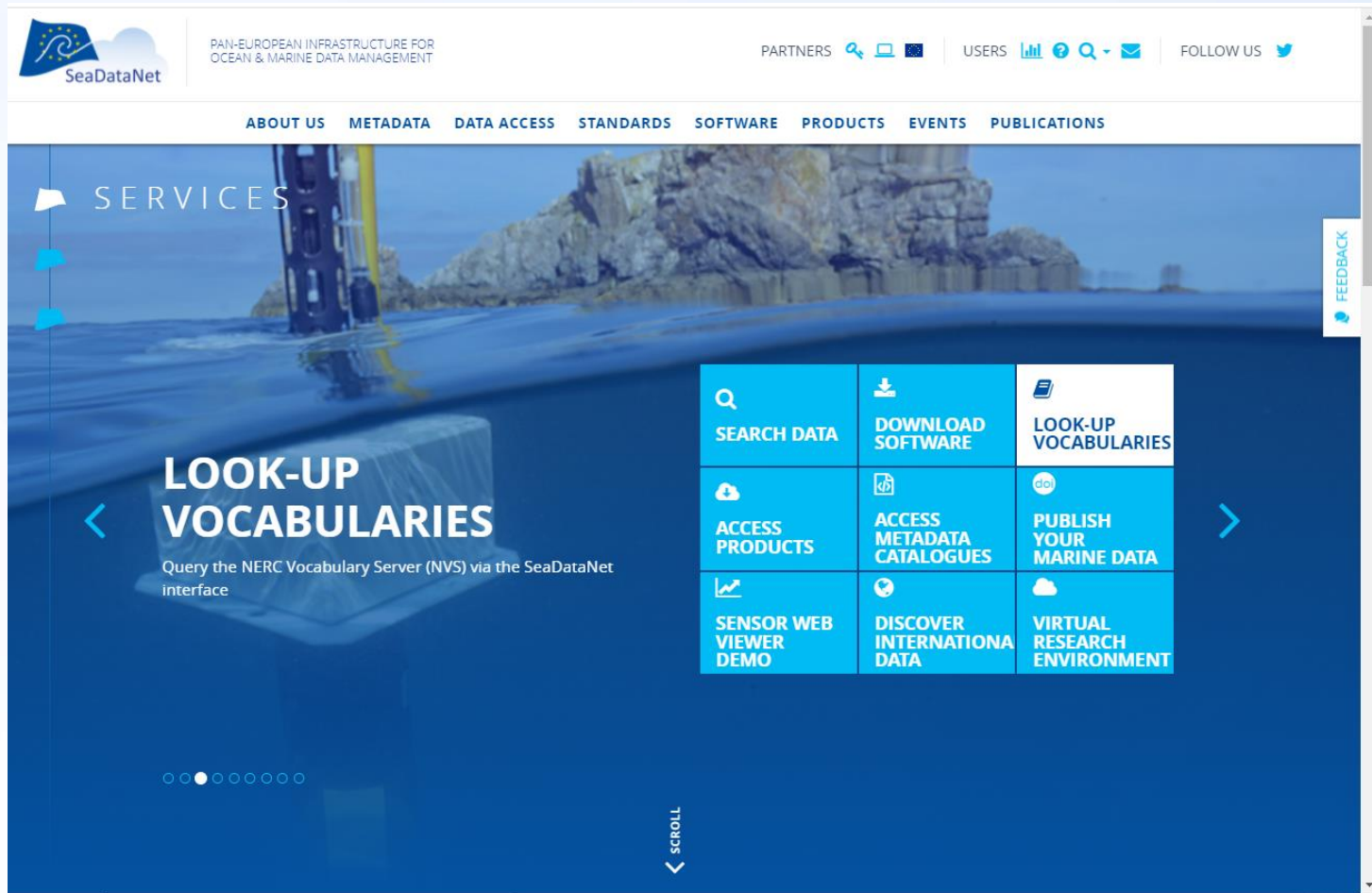


Define the data policy

- Most of the SeaDataNet data are publicly available (CC-by licence) for download (91%)
- 9% of restricted data are managed too : restriction like moratorium or other access restrictions
 - The metadata of restricted data are available but
 - Their distribution is under the responsibility of each data centre that have included them in the SeaDataNet infrastructure : case per case negotiation with the data user



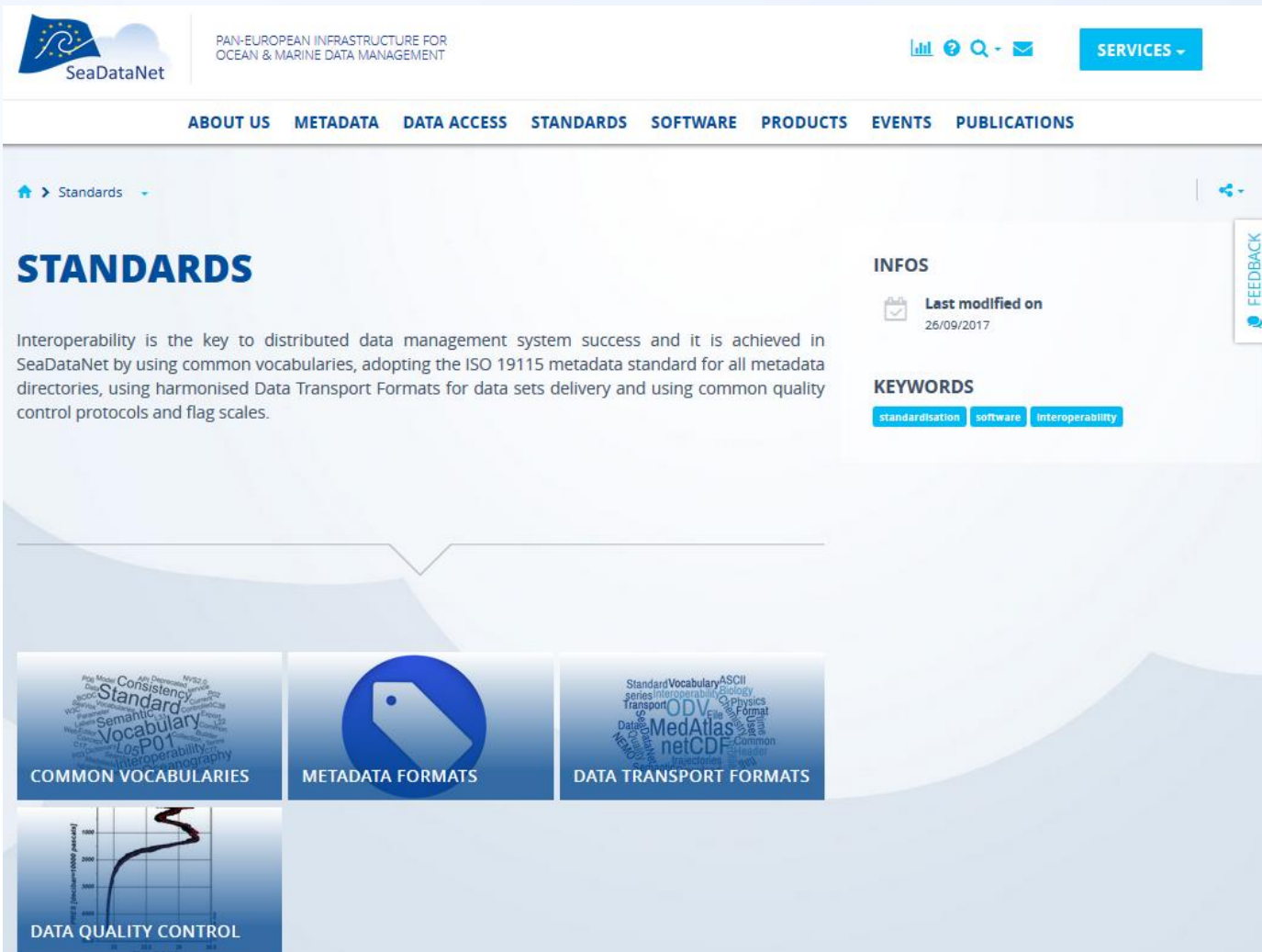
SeaDataNet portal- <https://www.seadatanet.org/>




The screenshot shows the SeaDataNet portal homepage. At the top left is the SeaDataNet logo and the text "PAN-EUROPEAN INFRASTRUCTURE FOR OCEAN & MARINE DATA MANAGEMENT". To the right are links for "PARTNERS", "USERS", and "FOLLOW US". Below this is a navigation menu with "ABOUT US", "METADATA", "DATA ACCESS", "STANDARDS", "SOFTWARE", "PRODUCTS", "EVENTS", and "PUBLICATIONS". The main content area features a large blue banner with a background image of a boat's deck. On the left, a "SERVICES" menu is visible. The central focus is a "LOOK-UP VOCABULARIES" section with the text "Query the NERC Vocabulary Server (NVS) via the SeaDataNet interface". To the right of this section is a 3x3 grid of service tiles: "SEARCH DATA", "DOWNLOAD SOFTWARE", "LOOK-UP VOCABULARIES", "ACCESS PRODUCTS", "ACCESS METADATA CATALOGUES", "PUBLISH YOUR MARINE DATA", "SENSOR WEB VIEWER DEMO", "DISCOVER INTERNATIONAL DATA", and "VIRTUAL RESEARCH ENVIRONMENT". A "FEEDBACK" button is on the far right, and a "SCROLL" arrow is at the bottom center.

