IMDIS 2018 - Barcelona, 5-7 November
International Conference on Marine Data and Information Syste...

Interoperability of new data type with SeaDataNet infrastructure:

Case of Flow Cytometry data (FCM)

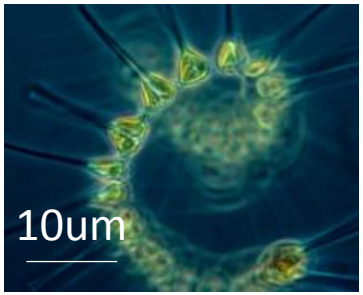
Soumaya LAHBIB, Gwenaëlle MONCOIFFÉ, Gérald GREGORI, Maurice LIBES, Michel DENIS,

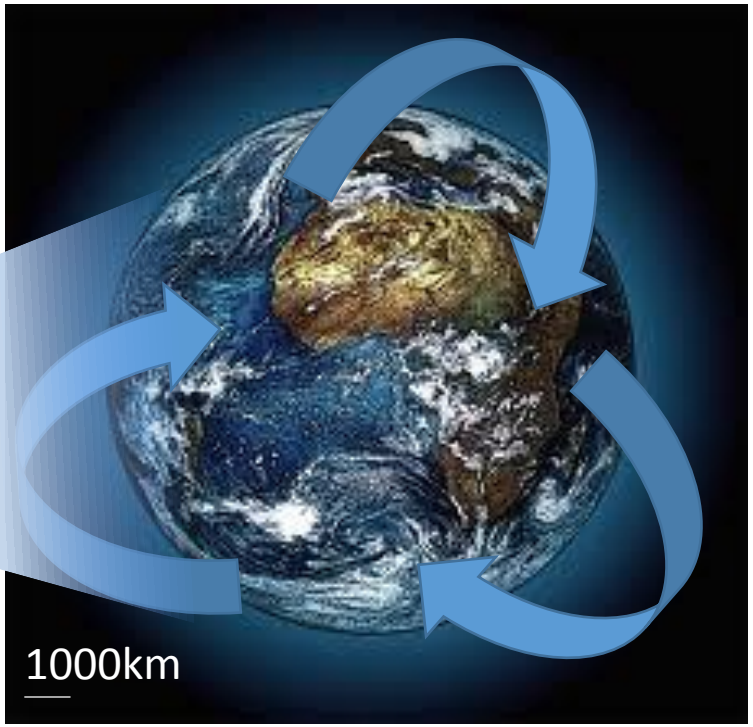
Pierre MARREC, Simon CLAUS, Michèle FICHAUT, Dick SCHAAP and Melilotus THYSSEN

Presented by Yolanda DEL AMO

Phytoplankton

Thousands of species < 1000 µm catalyze the most important geochemical processes for sustaining life on earth AND at a minute scale.

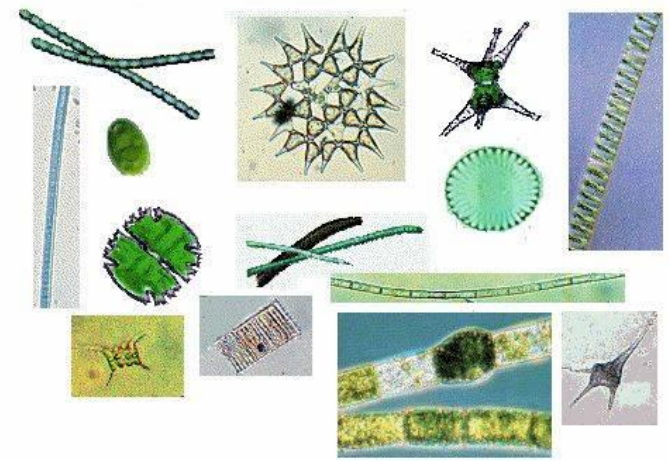




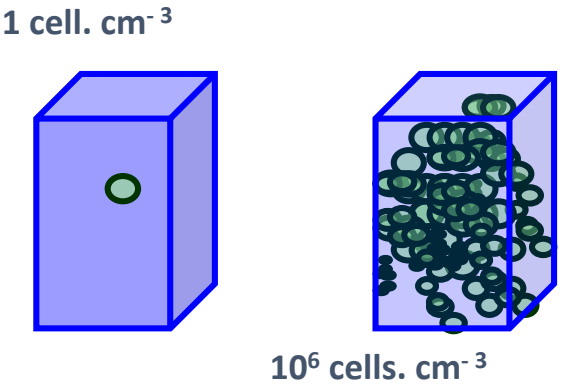
Phytoplankton produces between 45 and 57 Pg C Yr^{-1} of the NPP on earth (~45%) but represents <2% of its biomass.
Very high turn-over rate !

