

IMDIS 2016

International Conference on Marine Data and Information Systems



CHROME

High resolution and automated flow cytometry data management

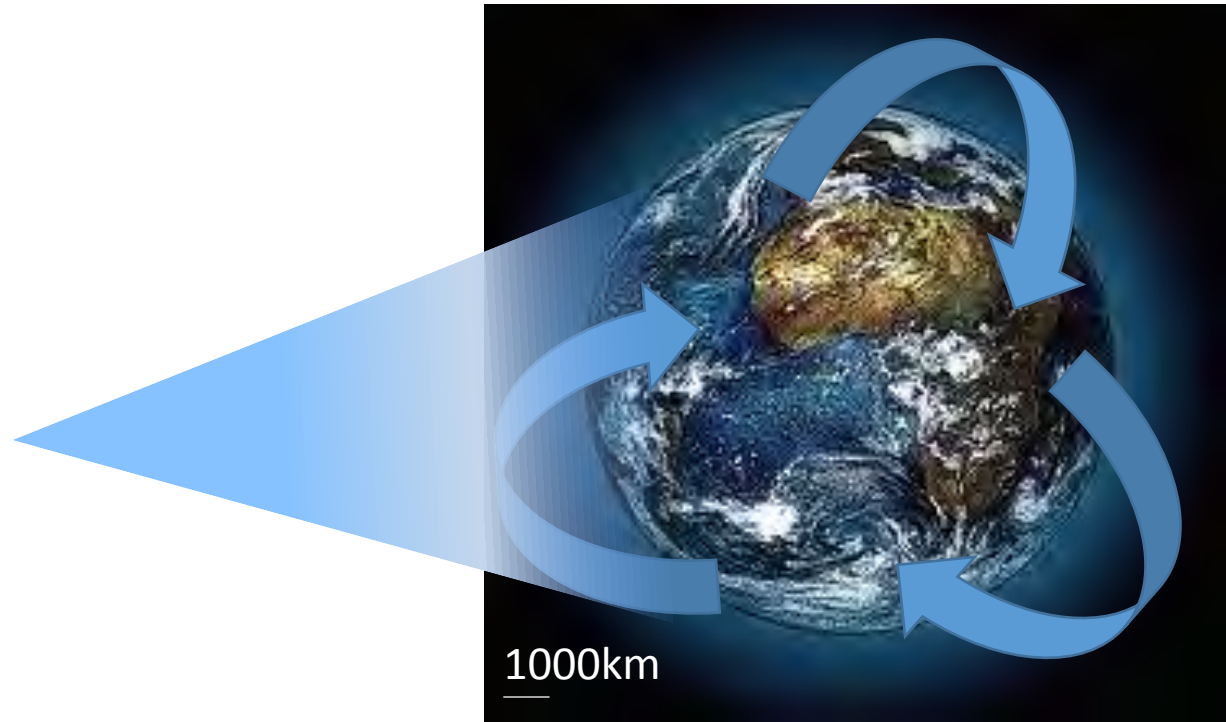
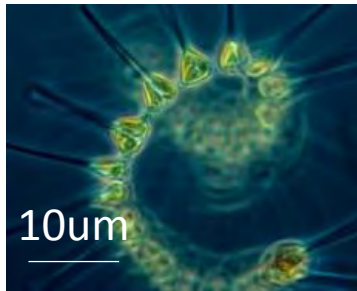
Gdansk (Poland) - October 11-13, 2016

Soumaya Lahbib (Data manager MIO), M. Dugenne (PhD candidate MIO),
M. Libes (Computer Engineer OSU), C. Sammari (Professor INSTM), M. Belhassen (Scientist INSTM),
G. Grégori (Research scientist MIO), Melilotus Thyssen (Research scientist MIO)



Phytoplankton

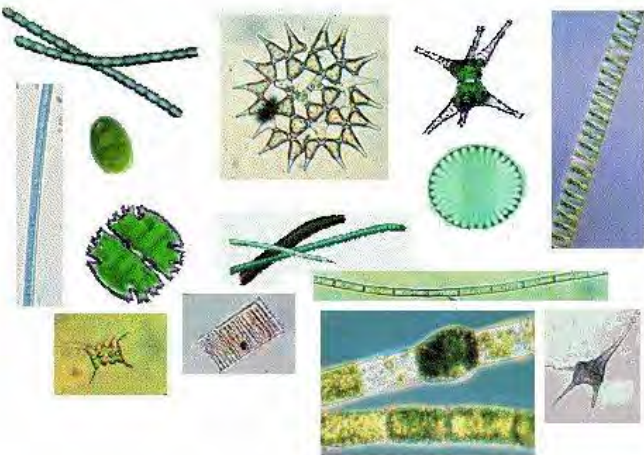
Thousands of species $< 1000 \mu\text{m}$ catalyze the most important geochemical processes for sustaining life on earth AND at a minute scale.