- Metadata
- Data
- Standards
- Tools
- Products

Standards



The screenshot shows the SeaDataNet website's 'Standards' page. At the top left is the SeaDataNet logo and the text 'PAN-EUROPEAN INFRASTRUCTURE FOR OCEAN & MARINE DATA MANAGEMENT'. To the right are navigation icons for a bar chart, help, search, and email, followed by a 'SERVICES' button. A horizontal menu contains links for 'ABOUT US', 'METADATA', 'DATA ACCESS', 'STANDARDS', 'SOFTWARE', 'PRODUCTS', 'EVENTS', and 'PUBLICATIONS'. Below the menu, a breadcrumb trail shows 'Home > Standards'. The main heading is 'STANDARDS'. The introductory text states: 'Interoperability is the key to distributed data management system success and it is achieved in SeaDataNet by using common vocabularies, adopting the ISO 19115 metadata standard for all metadata directories, using harmonised Data Transport Formats for data sets delivery and using common quality control protocols and flag scales.' To the right, an 'INFOS' section indicates the page was 'Last modified on 26/09/2017'. Below that, a 'KEYWORDS' section features three tags: 'standardisation', 'software', and 'interoperability'. A vertical 'FEEDBACK' button is on the far right. The bottom of the page features four blue boxes: 'COMMON VOCABULARIES' with a word cloud, 'METADATA FORMATS' with a tag icon, 'DATA TRANSPORT FORMATS' with another word cloud, and 'DATA QUALITY CONTROL' with a line graph.

 PAN-EUROPEAN INFRASTRUCTURE FOR OCEAN & MARINE DATA MANAGEMENT

ABOUT US METADATA DATA ACCESS STANDARDS SOFTWARE PRODUCTS EVENTS PUBLICATIONS

Home > Standards

STANDARDS

Interoperability is the key to distributed data management system success and it is achieved in SeaDataNet by using common vocabularies, adopting the ISO 19115 metadata standard for all metadata directories, using harmonised Data Transport Formats for data sets delivery and using common quality control protocols and flag scales.

INFOS

Last modified on 26/09/2017

KEYWORDS

standardisation software interoperability

FEEDBACK

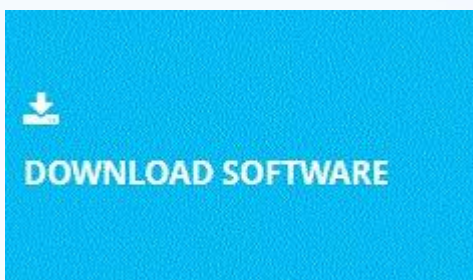
COMMON VOCABULARIES

METADATA FORMATS

DATA TRANSPORT FORMATS

DATA QUALITY CONTROL

Tools



SOFTWARE

INFOS

 Last modified on
15/09/2019

A major objective and challenge in SeaDataNet is to provide an integrated and harmonised overview and access to data resources, managed by distributed data centres. Moreover it is an objective to provide users common means for analysing and presenting data and data products. Therefore the Technical Task Team of SeaDataNet has designed an overall system architecture, and is developing common software tools for data centres and users.

Common software tools are being developed and freely made available to **Data Centres and/or End Users** for:

- Editing and generating XML metadata entries: [MIKADO javatool](#)
- Tool for the generation of spatial objects from vessel navigation during observations: [EndsAndBends](#)
- SeaDataNet file format converter : [OCTOPUS](#)
- Conversion of any ASCII format to the SeaDataNet ODV4 ASCII format: [NEMO javatool](#)
- Connecting systems of Data Centres to the SeaDataNet portal for data access: [Replication Manager javatool](#)
- Analysing and visualising of data sets: [Ocean Data View \(ODV\) software package](#)
- Interpolation and variational analysis of data sets: [DIVA software package](#)



Capacity building

- Around every 2 years: Training for all data centres connected to SDN infrastructure



End of part 2



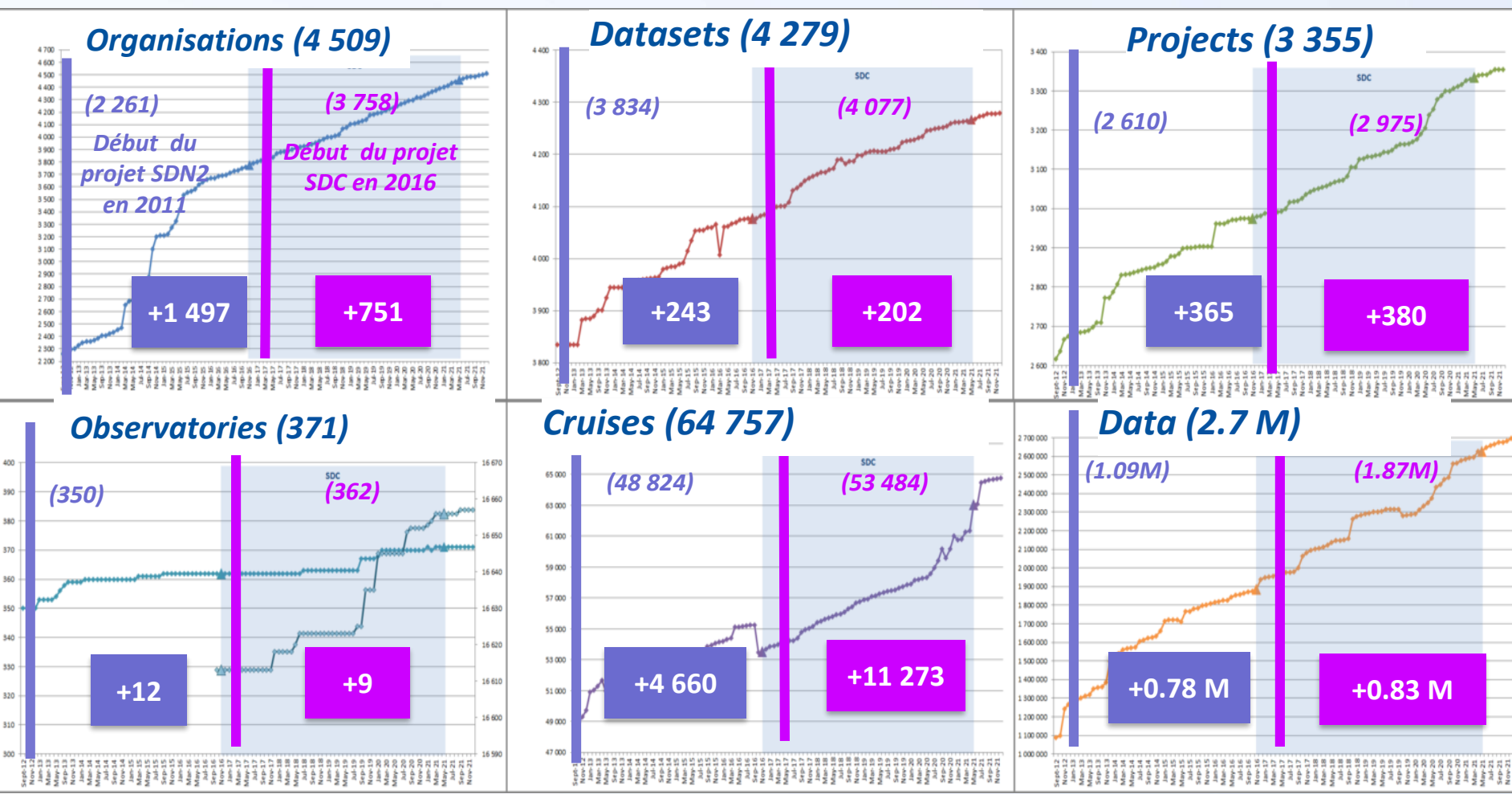
Part 3

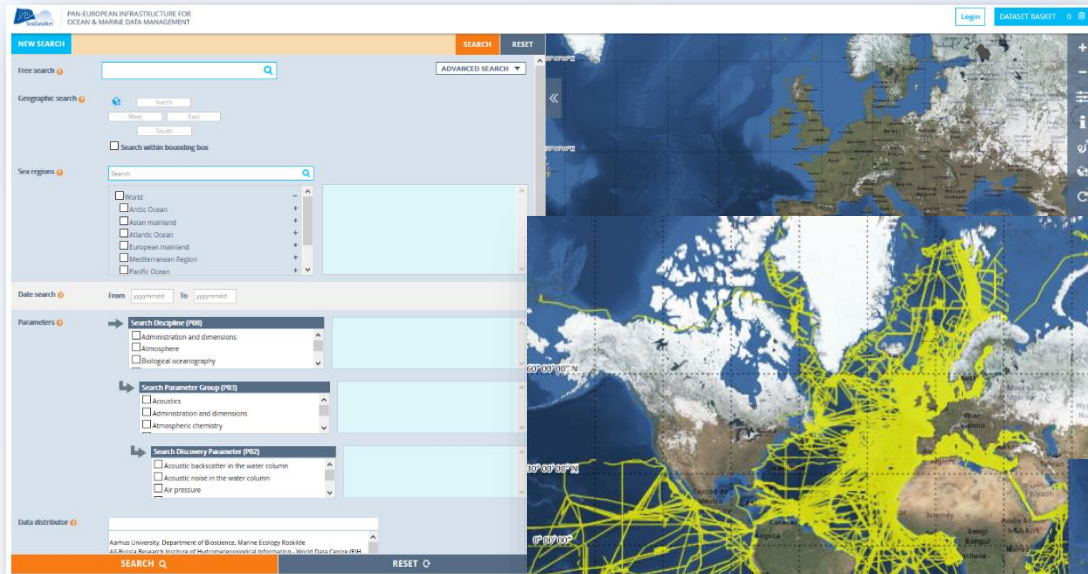
The SeaDataNet infrastructure

- Current content
- Example of proposed services

Current content of the SDN infrastructure

- Content of the catalogues





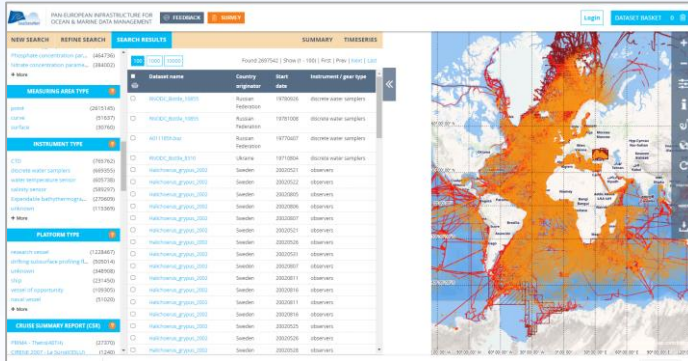
← Search interface
cdi.seadatanet.org/search