Phytoplankton observation is complex

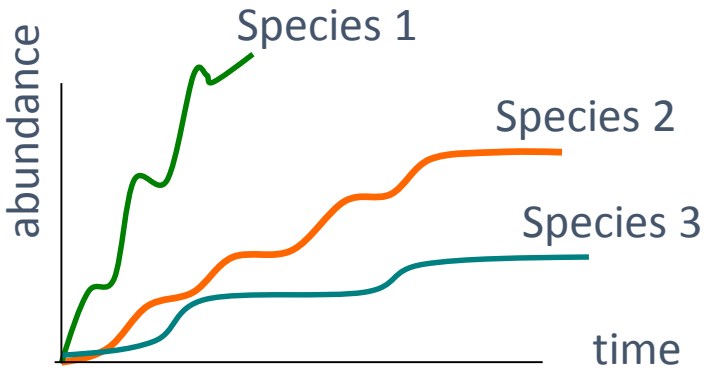
Morphology and size



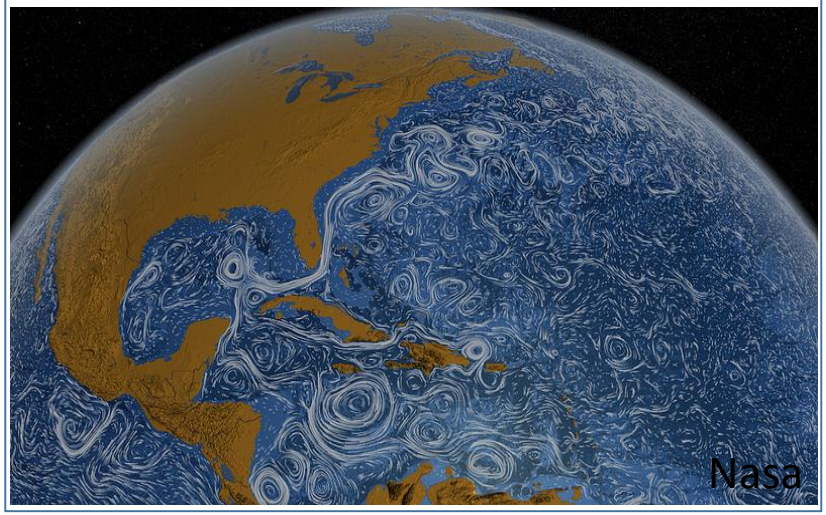
Abundances



Growth rates

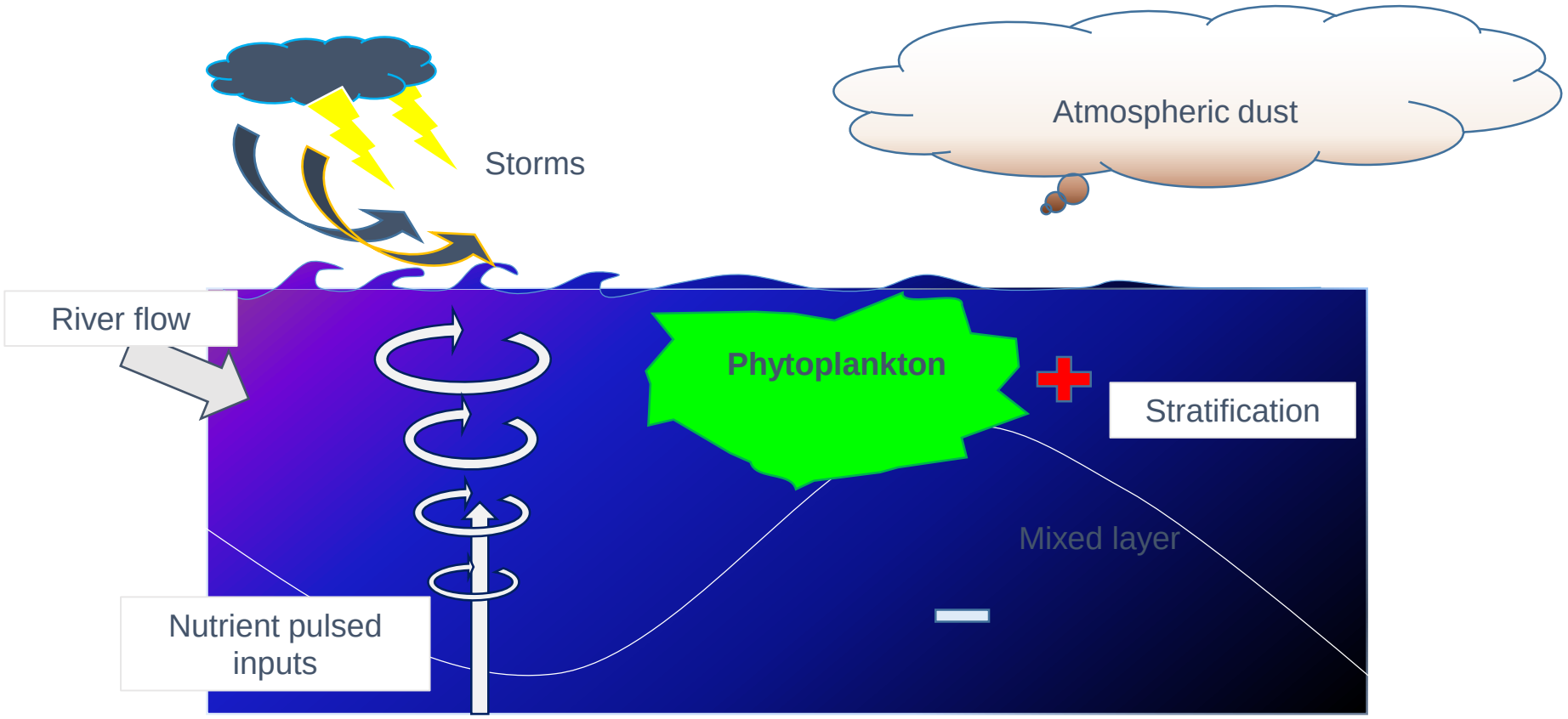


Turbulence



Serious lack in understanding and quantifying the role of phytoplankton in the biogeochemical processes

Short term variation and sporadic events impacts are nearly unknown

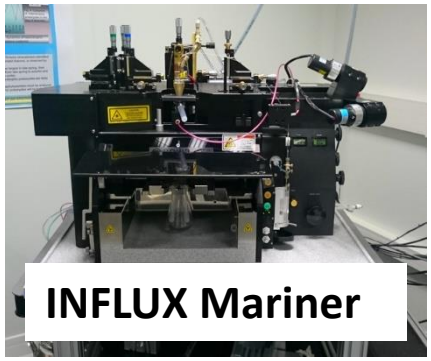


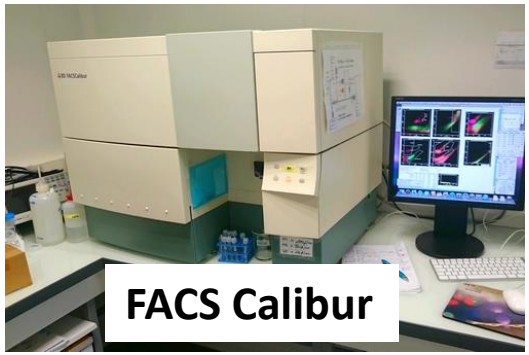
Several scientific experiences

CytoSense

CytoSub in a buoy

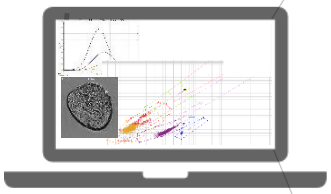
PRECYM
Plateforme REgionale de CYtométrie pour la Microbiologie
Flow cytometry made in MIO





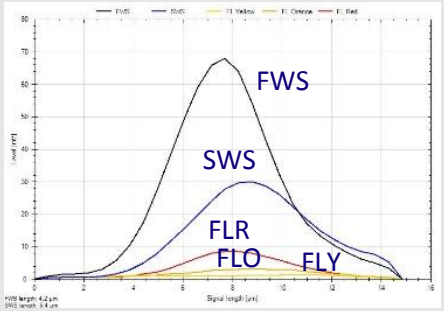
Phytoplankton functional groups resolution