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

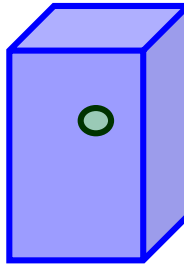
Phytoplankton observation is complex

Morphology and size

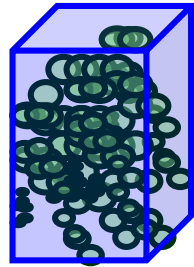


Abundances

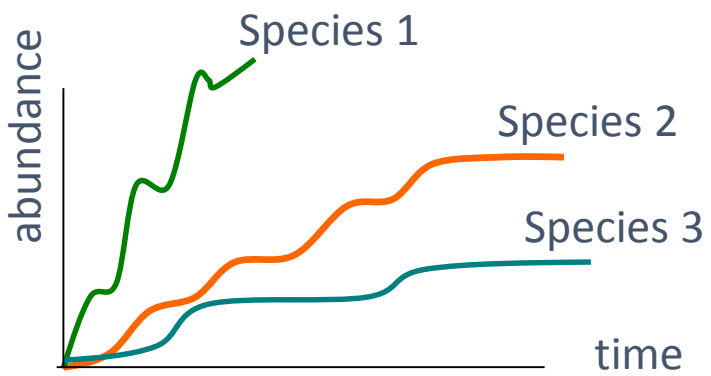
1 cell. cm⁻³



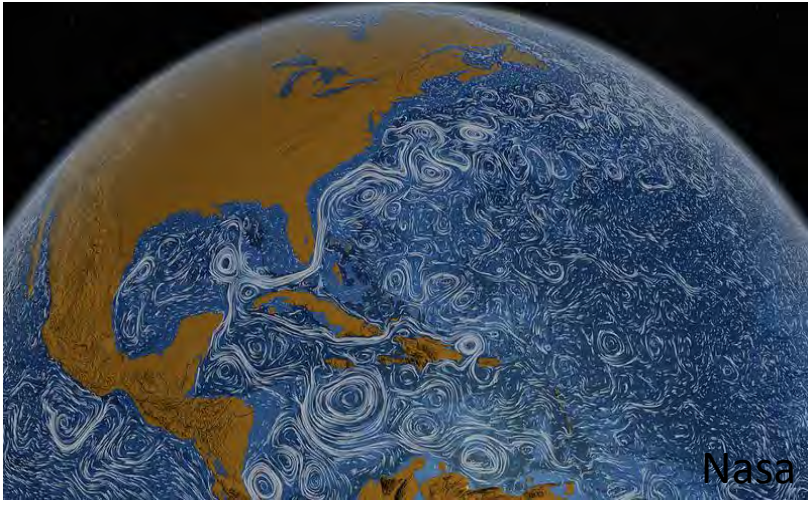
10⁶ cells. cm⁻³