Trajectories →

Vertical profiles or time series →

- since 1800 ? 2022
- 2.69 Millions of CDI for physics, chemistry, biology, geology and geophysics
- 91 % free access after Marine-ID connection

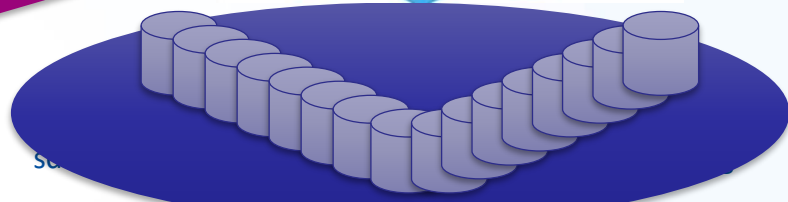
Service to users: Discovery and Access to data



Data discovery



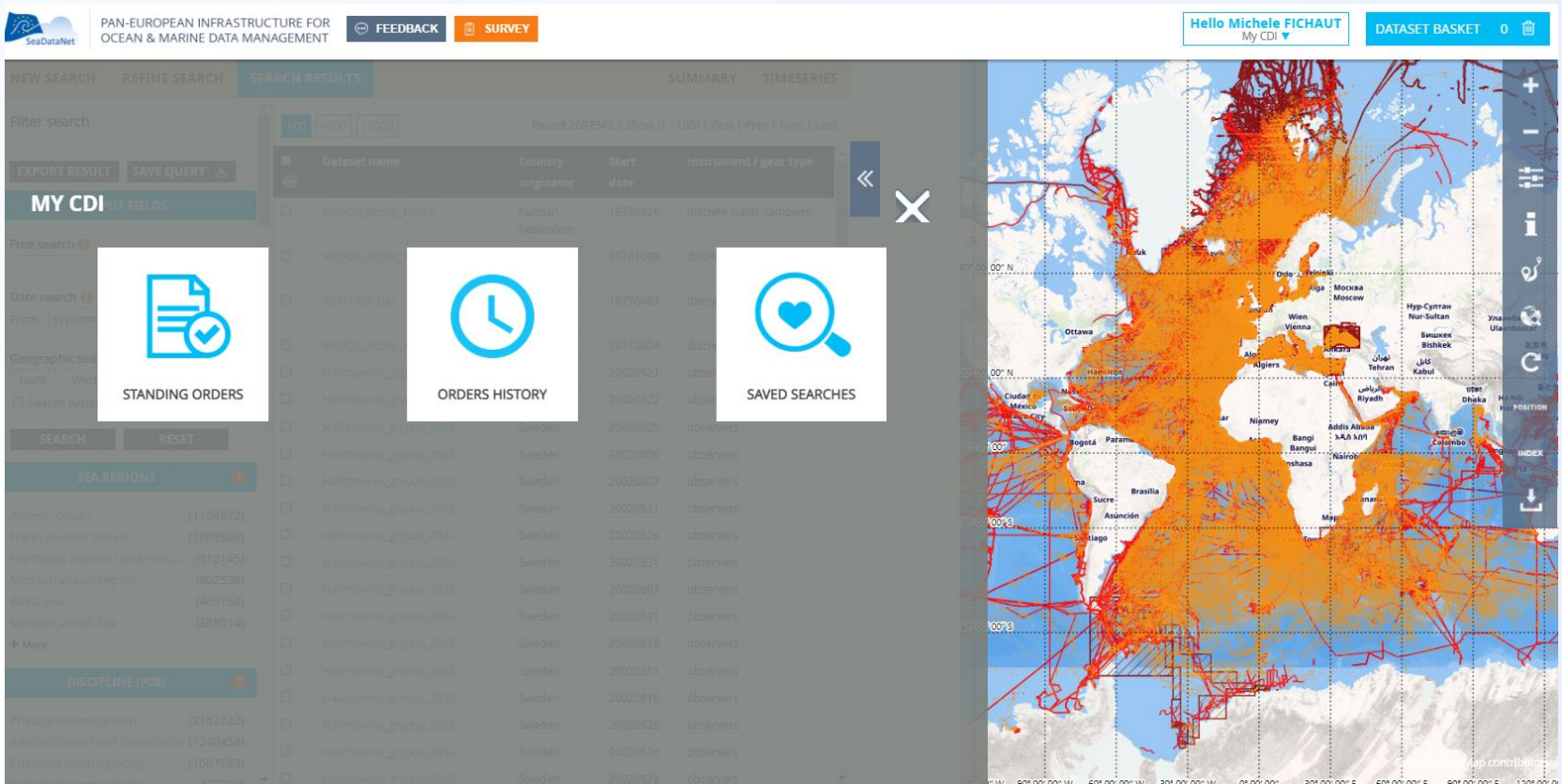
At the end of the 1st project (2011): 40 data centres, data distributed in NODCs
2nd project (2016): 80 data centres, data distributed in NODCs
Currently (2022): 113 data centres, public data duplicated in the Cloud



Data sources =
Data centres fed by
more than 800 collecting laboratories

User personal environment: MySeaDataCloud

- To be able to save and share searches, to follow the status of data requests, to have an history of data downloads



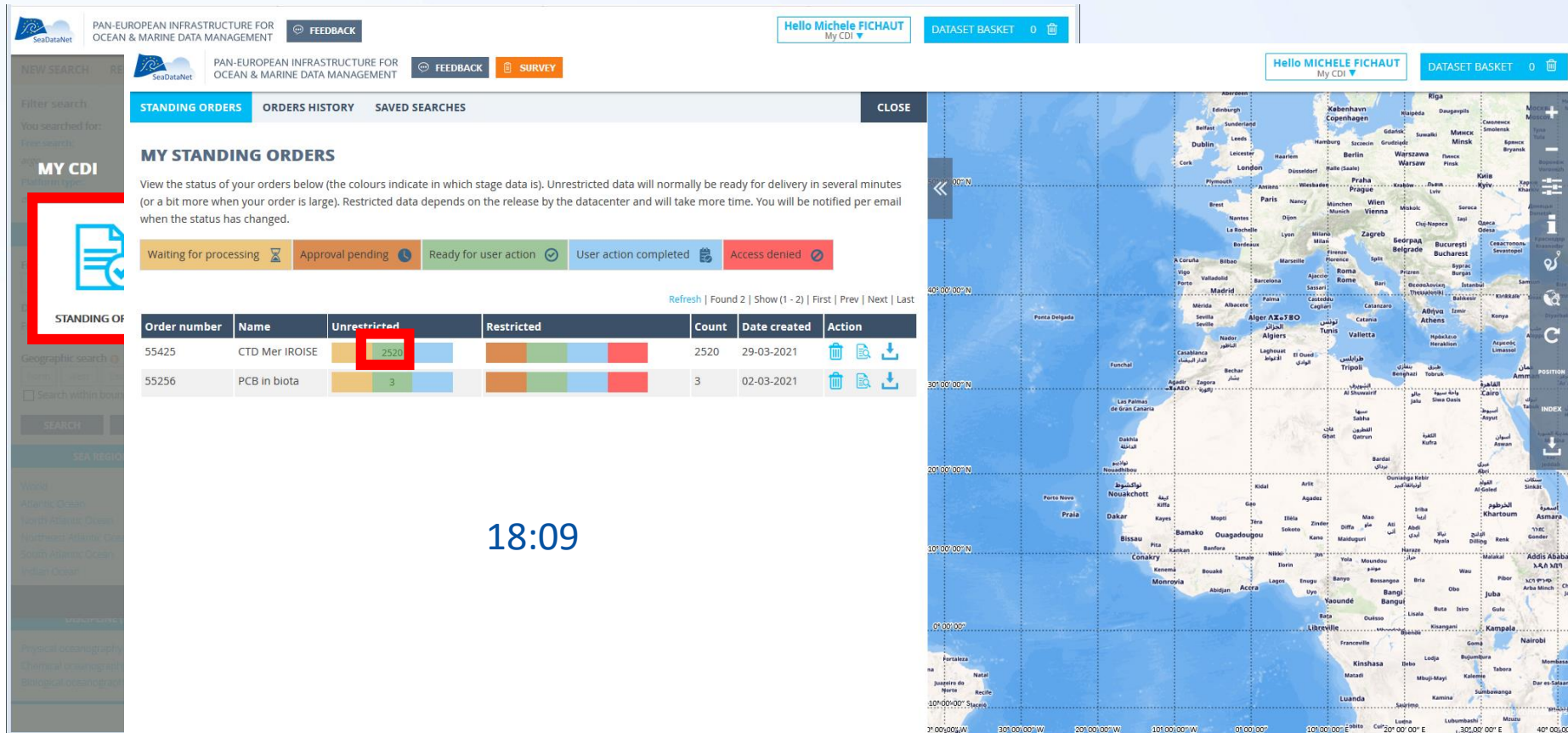
The screenshot displays the MySeaDataCloud user interface. At the top, there is a navigation bar with the SeaDataNet logo, the text "PAN-EUROPEAN INFRASTRUCTURE FOR OCEAN & MARINE DATA MANAGEMENT", and buttons for "FEEDBACK" and "SURVEY". On the right, the user is identified as "Hello Michele FICHAUT" with a "My CDI" dropdown and a "DATASET BASKET" showing 0 items.