Benchtop FCM

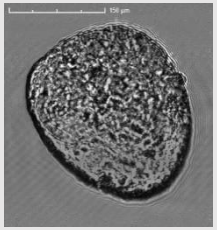


Automated FCM

1 cell

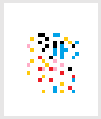


Optical properties

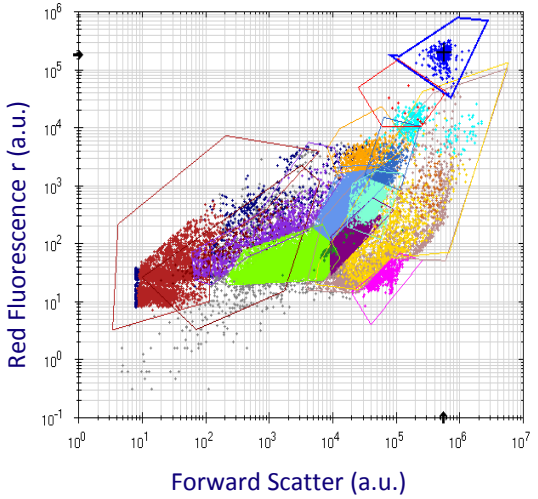


Photo

1x10^6 cells



Phytoplankton assemblage



- Phytoplankton functional groups/Phytoplankton abundance per group
- Fluorescences/scatter per cell/Size estimation after calibration of scatter
- Phytoplankton images (taxonomical identification >20 µm)

SeaDataCloud

Ingesting, validating, long-term storage and access of Flow Cytometery (FCM) data

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

Ingesting, validating, long-term storage and access of Flow Cytometery (FCM) data

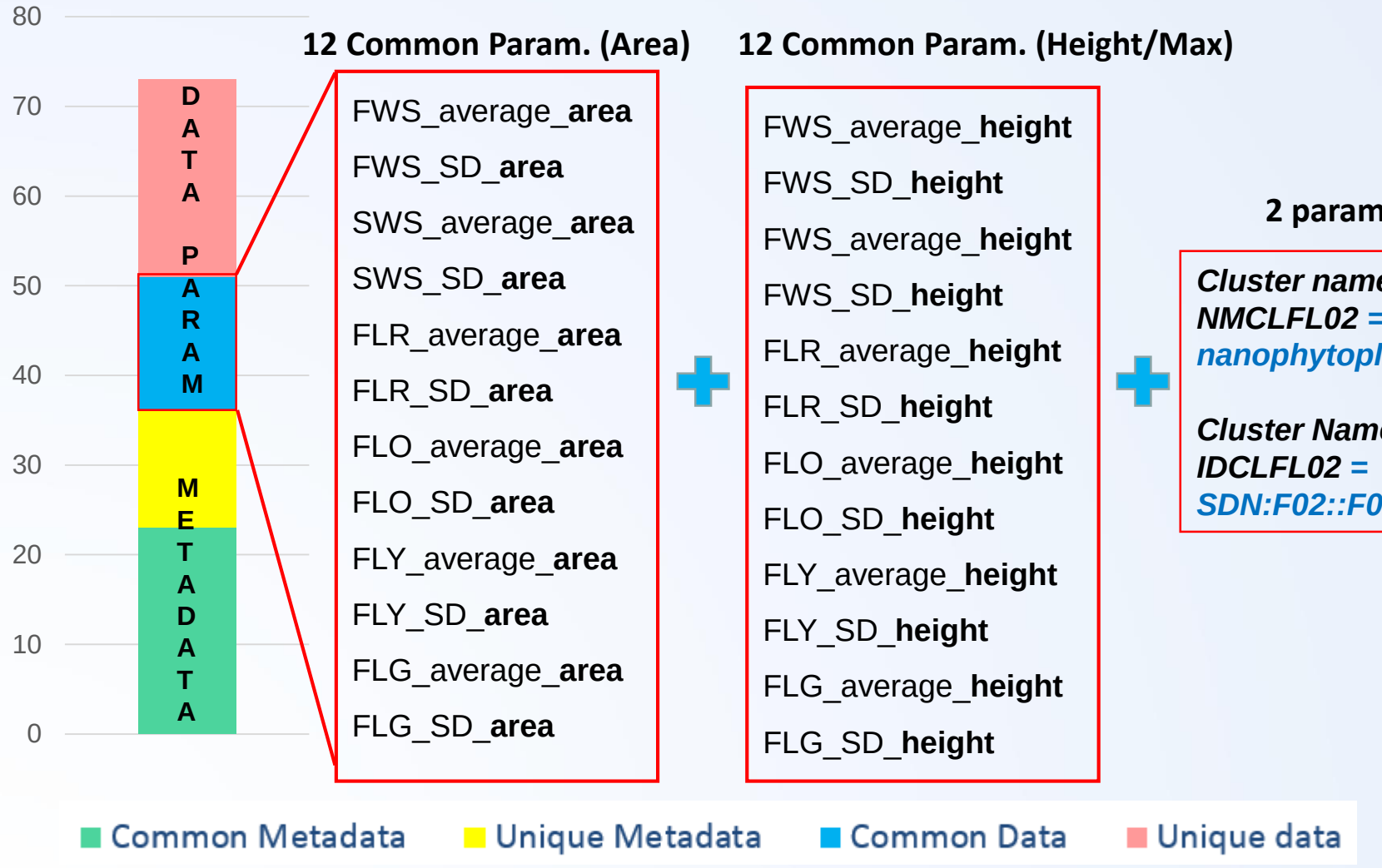
1. FCM Common Vocabulary
2. Data Transport Format for FCM data
3. Ingestion into SeaDataNet Infrastructure

SeaDataCloud

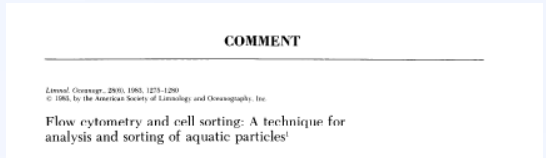
1

FCM Common Vocabulary

- Analysis of the existing codes (P01 list)
- FCM common parameters (JericoNext)
- Literature review and bibliography (1983-2017)
- Questionnaire (58 questions) to 180 FCM users


80
70
60
50
40
30
20
10
0
DATA
PARAM
METADATA
12 Common Param. (Area)
FWS_average_area
FWS_SD_area
SWS_average_area
SWS_SD_area
FLR_average_area
FLR_SD_area
FLO_average_area
FLO_SD_area
FLY_average_area
FLY_SD_area
FLG_average_area
FLG_SD_area
12 Common Param. (Height/Max)
FWS_average_height
FWS_SD_height
FWS_average_height
FWS_SD_height
FLR_average_height
FLR_SD_height
FLO_average_height
FLO_SD_height
FLY_average_height
FLY_SD_height
FLG_average_height
FLG_SD_height
2 parameters
Cluster name=>
NMCLFL02 = Eukaryote nanophytoplankton
Cluster Name ID=>
IDCLFL02 = SDN:F02::F0200005
Common Metadata
Unique Metadata
Common Data
Unique data

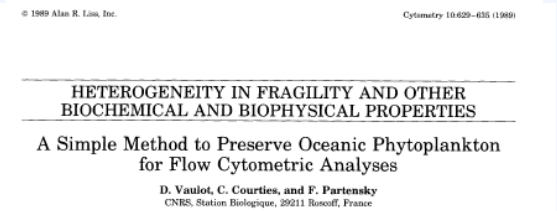
Bibliography (1983-2017)



COMMENT

Limnol. Oceanogr., 29(6), 1983, 1273–1280
© 1983, by the American Society of Limnology and Oceanography, Inc.