Growth rates



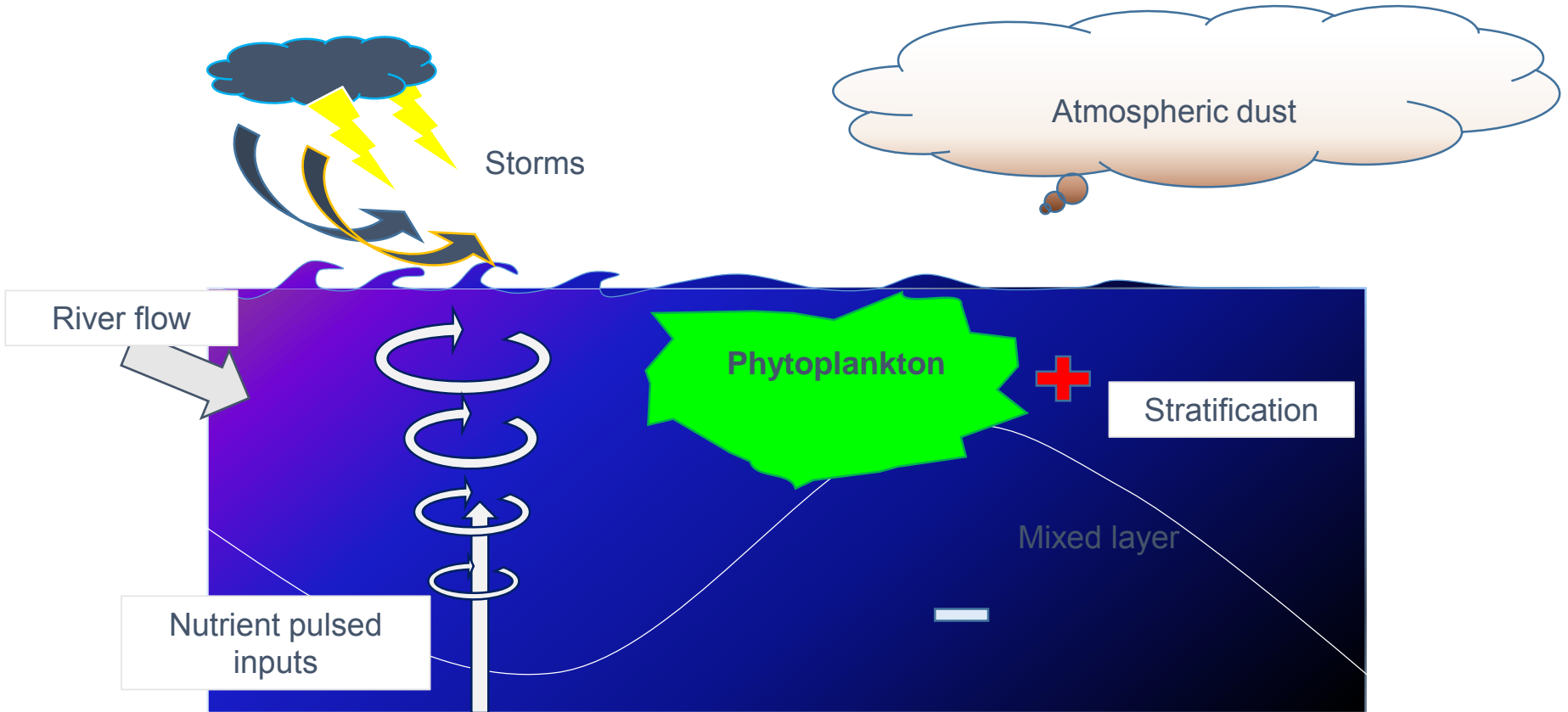
Turbulence



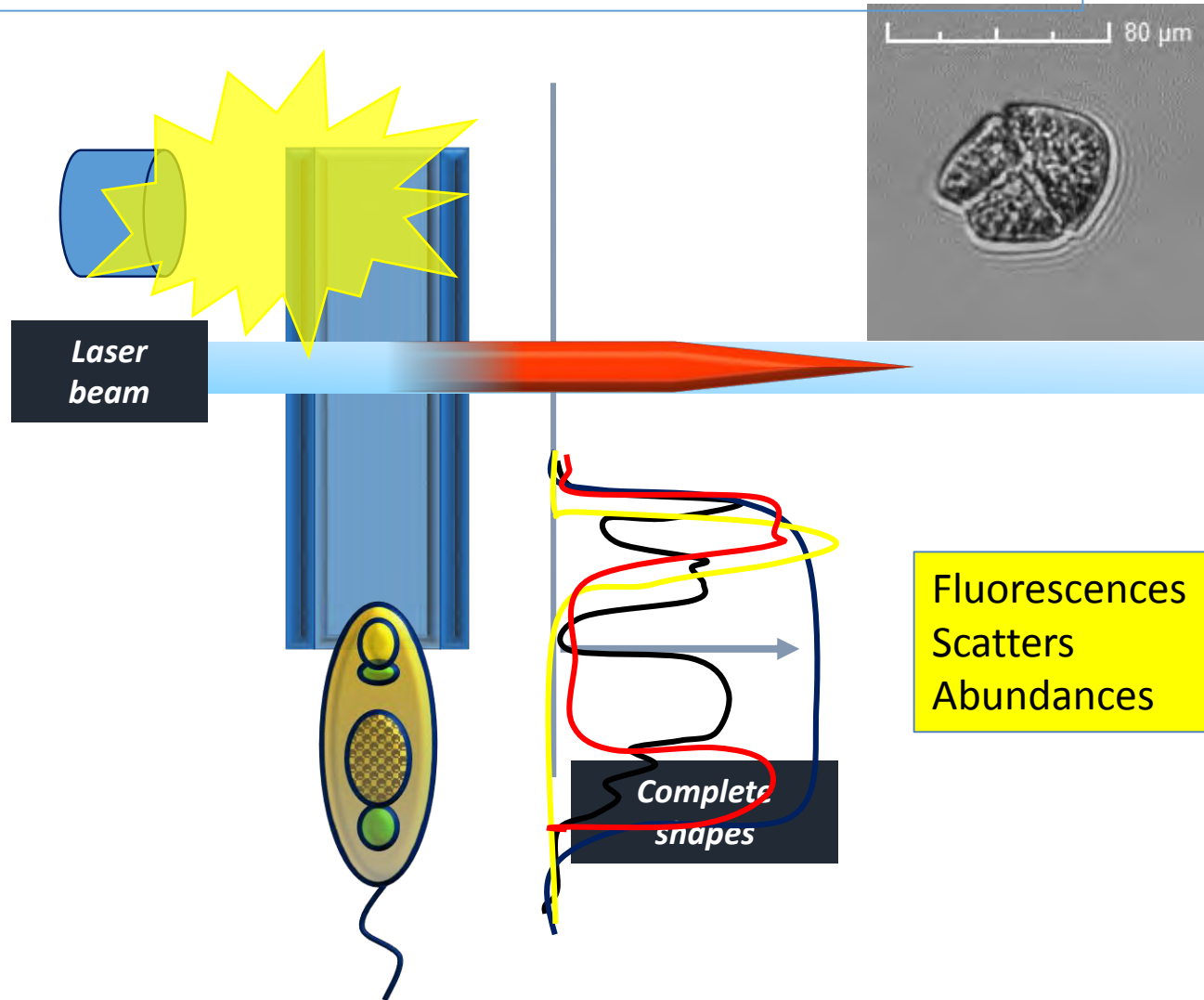
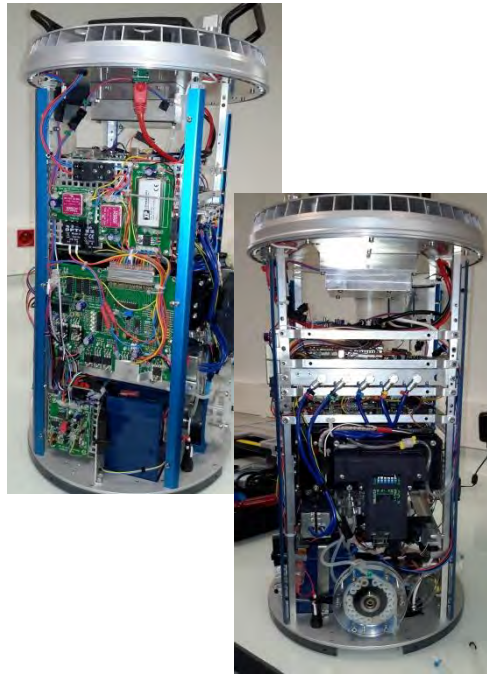
Nasa