The main interface is divided into several sections:

- Search Results:** A table listing datasets with columns for "Dataset name", "Country originator", "Start date", and "Instrument / gear type". The table shows 100 results, with the first few rows visible.
- Navigation and Actions:** A sidebar on the left includes "NEW SEARCH", "REFINE SEARCH", and "SEARCH RESULTS". Below the search results, there are buttons for "EXPORT RESULT" and "SAVE QUERY".
- Personal Environment:** A central panel titled "MY CDI" contains three main icons: "STANDING ORDERS" (document with checkmark), "ORDERS HISTORY" (clock), and "SAVED SEARCHES" (magnifying glass with heart).
- Map:** A large map on the right side shows a global view with a grid and various data points plotted, primarily concentrated in the Atlantic and Indian Oceans.

User personal environment: MySeaDataCloud

- Standing orders



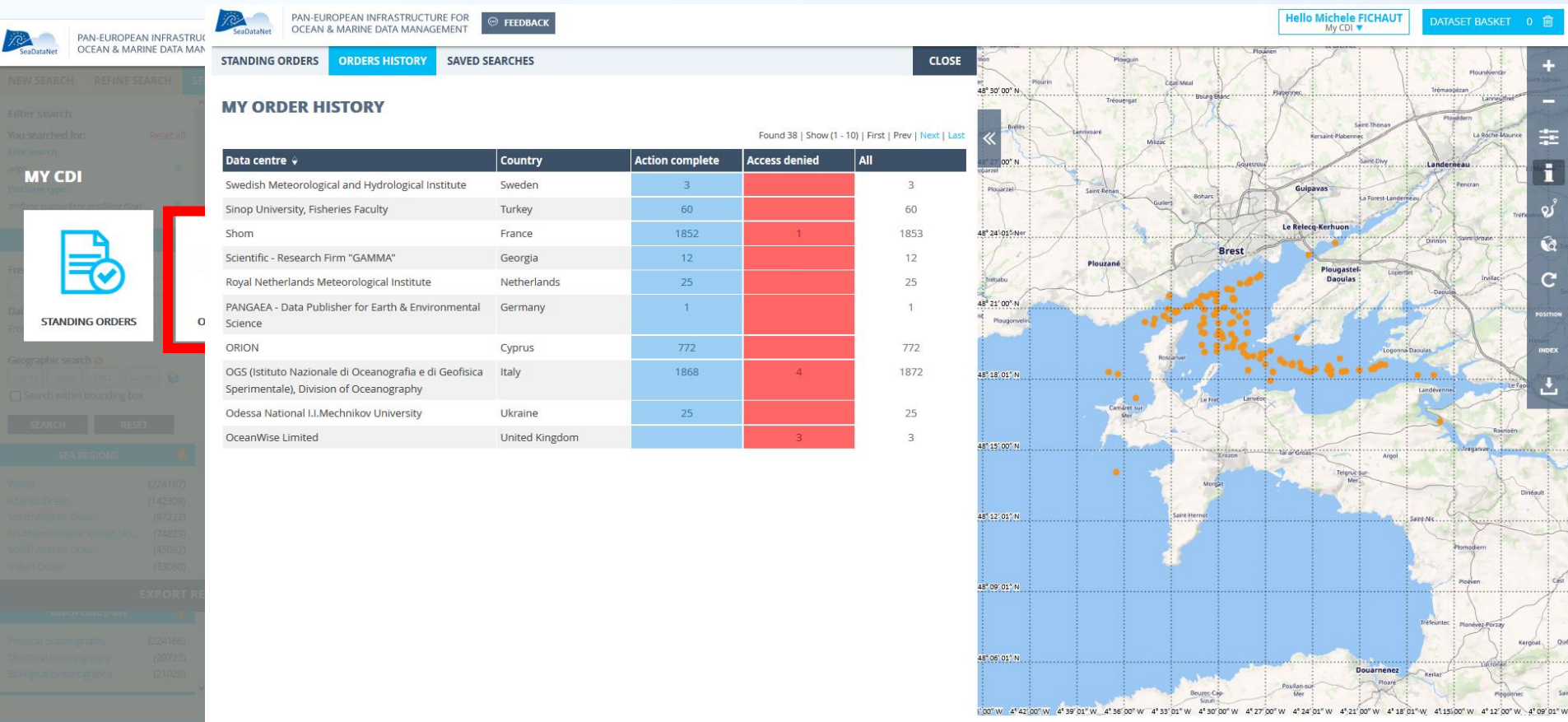
The screenshot displays the MySeaDataCloud user interface. At the top, there are navigation elements including the SeaDataNet logo, user information for 'Hello Michele FICHAUT', and a 'DATASET BASKET' with 0 items. The main content area is titled 'MY STANDING ORDERS' and includes a status legend with categories: 'Waiting for processing', 'Approval pending', 'Ready for user action', 'User action completed', and 'Access denied'. Below this is a table of standing orders.

Order number	Name	Unrestricted	Restricted	Count	Date created	Action
55425	CTD Mer IROISE	2526		2520	29-03-2021	[Icons]
55256	PCB in biota	3		3	02-03-2021	[Icons]

On the right side of the interface, there is a map of Europe and Africa. A red box highlights the 'STANDING ORDERS' tab in the left sidebar. The time '18:09' is displayed in the bottom center of the screenshot.

User personal environment: MySeaDataCloud

- Orders history



The screenshot displays the MySeaDataCloud user interface. At the top, there is a navigation bar with 'STANDING ORDERS', 'ORDERS HISTORY' (highlighted), and 'SAVED SEARCHES'. A 'CLOSE' button is visible on the right. The user's name 'Hello Michele FICHAUT' and a 'DATASET BASKET' with 0 items are shown in the top right corner.

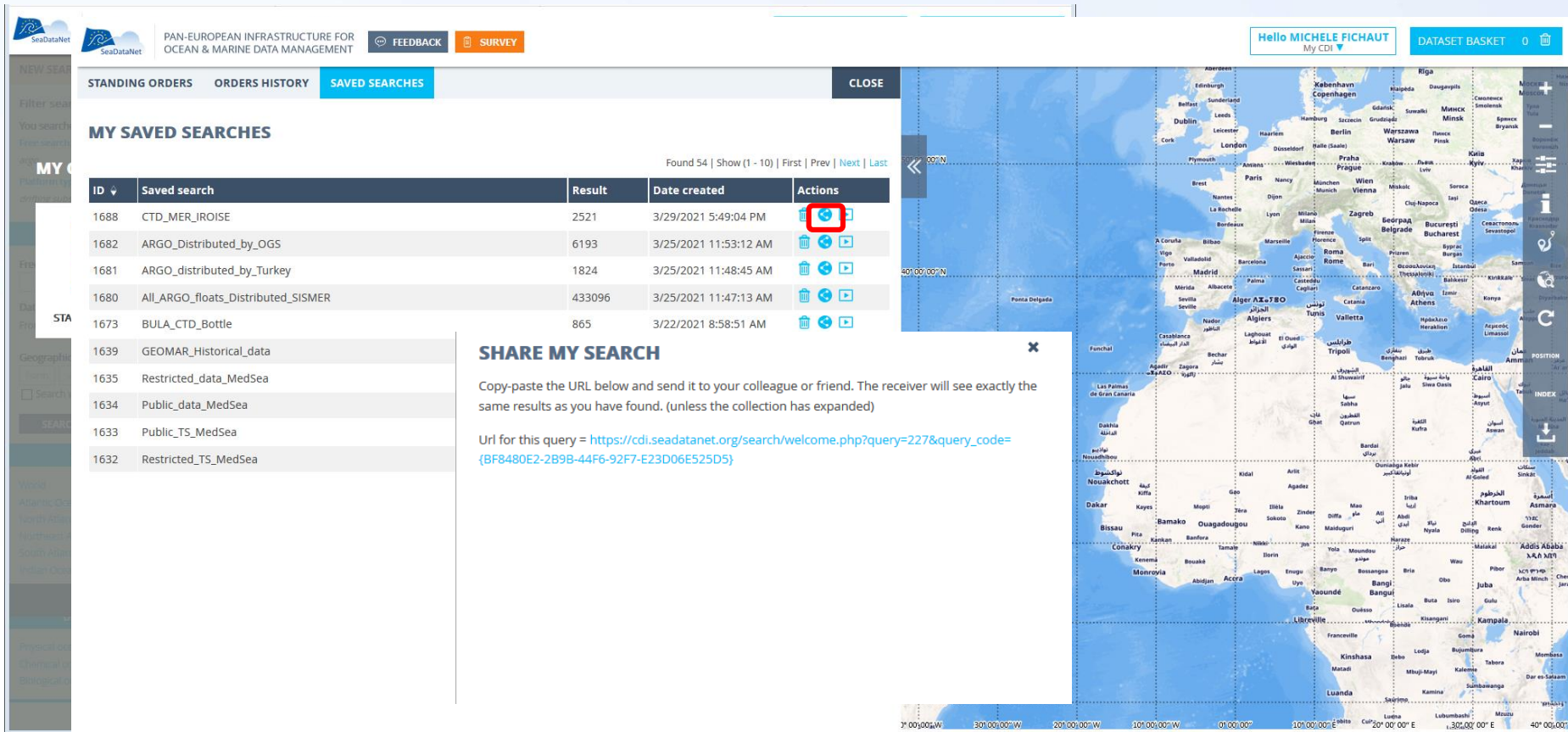
The main section is titled 'MY ORDER HISTORY' and shows a table of orders. The table has columns for 'Data centre', 'Country', 'Action complete', 'Access denied', and 'All'. The data is as follows:

Data centre	Country	Action complete	Access denied	All
Swedish Meteorological and Hydrological Institute	Sweden	3		3
Sinop University, Fisheries Faculty	Turkey	60		60
Shom	France	1852	1	1853
Scientific - Research Firm "GAMMA"	Georgia	12		12
Royal Netherlands Meteorological Institute	Netherlands	25		25
PANGAEA - Data Publisher for Earth & Environmental Science	Germany	1		1
ORION	Cyprus	772		772
OGS (Istituto Nazionale di Oceanografia e di Geofisica Sperimentale), Division of Oceanography	Italy	1868	4	1872
Odessa National I.L.Mechnikov University	Ukraine	25		25
OceanWise Limited	United Kingdom		3	3

Below the table, there is a map of the Brest region in France, showing various locations like Plouzané, Brest, Gulpavas, and Plougastel-Daoulas. The map includes a grid of latitude and longitude coordinates and a sidebar with navigation controls.