Flow cytometry and cell sorting: A technique for analysis and sorting of aquatic particles[1]



© 1989 Alan R. Liss, Inc. Cytometry 10:629–635 (1989)

HETEROGENEITY IN FRAGILITY AND OTHER BIOCHEMICAL AND BIOPHYSICAL PROPERTIES

A Simple Method to Preserve Oceanic Phytoplankton for Flow Cytometric Analyses

D. Vaulot, C. Courties, and F. Partensky
CNRS, Station Biologique, 29211 Roscoff, France

M. Thyssen et al. / Journal of Experimental Marine Biology and Ecology 416 (2011) 95–107

performed daily at noon in each mesocosm with a HANNA multi-parameter water quality meter (model H98828). These measurements showed that the water column was homogeneous during the whole experiment.

Samples for phytoplankton analysis using flow cytometry were collected every 6 h from 14:30 on August 20 to 14:30 on August 29 (sampling times were 2:30, 8:30, 14:30 and 20:30). Collecting data every 6 h is the minimal sampling frequency accepted in order to observe a 12:00 cell cycle (Nyquist, 1928), i.e. two cellular divisions per day, for any of the observed phytoplankton groups, which are commonly observed in natural environments (Binder and Durand, 2002; Jacquet et al., 2002; Thyssen et al., 2008). Samples for nutrient and chlorophyll a (chl a) analysis were collected once a day at 8:00.

2.2. Flow cytometry

Samples were collected using 1 dm³ dark containers and directly transferred into 12 cm³ vials for the Cytosense analyses, and 5 cm³ vials for the EPICS ALTRA flow cytometer analyses, both prefilled with glutaraldehyde (0.18% final concentration). The samples were immediately stored at −80 °C for less than a month. Flow cytometry analyses were conducted using two different types of instruments in order to achieve accurate estimations of cell counts from the smallest picophytoplankton to the largest microphytoplankton, and to collect cellular information using their light scattering properties. Forward light scatter (FWS) and sideward light scatter (SWS)) and their auto fluorescing properties (red fluorescence from chlorophyll (FLR) and orange fluorescence from phycoerythrin (FLO)). The picophytoplankton cells (Pico, diameter < 2 μm) and the smallest nanophytoplankton cells (Nano 1, ~2 μm) were analysed using an Epics Altra flow cytometer (Beckman Coulter) equipped with a 488 nm laser operated at 15 mW. Samples were thawed at room temperature and analysed immediately. Fluorescent beads (Fluoresbrite YG microspheres of 1.0 μm, Polysciences™) were systematically added to each sample, as an internal standard, in order to normalize the fluorescence emission and light scatter signal obtained from the Epics Altra flow cytometer. Abundance estimations were derived from the cell counts and the corresponding analysed volumes defined by the acquisition time and sample flow rate. The flow rate was obtained from weighing the vials before and after analysis and dividing the mass uptake by the sample density. Size was estimated by analysing bead suspensions of different bead sizes and determining the relationship between size and forward scatter (Venkatapathi et al., 2006). The FLR (675 ± 10 nm) and the FWS of the cells were recorded as the signal peak, thus giving little information on their shape, although the instrument is able to analyse the time of flight which gives an indication of their length. FLR and FWS peak values from the EPICS ALTRA are further defined as FLR_A and FWS_A.

Cells larger than ~2 μm were analysed using a Cytosense flow cytometer from Cytobuoy bv, equipped with a 488 nm laser operated at 15 mW. The pulse shape of FLR (668–734 nm), FLO (601–668 nm) and the FWS signals from the cells were recorded, allowing complex cells to be differentiated and chain forming cells to be accounted for. Integrated values of the Cytosense FLR and FWS signals are further defined as FLR_C and FWS_C. Abundances were directly estimated from the analysis of the samples through a stable peristaltic pump, routinely tested by using bead suspensions of known concentration. FluoSpheres polystyrene beads (Invitrogen), namely 2 μm red fluorescing and 10 μm orange fluorescing beads, were used as an internal standard to normalise scatter and fluorescence signals. Two protocols were used to optimise the abundance estimation of the small and large cells respectively. Cells < ~10 μm were analysed with a peristaltic pump speed of 1.08 mm³ s⁻¹ and a trigger level of 7 mV on FWS. Cells > ~10 μm were analysed using a peristaltic pump speed of

2.3. Chlorophyll a and nutrient analysis

Chlorophyll a (chl a) content was determined by High Performance Liquid Chromatography (HPLC). A volume of 400–600 cm³ was filtered onto a 25 mm Whatman GF/F filter. Filters were stored at −80 °C. Pigments were then extracted and analysed by HPLC after Zapata et al. (2000). Nitrate + nitrite ($NO_3^- + NO_2^-$), phosphate (PO_4^{3-}), and silicic acid ($Si(OH)_4$) concentrations were determined from 25 cm³ precombusted GF/F filtered seawater samples collected (60 cm³ triplicates) from each mesocosm at 8:00 am. An equal sample volume was discarded prior to storing the sample in plastic acid cleaned 100 cm³ bottles, kept frozen at −20 °C until analysis at EMER within 1 month, using a Bran Luebbe AutoAnalyzer 3 system based on the method by Grasshof et al. (1983).

2.4. Statistical analysis

Statistical analyses were run under R freeware (http://cran.r-project.org/). For each phytoplankton cluster, abundances, average $FWS_{A\ and\ C}$ and $FLR_{A\ and\ C}$ values per cell were calculated in order to identify differences between treatments during 3 different stages of phytoplankton development, a set of statistical analysis was run. For each defined phytoplankton stage, a normality test (Shapiro test) followed by a test of sphericity (Mauchly test) was run in order to define the best variance test. When data followed a normal distribution and sphericity was observed, a RM-ANOVA (repeated measures) was used. When normality was validated but not sphericity, or when normality was not validated, a Freidman rank test was run. Relative phytoplankton average abundances, relative average $FWS_{A\ and\ C}$ and relative average $FLR_{A\ and\ C}$ were calculated to show the differences between NTNUV (control) and the treated mesocosms (HTNUV, NTHUV and HTHUV) during the 3 different stages of the phytoplankton development, whilst considering the respective NTNUV value, as running post-hoc tests for each cluster and each phytoplankton stage, would have lead to complex interpretations. Significant differences were identified using a paired Wilcoxon signed-rank test. Periodic processes in the dynamics of abundance, average $FWS_{A\ and\ C}$ and $FLR_{A\ and\ C}$ values per cell were verified using computing periodograms with a Fast Fourier transformation smoothing the results with a series of modified Daniell smoothers (moving averages giving half weight to the end values, Daniell, 1946), generating spectral plots. These algorithms were computed on the average values between duplicates.