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

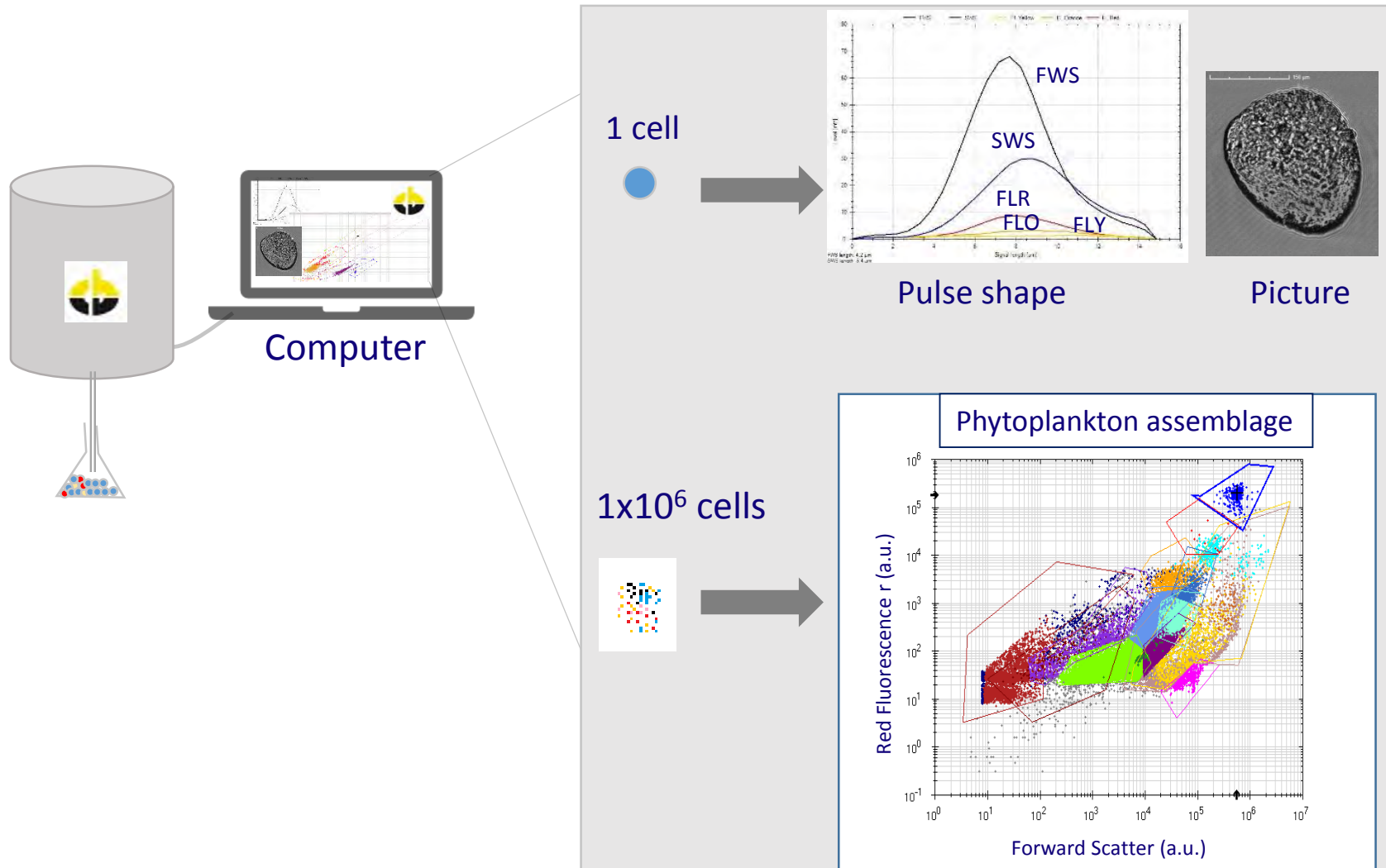
Short term variation and sporadic events impacts are nearly unknown



New technology for the resolution of phytoplankton functional diversity at hourly and regional scales



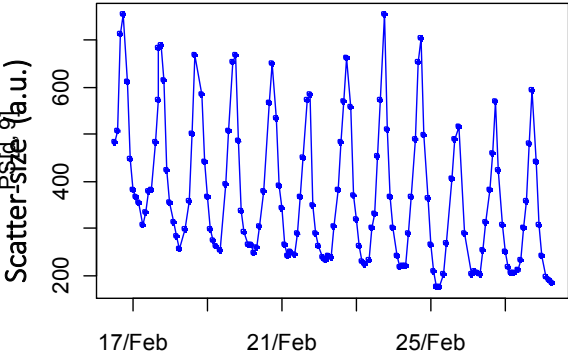
Phytoplankton functional groups resolution



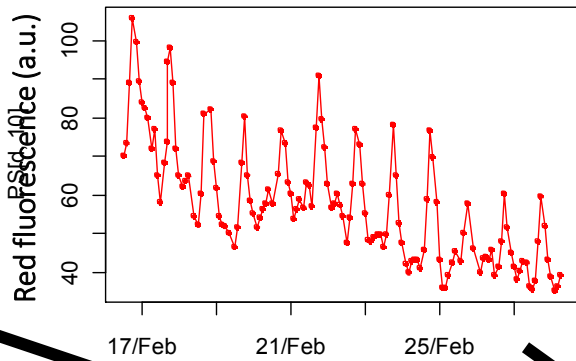
- Phytoplankton functional groups/Phytoplankton abundance per group
- Fluorescences/scatter per cell/Size estimation after calibration of scatter
- Phytoplankton images (taxonomical identification >20 µm)

Additional information extracted from the single cell approach:

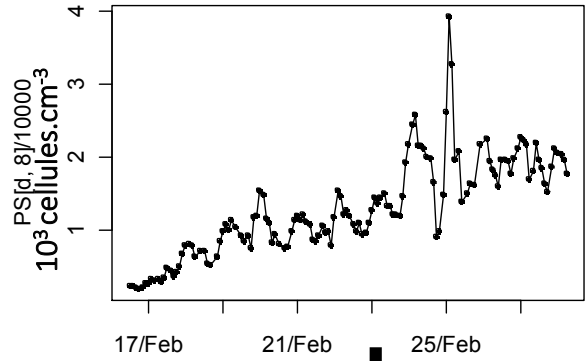
Scatter



Fluorescence

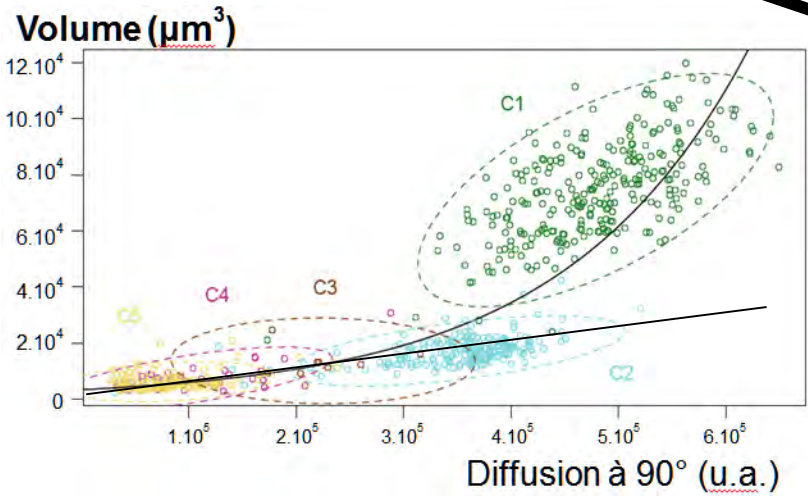
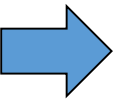
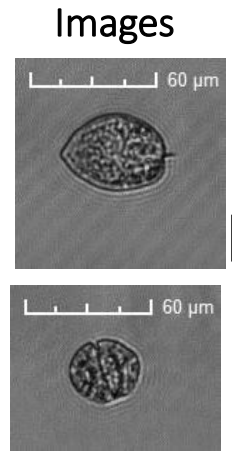


Abundances



format = c(dates = "D/MM", times = "H:M"), origin = c(month = 1, day = 1, year = 2015)

Biovolume



size-structured matrix population model:
In situ growth rate per phytoplankton cluster

André *et al.*, 1999; Sosik *et al.*, 2003;
Thyssen *et al.*, 2009, 2010,
Dugenne *et al.*, 2014, 2015.

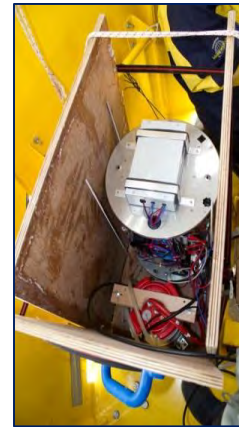
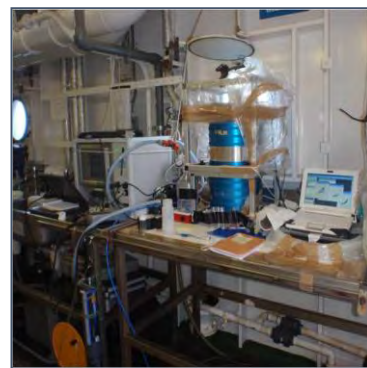
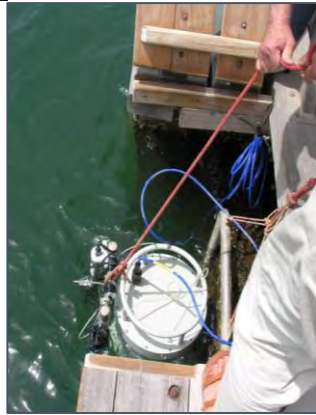
Several scientific experiences were conducted with a relative autonomie of 6 months

Scientific vessels

Coastal platforms

Ships of opportunity

Buoys



Malkassian et al. 2011

Dugenne et al. 2014

Thyssen et al. 2008, 2009a.b, 2012, 2014, 2015





CNRS UPMC INSU
Station Biologique
Roscoff

Institut de
Mathématiques
de Marseille,
UMR 7373



INGV

CHROME

Continuous High Resolution Observation of the
Mediterranean Sea:

<https://chrome.mio.univ-amu.fr/>

*Understanding of the ecological and
biogeochemical functioning in relation to
meso-scale dynamics at the Mediterranean
sub-basin scale and weekly scale.*



A*MIDEX CHROME Project

Data Acquisition = one analysis every 30 min.