User personal environment: MySeaDataCloud

- Saved searches



The screenshot shows the MySeaDataCloud user interface. At the top, there is a navigation bar with 'STANDING ORDERS', 'ORDERS HISTORY', and 'SAVED SEARCHES' (which is highlighted). Below this, a table titled 'MY SAVED SEARCHES' lists several searches with columns for ID, Saved search, Result, Date created, and Actions. A red box highlights the share icon in the Actions column for the first search (ID 1688).

ID	Saved search	Result	Date created	Actions
1688	CTD_MER_IROISE	2521	3/29/2021 5:49:04 PM	[Share icon highlighted]
1682	ARGO_Distributed_by_OGS	6193	3/25/2021 11:53:12 AM	[Share icon]
1681	ARGO_distributed_by_Turkey	1824	3/25/2021 11:48:45 AM	[Share icon]
1680	All_ARGO_floats_Distributed_SISMER	433096	3/25/2021 11:47:13 AM	[Share icon]
1673	BULA_CTD_Bottle	865	3/22/2021 8:58:51 AM	[Share icon]

A 'SHARE MY SEARCH' dialog box is open, providing instructions and a URL for sharing the search results.

SHARE MY SEARCH

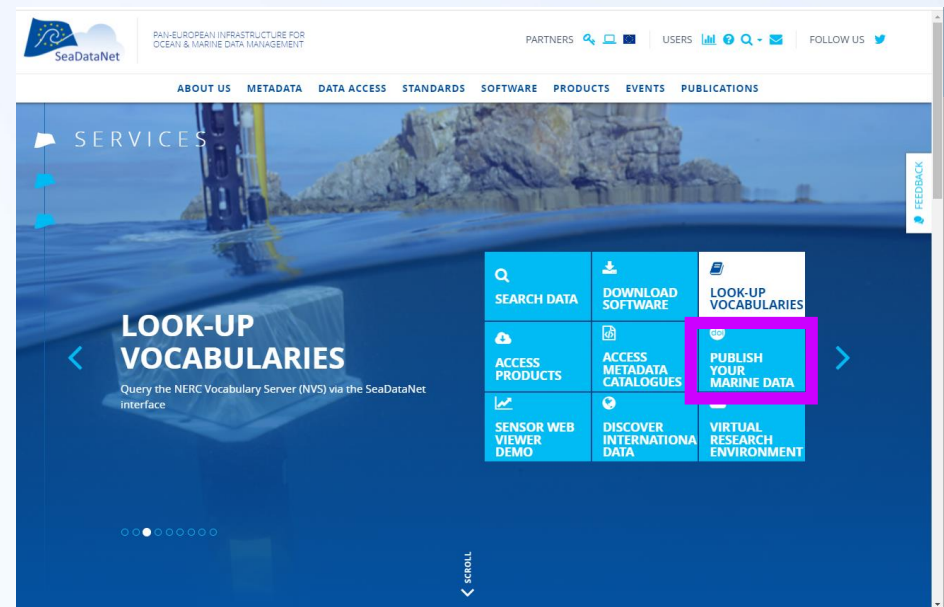
Copy-paste the URL below and send it to your colleague or friend. The receiver will see exactly the same results as you have found. (unless the collection has expanded)

Url for this query = https://cdi.seadatanet.org/search/welcome.php?query=227&query_code={BF480E2-2B98-44F6-92F7-E23D06E525D5}

Publishing service

SEANOE Sea scientific open data publication

- To offer scientists a place to publish their datasets and get DOI (Digital Object Identifier)
 - Useful for publications based on datasets, publisher often require a DOI on the studied dataset



Publishing service : landing page

☐ Cruises PRIMO 1994 (CNR IT): Hydrographic measurements in the Sicily Channel and in the southern Tyrrhenian Sea (spring and fall 1994)

Date 2021-12-13

Temporal extent 1994-05-20 -1994-10-18

Author(s) [Sparnocchia Stefania](#)^{ORCID}¹, Borghini Mireno²

Affiliation(s) 1 : CNR-ISMAR, Trieste, Italy
2 : CNR-ISMAR, La Spezia, Italy

DOI [10.17882/85185](https://doi.org/10.17882/85185)

Publisher SEANOE

Keyword(s) CTD, Sicily Channel, Tyrrhenian Sea, Western Mediterranean

Abstract This data set contains the CTD data collected from the RV URANIA of the CNR (Italy) during the PRIMO-94 and PRIMO-94B cruises (20-29 May 1994 and 5-18 October 1994). These cruises were part of an intensive field program in the Sicily Channel and in the southern Tyrrhenian basin conducted by the Stazione Oceanografica of CNR in different periods from 1993 to 1995. Data have been used in several studies (see References).

CTD profiles were collected using a Neil-Brown MKIII CTD. The probe was calibrated in temperature and conductivity at the SACLANT Center of La Spezia, before and after each cruise, and at sea in salinity and oxygen, against water samples. Declared instrumental precisions were 0.002 °C for temperature and 0.005 for salinity (PSS-78).

The data set is provided per cruise as ODV Spreadsheet files in TXT format, containing:

- Cruise name
- Station number
- Type of acquisition (here C)
- Date in mon/day/yr and Time in hh:mm:ss
- Coordinates in Longitude [degrees_east] and Latitude [degrees_north]
- Bottom depth [m]
- Depth, salt water [m]
- Temperature, IPTS-68 [degC]
- Conductivity [S/m]
- Temperature, ITS-90 [degC]
- Salinity, PSS-78 (Practical Salinity)
- Dissolved oxygen [ml/l]

Click to download the data

[DATA](#)

Utilisation These data are published without any warranty, express or implied. The user assumes all risk arising from their use. These data are intended to be quality controlled, but it is possible that they contain errors. It is the unique responsibility of the user to assess if the data are appropriate for their use, and to interpret the data, data quality, and data accuracy accordingly. Authors welcome users to ask questions and report problems.

Acknowledgements This data set was collected by the group known as the Stazione Oceanografica (Oceanographic Station) of the CNR, Pozzuolo di Lerici, La Spezia, led by Mario Astraldi and Gian Pietro Gasparini. We are grateful to Mr. Carlo Galli, Mr. Eglisto Lazzoni and Mr. Domenico Bacciola for their remarkable contribution in the field and in the laboratory work. The research was funded by the EU through the MAST program (Contracts MAS2-CT93-0061 GEODYME and MAS2-CT93-0066 EUROMODEL II-MTP). The experiment in the Sicily Channel was a contribution to the IOC Programmes POEM and PRIMO.

Sensor metadata Neil-Brown MK III CTD

Data	File	Size	Format	Processing	Access
	CTD Data from PRIMO-94	9 MB	ODV		Open access
	CTD Data from PRIMO-94B	9 MB	ODV		Open access