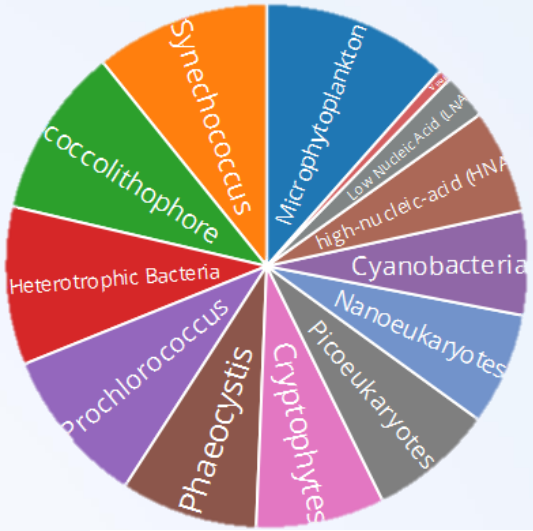
3. Results

3.1. UVR, temperature, salinity, chlorophyll a and nutrient concentrations

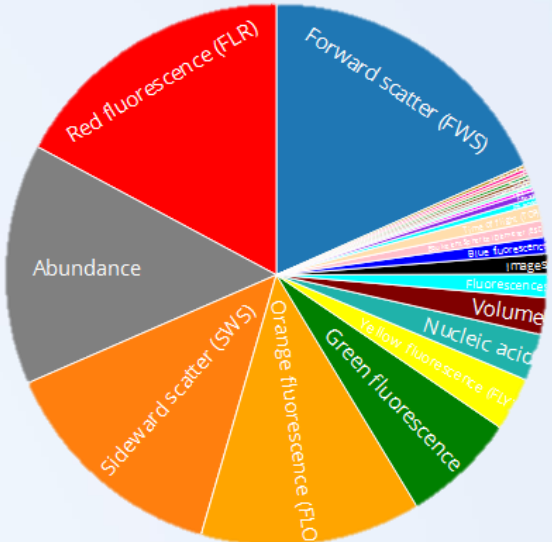
The photoactive depth (Z_{eb}, 10% of surface incident light) represents the depth at which UVR has significant biological effects (Neale et al., 2003). Z_{eb} reached depths between 2.7 and 5.7 cm and between 2.6 and 3.6 cm for radiation at 305 nm and 313 nm, respectively (Fig. 1A, B). Fig. 1C and D shows the 305 nm and 313 nm average irradiances in the water column from surface to Z_{eb}, calculated according to Macintyre and Cullen (1996). Average water column UVB irradiance increase in the HUV mesocosms were 77.8 ± 10.7% and 45.4 ± 16.8% for 305 and 313 nm, respectively (Fig. 1C, D), as compared to NUV treatments.

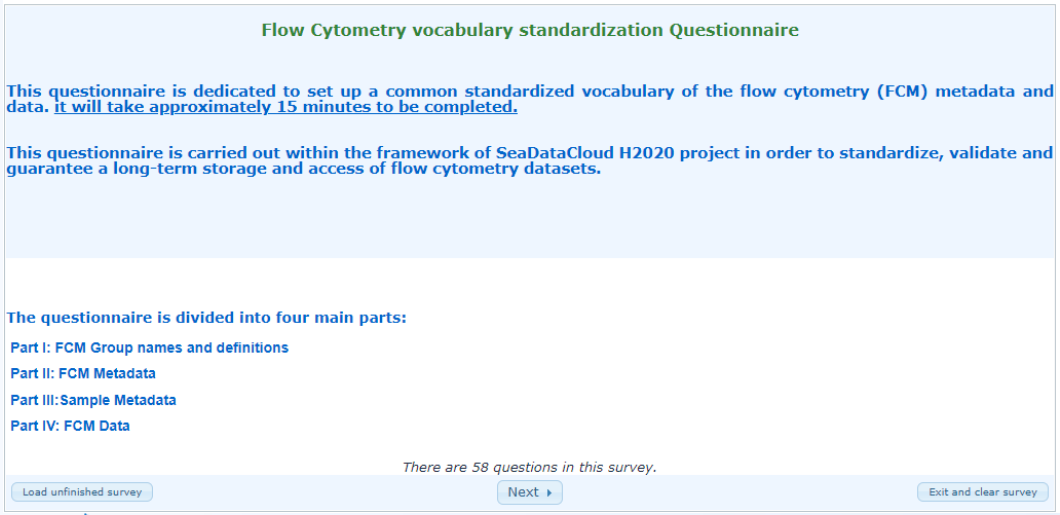
The initial temperature in all the mesocosms was ~13 °C and was increased by 2 °C from day 2 to day 4. At day 4, temperature stabilised at ~15 °C in the normal temperature and at ~18 °C in the high temperature treatment mesocosms on day 5 (Fig. 2A). Salinity values varied between 24.14 in HTNUV on day 6 and 25.19 in HTHUV on day 1 (data not shown). Chlorophyll a concentrations increased from day 1 up to day 5 in HTHUV and NTNUV, reaching maximal values of 8.90 ± 1.6 μg l⁻¹

Common functional groups



Parameters





Flow Cytometry vocabulary standardization Questionnaire

This questionnaire is dedicated to set up a common standardized vocabulary of the flow cytometry (FCM) metadata and data. it will take approximately 15 minutes to be completed.

This questionnaire is carried out within the framework of SeaDataCloud H2020 project in order to standardize, validate and guarantee a long-term storage and access of flow cytometry datasets.

The questionnaire is divided into four main parts:

Part I: FCM Group names and definitions

Part II: FCM Metadata

Part III:Sample Metadata

Part IV: FCM Data

There are 58 questions in this survey.

Load unfinished survey | Next ▸ | Exit and clear survey

180 FCM users all around the world

SeaDataCloud

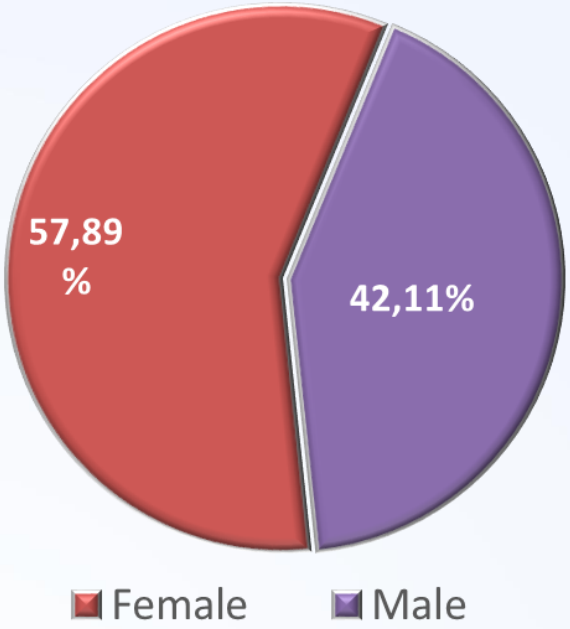
➔ 38 answers (2 months)