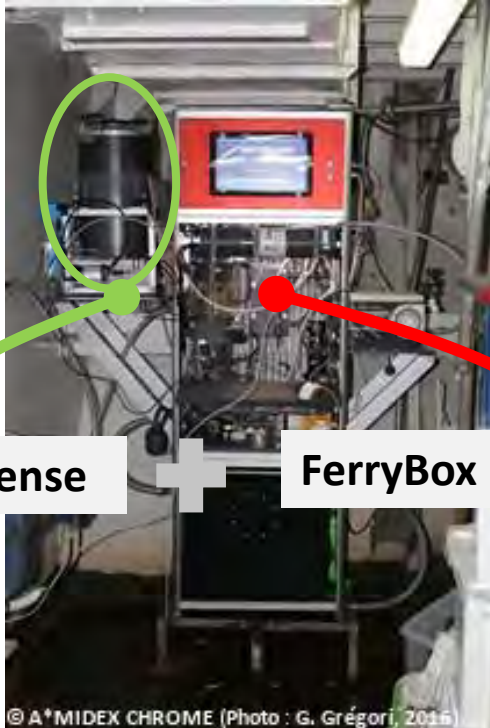


C/F CARTHAGE trajectories

./30min

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

Pont 1 – C/F CARTHAGE



./1min

- Temperature
- Salinity
- Fluorescence
- Turbidity
- pH
- pCO₂
- Oxygen



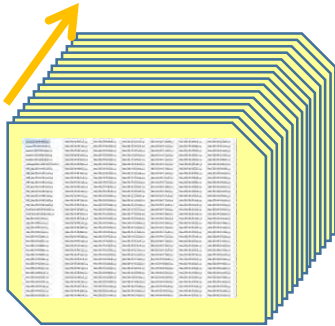
© A*MIDEX CHROME (Photo : G. Grégori, 2016)

Data acquisition & analysis



Output

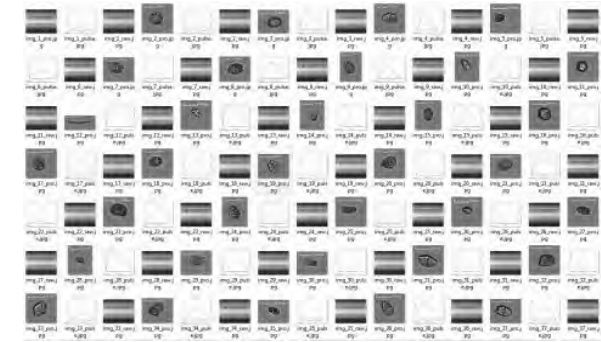
Manual clustering



Backup



Measurements by CytoSense



Seperate statistical CSV files:
Average values of optical properties and Counts
+ Pictures



- 1. Huge quantity of data
- 2. Data memory size consuming
- 3. Not a dedicated FCM database



THE CYTOBASE DATABASE

FCM Data management Workflow

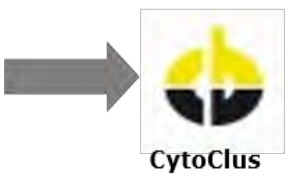


Table Données
Table photo
Collection photos

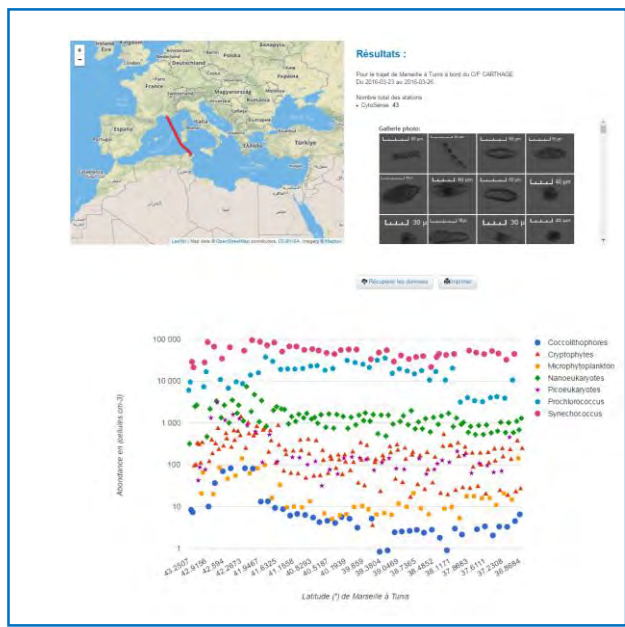


Acquisition

Analysis

Consolidation

Expert QC



Accessibility



Integration

Data consolidation

Cytobase Input Processor (M. Dugenne, 2015)