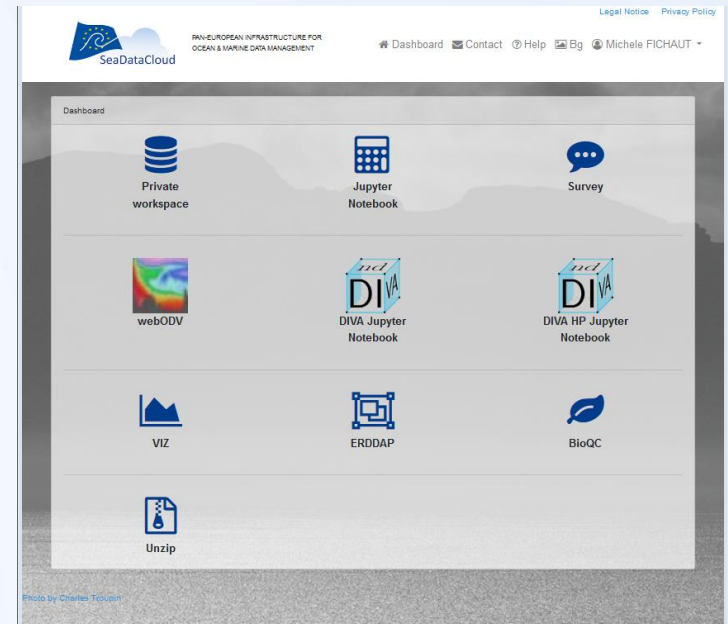
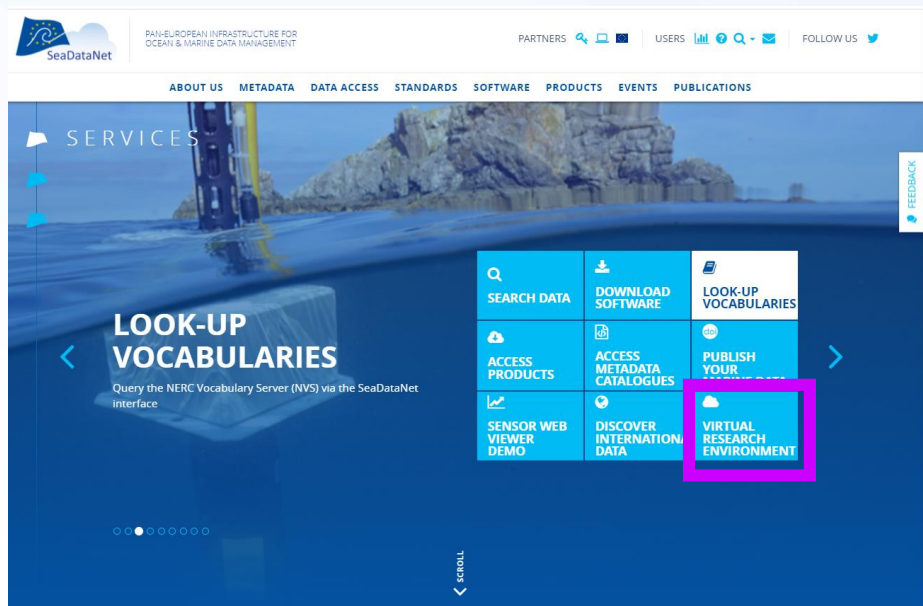


How to cite

Sparnocchia Stefania, Borghini Mireno (2021). Cruises PRIMO 1994 (CNR IT): Hydrographic measurements in the Sicily Channel and in the southern Tyrrhenian Sea (spring and fall 1994). SEANOE. <https://doi.org/10.17882/85185>

Virtual Research Environment for users

- A virtual environment in the Cloud,
 - With data and tools
 - Possibility to add his/her own data



End of part 3



Part 4

FAIR Principles and data FAIRness

- Presentation of the FAIR principles
- Evaluation of the data FAIRness
- Example of SDN French physical/chemical cruises data distributed in SeaDataNet

FAIR data and metadata

- 2016, Paper « The FAIR Guiding Principles for scientific data management and stewardship” par Wilkinson, Dumontier, & Aalbersberg
<https://doi.org/10.1038/sdata.2016.18>
- The FAIR Principles = a set of guidelines for managing research data
- aiming to make them **F**indable, **A**ccessible, **I**nteroperable and **R**e-usable by both humans and machines

Findable

Metadata and data should be findable for both humans and computers

F

Interoperable

Data needs to work with applications or workflows for analysis, storage and processing

I

Accessible

Once found, users need to know how the data can be accessed

A

Reusable

The goal of FAIR is to optimise data reuse via comprehensive well-described metadata

R



FAIR principles

F : Findable

- ✓ F1. (Meta)data are assigned a globally unique and persistent identifier
- ✓ F2. Data are described with rich metadata (defined by R1 below)
- ✓ F3. Metadata clearly and explicitly include the identifier of the data they describe
- ✓ F4. (Meta)data are registered or indexed in a searchable resource

A : Accessible

- ✓ A1. (Meta)data are retrievable by their identifier using a standardised communications protocol.
 - ✓ A1.1. The protocol is open, free, and universally implementable.
 - ✓ A1.2. The protocol allows for an authentication and authorisation procedure, where necessary
- ✓ A2. Metadata are accessible, even when the data are no longer available

I : Interoperable

- ✓ I1. (Meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
- ✓ I2. (Meta)data use vocabularies that follow FAIR principles
- ✓ I3. (Meta)data include qualified references to other (meta)data

R : Reusable

- ✓ R1. (Meta)data are richly described with a plurality of accurate and relevant attributes
 - ✓ R1.1. (Meta)data are released with a clear and accessible data usage license
 - ✓ R1.2. (Meta)data are associated with detailed provenance
 - ✓ R1.3. (Meta)data meet domain-relevant community standards

RDA - FAIR Data Maturity Model (FDMM) WG

- Specification and guidelines:
<https://doi.org/10.15497/RDA00050>
- Indicators that aim to formulate measurable aspects of each principle that can be used by assessment approaches
- Generic self-assessment model to measure the level of maturity of a dataset

FAIR Data Maturity Model indicators



- 41 indicators have been defined by the WG RDA-FDMM

- **F1** : 7 indicators
- **A** : 12 indicators
- **I** : 12 indicators
- **R** : 10 indicators

Indicators are classified as

- **Essential**: utmost importance
- **Important**: substantially increase FAIRness
- **Useful**: nice to have

Priority	Principle				Grand Total
	Findable	Accessible	Interoperable	Reusable	
Essential	7	8	0	5	20
Important	0	3	7	4	14
Useful	0	1	5	1	7
Grand Total	7	12	12	10	41

FAIR Data Maturity Model indicators



- 7 indicators for FINDABLE (all essentials) - identifiers
 1. **F1-O1M**: Metadata identified by a persistent identifier
 2. **F1-O1D**: Data identified by a persistent identifier
 3. **F1-O2M**: Metadata is identified by a globally unique identifier
 4. **F1-O2D**: Data is identified by a globally unique identifier
 5. **F2-O1M**: Rich metadata is provided to allow discovery
 6. **F3-O1M**: Metadata includes the identifier of the data
 7. **F4-O1M**: Metadata is offered in such a way that it can be harvested and indexed

FAIR Data Maturity Model indicators



- 12 indicators for ACCESSIBLE (8 essential, 3 important, 1 useful)
 8. **A1-01M**: Metadata contains information to enable the user to get access to the data
 9. **A1-02M**: Metadata can be accessed manually (i.e. with human intervention)
 10. **A1-02D**: Metadata identifier resolves to a metadata record
 11. **A1-03M**: Metadata identifier resolves to a metadata record
 12. **A1-03D**: Data identifier resolves to a digital object
 13. **A1-04M**: Metadata is accessed through standardised protocol (e.g. HTTP, FTP, ...)
 14. **A1-04D**: Data is accessed through standardised protocol (e.g. HTTP, FTP, ...)
 15. **A1-05D**: Data can be accessed automatically (i.e. by a computer program)
 16. **A1.1-01M**: Metadata is accessed through a free access protocol
 17. **A1.1-01D**: Data is accessed through a free access protocol
 18. **A1.2-01D**: Data is accessed through an access protocol that's support Authentication and Authorisation
 19. **A2-01M**: Metadata is guaranteed to remain available after data is no longer available

FAIR Data Maturity Model indicators



- 12 indicators for INTEROPERABLE (7 important, 5 useful)
 20. **I1-01M**: Metadata uses knowledge representation expressed in standardised format
 21. **I1-01D**: Data uses knowledge representation expressed in standardised format
 22. **I1-02M**: Metadata uses machine-understandable knowledge representation
 23. **I1-02D**: Data uses machine-understandable knowledge representation
 24. **I2-01M**: Metadata uses FAIR-compliant vocabularies
 25. **I2-01D**: Data uses FAIR-compliant vocabularies
 26. **I3-01M**: Metadata includes references to other metadata
 27. **I3-01D**: Data includes references to other data
 28. **I3-02M**: Metadata includes references to other data
 29. **I3-02D**: Data includes qualified references to other data
 30. **I3-03M**: Metadata includes qualified references to other metadata
 31. **I3-04M**: Metadata includes qualified references to other data

FAIR Data Maturity Model indicators



- 10 indicators for REUSABLE (5 essential, 4 important, 1 useful)
 - 32. R1-01M:** Plurality of accurate and relevant attributes are provided to allow reuse
 - 33. R1.1-01M:** Metadata includes information about the licence under which the data can be used
 - 34. R1.1-02M:** Metadata refers to a standard reuse licence
 - 35. R1.1-03M:** Metadata refers to a machine-understandable reuse licence
 - 36. R1.2-01M:** Metadata includes provenance information according to community-specific standards
 - 37. R1.2-02M:** Metadata includes provenance information according to a cross-community language
 - 38. R1.3-01M:** Metadata complies with a community standard
 - 39. R1.3-01D:** Data complies with a community standard
 - 40. R1.3-02M:** Metadata is expressed in compliance with to a machine-understandable community standard
 - 41. R1.3-02D:** Data is expressed in compliance with to a machine-understandable community standard