Completed answers (79%)

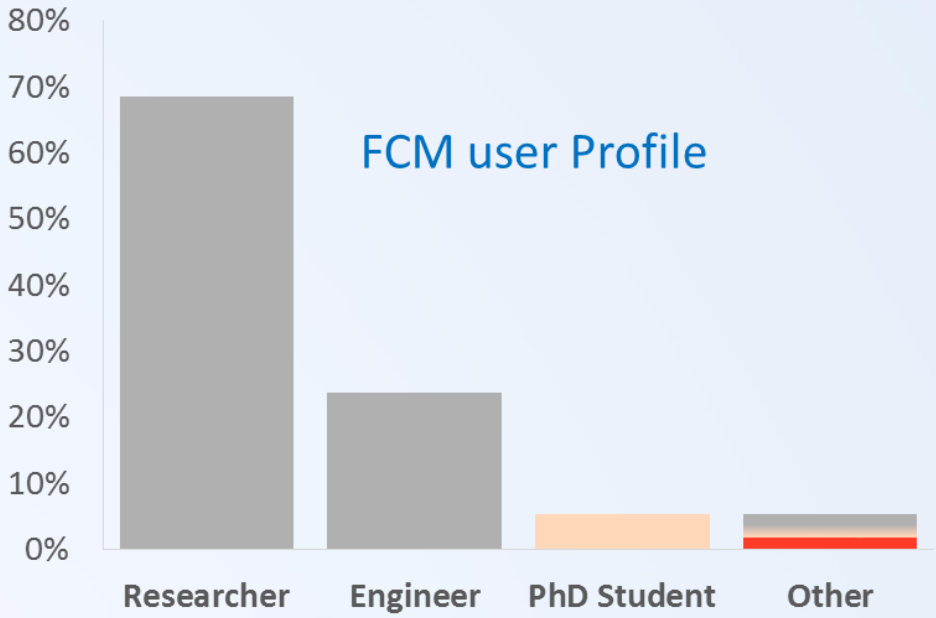
Uncompleted answers (21%)

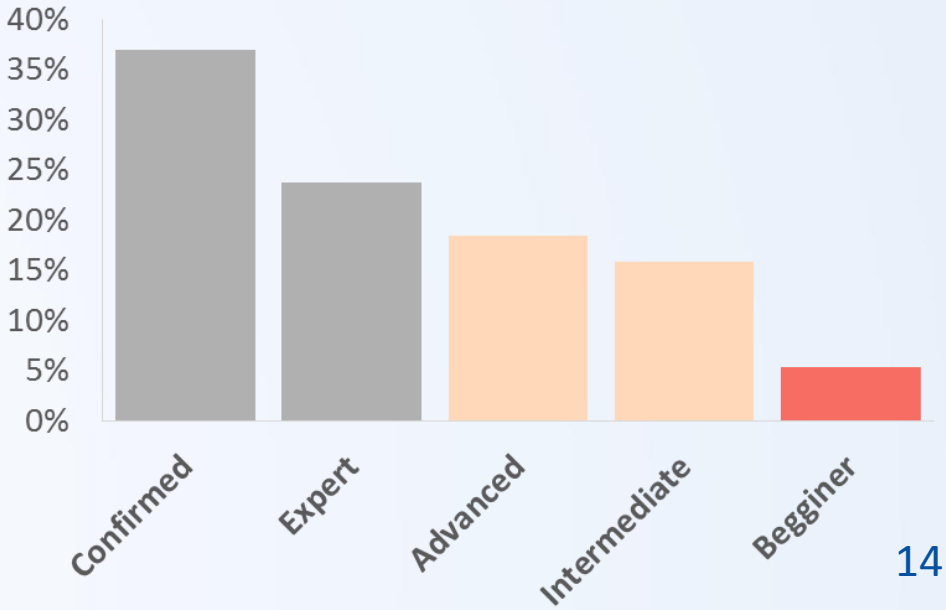
SeaDataCloud

Gender participation



FCM user Profile





Semantic model (BODC)

Chemical model	Biological model	Physical model
Measurement Substance Measurement Matrix Relationship Matrix Matrix Subcomponent	Measurement Organism Name Organism Specifics Measurement Matrix Relationship Matrix Matrix Subcomponent Method	Measurement Statistical Measurement Matrix Relationship Matrix Method
Concentration of carbon (total inorganic) {TCO2} per unit mass of the water body [dissolved plus reactive particulate phase]	**Abundance of Bacteria (ITIS: 202421: WoRMS 6) [Subgroup: heterotrophic] per unit volume of the water body by automated flow cytometry**	**Forward light scatter pulse shape area average per cluster in the water body by automated flow cytometry**

Forward scatter pulse shape area

Average

per cluster in the

Water body

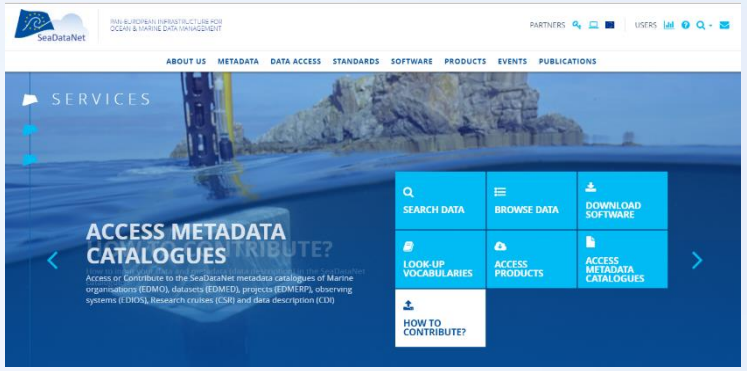
automated flow cytometry

The cluster name is managed in a separate vocabulary list (F02)

Common vocabulary results ➔ 44 codes

BODC WEBSERVICES V2 (LIBRARIES) CL12

Library	Thesaurus	Title	Alt Title	Version	Members	Modified
C16		SeaDataNet sea areas	SDN sea areas	9	127	11/7/2012 2:00:06 AM
C17		ICES Platform Codes	ICES Platforms	712	5607	3/20/2018 2:00:05 AM
C19		SeaVoX salt and fresh water body gazetteer	SeaVoX water bodies	17	263	2/21/2018 2:00:03 AM
C32		International Standards Organisation countries	ISO countries	7	251	1/14/2016 2:00:02 AM
C34		Activity purpose categories	Purpose categories	4	22	8/27/2011 3:00:05 AM
C35		European Nature Information System	EUNIS3 Habitats	1	56	2/19/2010 2:01:37 AM