FDMM evaluation methods

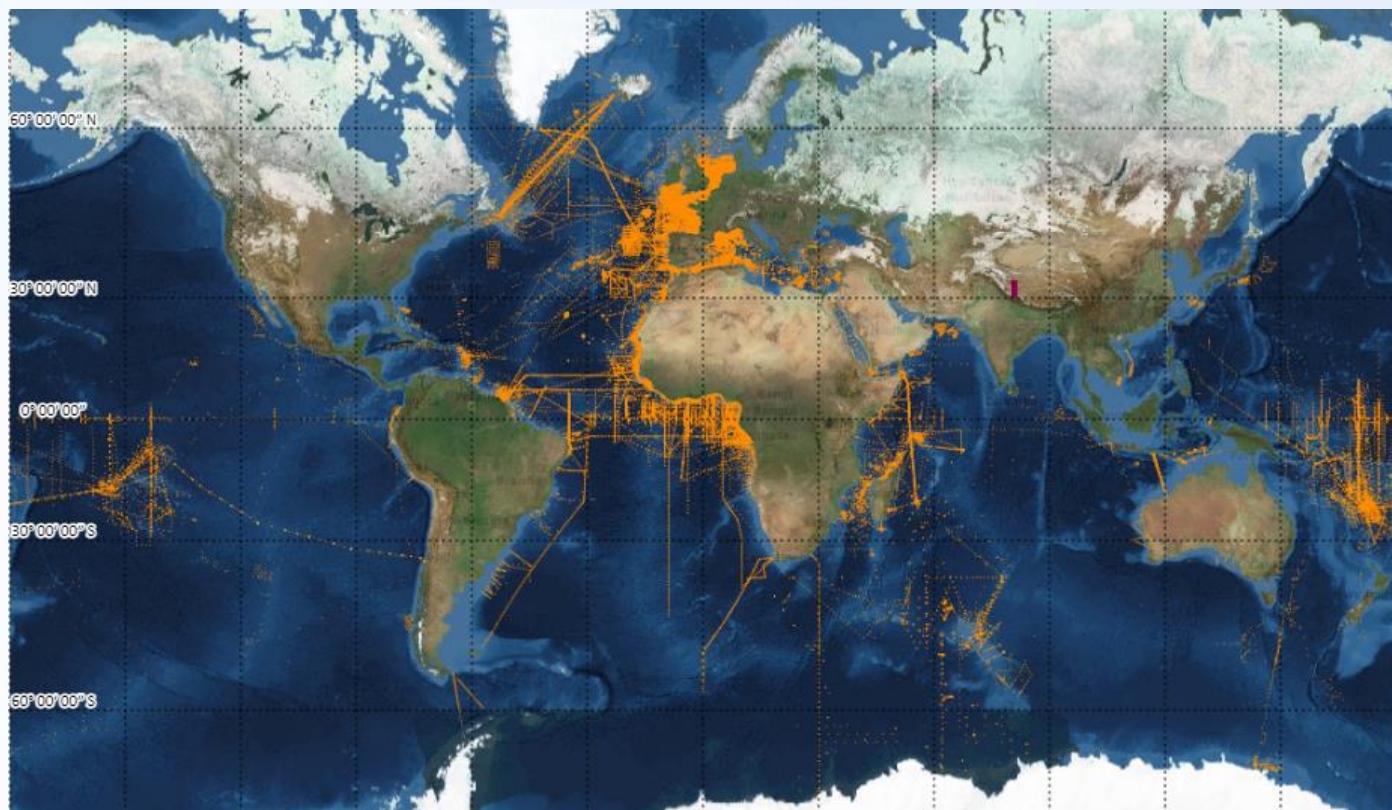


- 2 methods
 - Measuring progress
 - Delivering a measure of the extent to which a resource under evaluation meets the requirements of an expressed indicator following the scale
 - 0 : not applicable
 - 1 : not being considered yet
 - 2 : under consideration or in planning phase
 - 3 : in implementation phase
 - 4 : fully implemented
 - Measuring pass-or-fail
 - Determining whether a resource under evaluation meets the requirements of an expressed indicator on a binary pass-or-fail scale

FAIRness of the Physics-Chemistry data from the French cruises

- In SeaDataNet : 118 508 CDIs

14 717 XBTs
49 500 CTDs
51 997 Bottles
2 294 Currentmeters

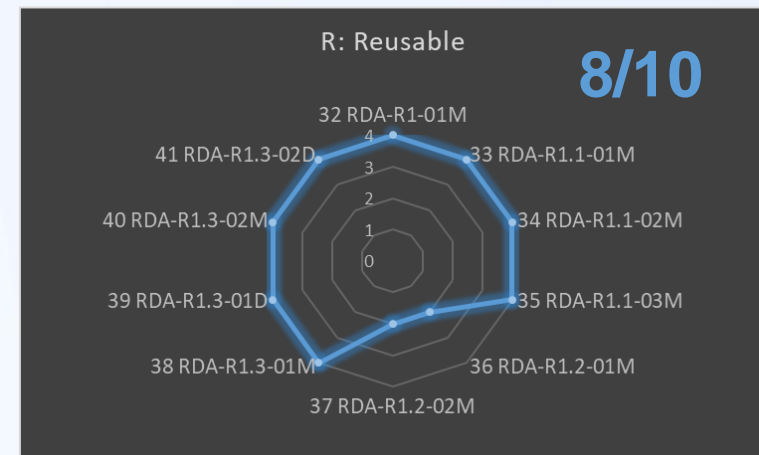
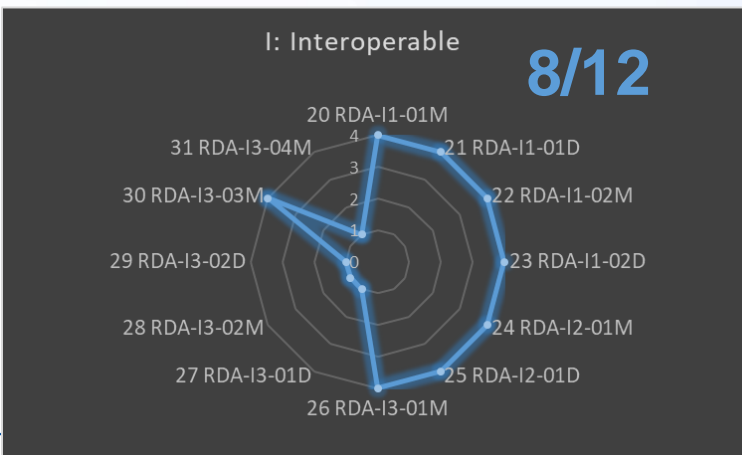
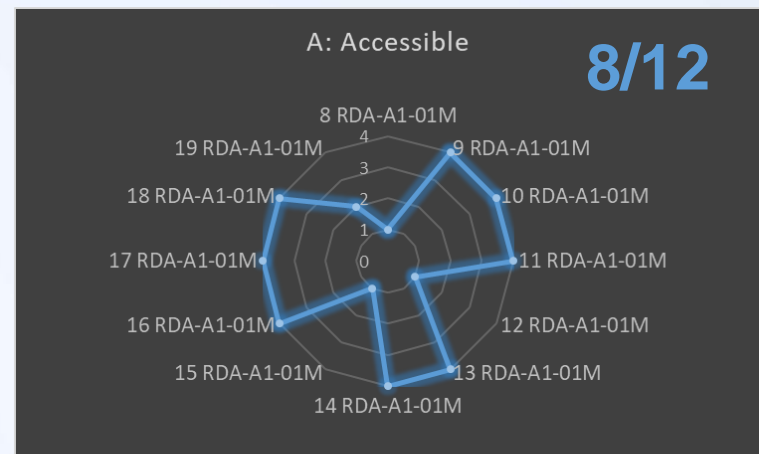
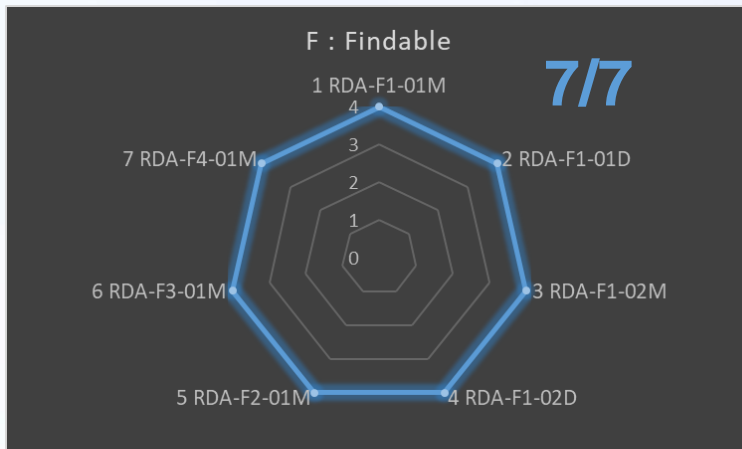


FAIRness of the Physics-Chemistry data from the French cruises

- Review of the 41 criteria previously presented and assignment of a level (0, 1, 2, 3 or 4) per criterion for this in situ dataset