F02	SeaDataCloud Flow Cytometry Standardised Cluster Names	SDC flow cytometry cluster names	2	11	2/3/2018 2:00:02 AM
P01	BODC Parameter Usage Vocabulary	BODC PUV	800	37732	3/14/2018 2:00:03 AM
P02	SeaDataNet Parameter Discovery Vocabulary	SeaDataNet PDV	107	435	2/13/2018 2:00:03 AM
L22	SeaVoX Device Catalogue	SeaVoX Device Catalogue	324	1280	3/6/2018 2:00:04 AM
P06	BODC data storage units	BODC units	99	346	2/16/2018 2:00:02 AM

2

Data Transport Format for FCM Data

SeaDataNet

PAN-EUROPEAN INFRASTRUCTURE FOR OCEAN & MARINE DATA MANAGEMENT

https://www.seadatanet.org/Standards/Data-Transport-Formats

ABOUT US METADATA DATA ACCESS STANDARDS SOFTWARE PRODUCTS EVENTS PUBLICATIONS

The SeaDataNet NetCDF (CF) format for profiles, time series and trajectories can be used next to the SeaDataNet ODV 4 ASCII format in the services of the SeaDataNet infrastructure.

Additional feature types have been defined for the storage of multiple trajectories data like moored ADCP (Feature type = timeseriesProfile) or shipborn ADCP (Feature type = trajectory profile).

SPECIFIC DOCUMENTATION FOR **BIOLOGICAL DATA**

DESCRIPTION OF THE SEADATANET DATA TRANSPORT FORMAT FOR BIOLOGICAL DATA

DOWNLOAD (991.96 KB)

TEMPLATE AND EXAMPLES OF BIOLOGICAL DATA FILES (ZIP FILE)

DOWNLOAD (550.84 KB)

SPECIFIC DOCUMENTATION FOR **MICROLITTER DATA**

Format and examples of of ODV data and CDI xml metadata files, they have been prepared in the frame of the EMODnet chemistry project

SPECIFIC DOCUMENTATION FOR **FLOWCYTOMETRY DATA**

Format and examples of ODV data and CDI xml metadata files

Data Transport Format for FCM Data

SeaDataNet

PAN-EUROPEAN INFRASTRUCTURE FOR OCEAN & MARINE DATA MANAGEMENT

https://www.seadatanet.org/Standards/Data-Transport-Formats

ABOUT US METADATA DATA ACCESS STANDARDS SOFTWARE PRODUCTS EVENTS PUBLICATIONS

TemplateandExampleODV_CDI_FlowCytoMetry - Excel (Échec de l'activation du produit)

FICHIER ACCUEIL INSERTION MISE EN PAGE FORMULES DONNÉES RÉVISION AFFICHAGE

C4

	A	B	C	D
1		This file contains : the **general description of an ODV Flow Cytometery (FCM) File** for SeaDataNet and the **detailed description of ODV file Fields** that can be used as well as **an example of FCM dataset** with the CDI metadata (Orange sheets) and the corresponding data in the ODV FCM format (Blue sheets).		
2				
3				
4		Every ODV sheet contains a set of FCM data and has a **HEADER** and a **DATA TABLE.** The DATA TABLE contains the data It exists in 2 type of fields: 6 **fixed** fields and a various number of **additional** fields. The HEADER describes the fixed and additional fields For each of the additional Fields it is required to describe: -**Subject:** This header contains the name of the column header of the additional field, and the prefix 'SDN:LOCAL:' indicating it is a local text string. -**Object**: This header contains the code from the 'SeaDataNet P01 vocabulary', corresponding to the local subject (codes available at http://vocab.nerc.ac.uk/collection/P01/current/). -**Units:** This header describes the units corresponding to the object, according to SeaDataNet P06 vocabulary (available at http://vocab.nerc.ac.uk/collection/P06/current/). -**Instruments:** This header should be entered optionally for the relevant datafields. It corresponds with SeaDataNet L22 vocabulary (available at http://vocab.nerc.ac.uk/collection/L22/current/). OBJECTIVE: This format enables NODC's to make FCM data accessible using SeaDataNet infrastructure and makes it possible for NODC's to use SeaDataNet to exchange FCM data		
5				

How to use | General ODV template 20180302 | ODV Fields_20180302 | FCM_ODV_exple_20180302 | FCM_CDI_fields_20180302 | FCM_CDI_fexple_20180302

sdn-userdesk@seadata

Ingestion into SDN

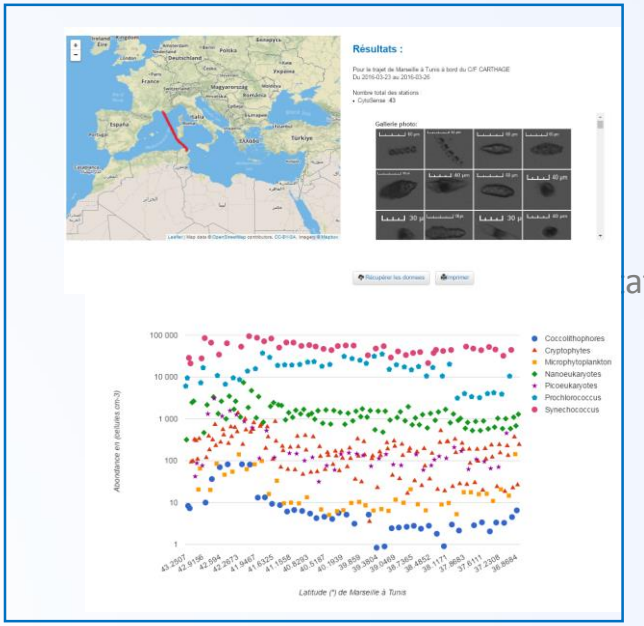
SeaDataCloud

Local FCM data management

CytoClus

Data files

Acquisition | Analysis | Consolidation | Expert QC



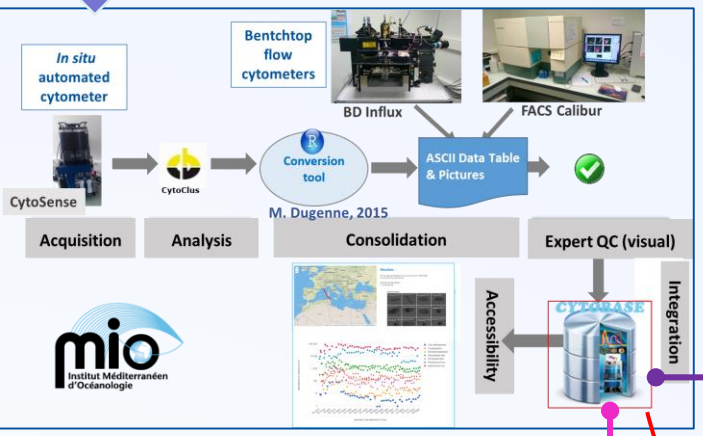
ation

Accessibility

Integration

Ingestion into SDN

SeaDataCloud



Metadata generation

MIKADO

Coupling table

Download Manager

SeaDataNet

SEADATANET PORTAL

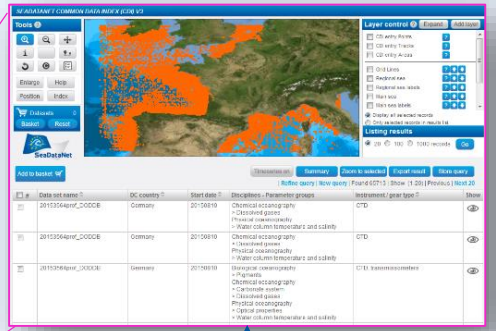
CSR

EDIOS

EDMED

EDMERP

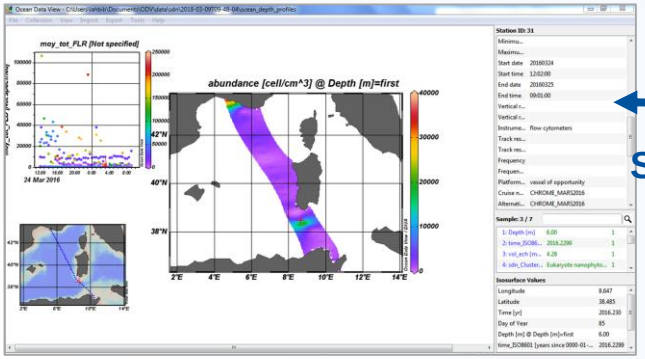
CDI



Ocean Data View

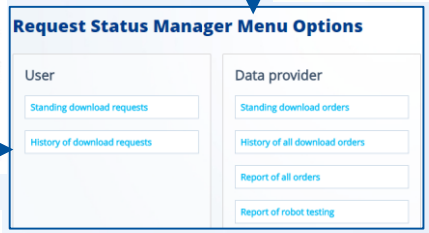
https://odv.awi.de

© 2017 Reiner Schlitzer



IMPORT

SDN Format



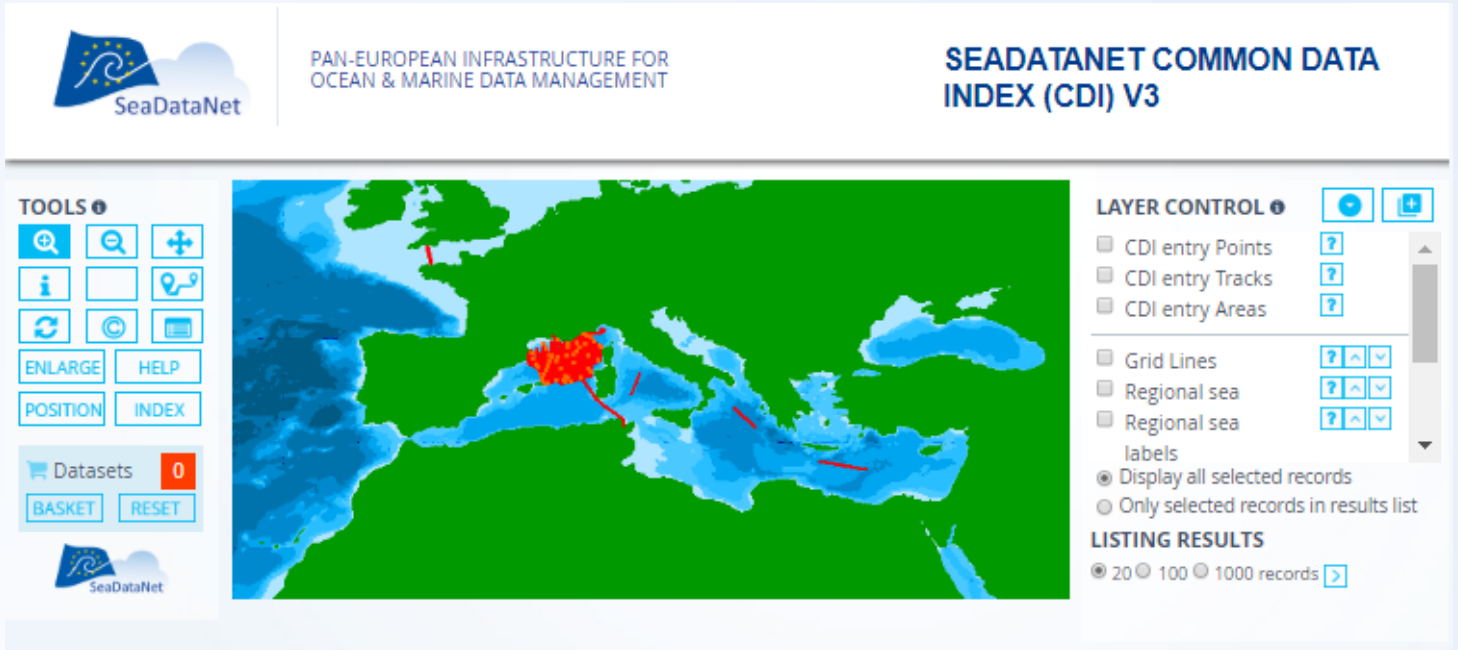
OCEAN DATA VIEW ASCII FORMAT

SeaDataCloud

SeaDataNet

PAN-EUROPEAN INFRASTRUCTURE FOR OCEAN & MARINE DATA MANAGEMENT

SEADATANET COMMON DATA INDEX (CDI) V3



ADD TO BASKET | TIMESERIES ON | SUMMARY | ZOOM TO SELECTED | EXPORT RESULT | STORE QUERY

| Refine query | New query | Found 67 | Show (1-20) | Previous | Next 20

#	Data set name	DC country	Start date	Disciplines - Topics	Instrument / gear type	Show
	BERRE MISE 2014_FCMW	France	20140626	Biological oceanography > Other biological measurements	flow cytometers	
	BioArgoMed_FCMW	France	20150706	Biological oceanography > Other biological measurements	flow cytometers	
	CEL2SAT_FCMW	France	20130521	Biological oceanography > Other biological measurements	flow cytometers	
	CHROME_MARS2016_FCMW	France	20160324	Biological oceanography > Other biological measurements	flow cytometers	
	DEWEX LEG1_FCMW	France	20130203	Biological oceanography > Other biological measurements	flow cytometers	

Conclusion

- FCM data are ingested into SeaDataNet infrastructure
- Whatever the instrument used ➔Common Vocabulary (CV)
- Decide on a group of experts interested in contributing to the vocabulary work and decide on a co-ordinator
- Update is possible/The BODC Vocabulary Editor webpage: https://www.bodc.ac.uk/resources/vocabularies/vocabulary_editor/
- BODC is setting up some repositories on GitHub for each individual collection and F02 will have its own too. So this could be used to share and discuss issues more widely.

Thank you for your attention