Results of the evaluation per FAIR principle

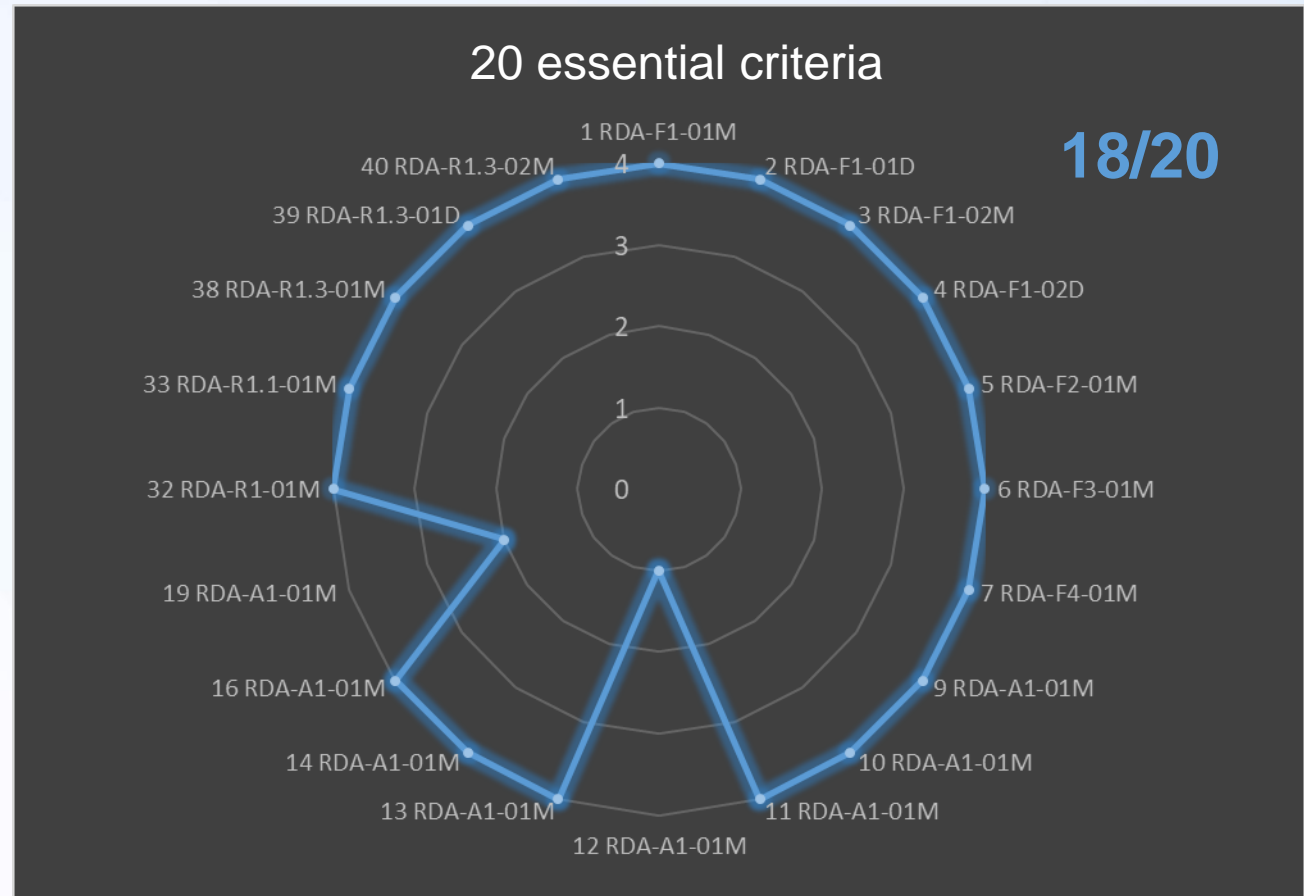
- Total 31/41



Per priority : Essential

MISSING

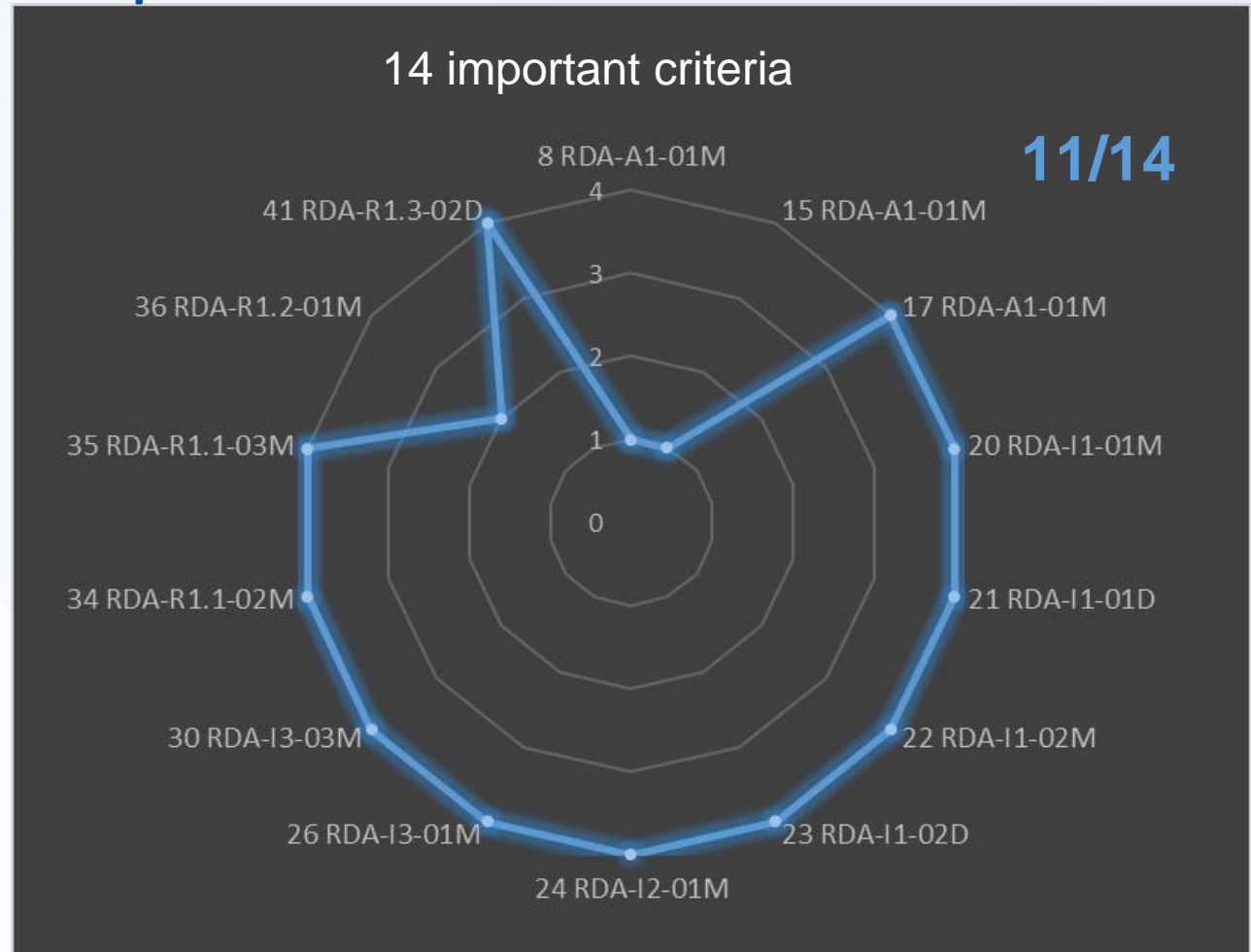
- 12 (A): direct access to data via an API for example
- 19 (A): Obsolete metadata are kept but the info is not accessible online at the moment



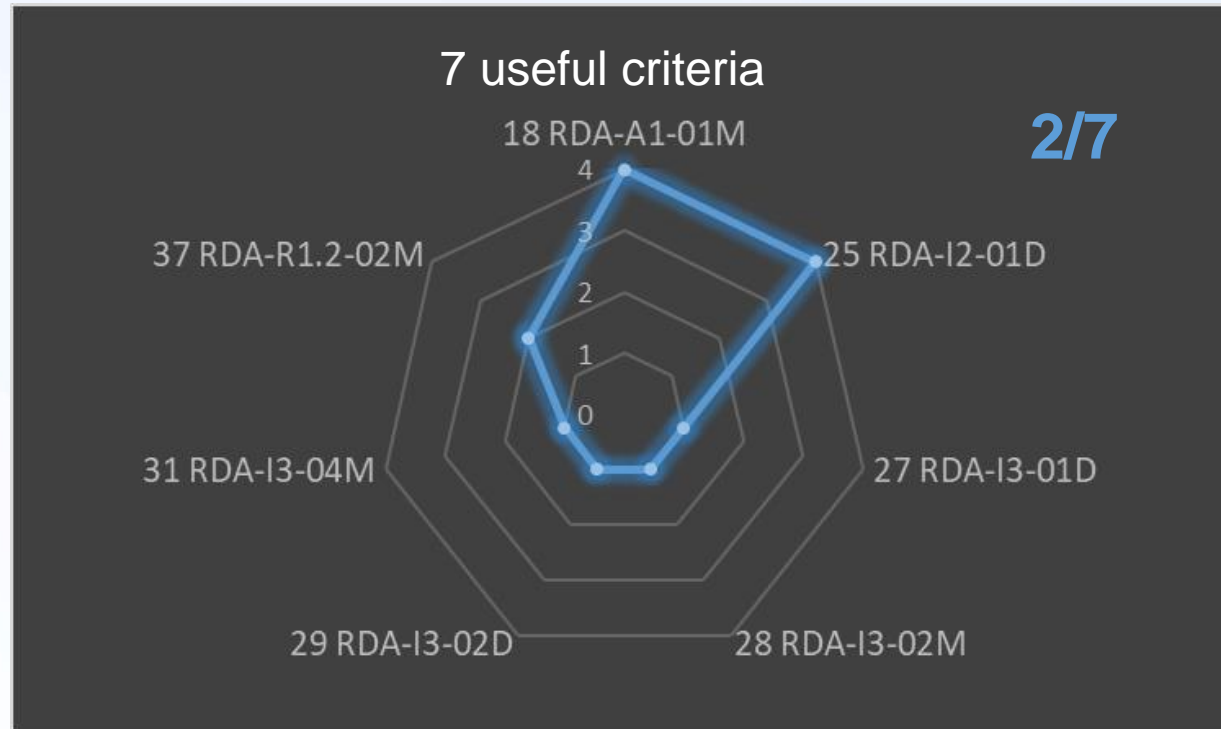
Per priority : Important

MISSING

- 8 (A): Direct access to data via metadata
- 15 (A): No API on data
- 36 (R): Provenance according to community standard



Per priority : Useful



MISSING

- 27 (I): Data with references to other data
- 28 (I): Metadata with references to other data
- 29 (I): Data with qualified references to other data
- 31 (I): Metadata with qualified references to other data

Conclusion

- Interesting exercise to see what improvements can be made to the datasets in particular with regard to the essential criteria which are “mandatory” to have FAIR data

End of part 4