Cytobase data

Parcourir... 11 fichiers sélectionnés.
Upload complete

Warning:

Upload successful



Create Inputs folder

Metadata

Project and samples context Raw data Size conversion

Project	Project date	PI	Cytometer ID
<input type="text" value="Enter project name"/>	<input type="text" value="2015-08-18"/>	<input type="text" value="Enter PI name"/>	<input type="text" value="Enter cytometer ID"/>
Station	Depth	Latitude	Longitude
<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
Filename	Filename model		
<input type="text" value="Filename"/>	<input type="text" value="BERRE_082013_3F_FLR9 2013-12-17 13u"/>		
Samples operator	Standards reference	Clustering method	Observation type
<input type="text" value="Enter name of operator"/>	<input type="text" value="Enter standards beads ref"/>	<input type="text" value="Automated"/>	<input type="text" value="In situ"/>

Data

2013-12-17T14:17:00Z	2013-12-17T14:17:00Z	16,056.60	Synechococcus	BERRE_082013_12S_FLR9 2013-12-17 14u17.cyz	1.99	FL Red	10
2013-12-17T14:17:00Z	2013-12-17T14:17:00Z	16,056.60	Cryptophytes	BERRE_082013_12S_FLR9 2013-12-17 14u17.cyz	1.99	FL Red	10
2013-12-17T14:43:00Z	2013-12-17T14:43:00Z	16,056.61	Beads 2 mu	BERRE_082013_16F_FLR9 2013-12-17 14u43.cyz	2.04	FL Red	10
2013-12-17T14:43:00Z	2013-12-17T14:43:00Z	16,056.61	Microphytoplankton	BERRE_082013_16F_FLR9 2013-12-17 14u43.cyz	2.04	FL Red	10
2013-12-17T14:43:00Z	2013-12-17T14:43:00Z	16,056.61	Picoeukaryotes 2	BERRE_082013_16F_FLR9 2013-12-17 14u43.cyz	2.04	FL Red	10
2013-12-17T14:43:00Z	2013-12-17T14:43:00Z	16,056.61	Picoeukaryotes 1	BERRE_082013_16F_FLR9 2013-12-17 14u43.cyz	2.04	FL Red	10

Please associate each selection set to trigger, PMT's amplification and standardized phytoplankton category

NB: All incompatible entries will be removed

Expert name

Cluster:

Trigger

Channel/Level:

PMT's amplification

SWS:

Standardized name

Cluster:

Associate

Data consolidation

Picture selection

Project and samples context Raw data Size conversion Image-In-Flow pictures Stations explorer

Add samples pictures

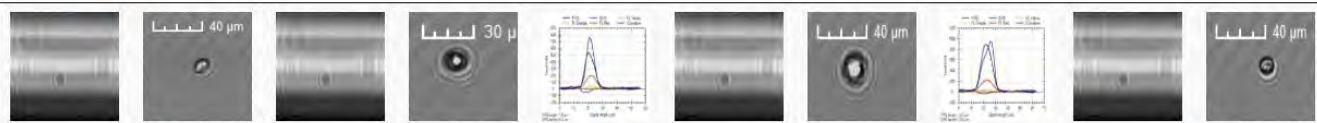
Select file: DEWEXL2FLR10 2013-04-05 15u04.cyz

Add pictures: Choisir les fichiers 15 fichiers

Upload complete

Check all

Show 4 entries Search:



Project and samples context Raw data Size conversion Image-In-Flow pictures Stations explorer


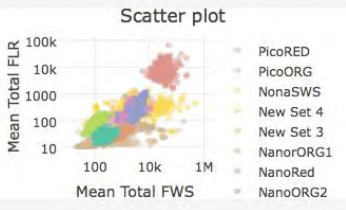
Stations Explorer

Select station: S1

X axis: Mean Total FWS

Y axis: Mean Total FLR

Scatter plot



Download Table

Data Table
Picture Table



CYTOBASE modeling

- User requirements acquisition
- Management rules clarification
- First modeling versions
- Validation and/or modifications

The screenshot shows the JMerise software interface. The main window displays a database model for 'MCD : MCD_V8'. It features several entities and their relationships:

- projet**: Attributes include `pk_id_proj` (Auto_increment), `nom_proj`, `pi_proj`, `date_deb_proj`, and `date_fin_proj`.
- mesures_echantillon**: Attributes include `pk_id_ech` (Auto_increment), `ref_ech`, `nom_ech`, `meth_sous_ech`, `date_analyse`, `depth`, `pres`, `ref_cont_beats`, `freq_channels`, `freq_level`, `time_acquisition`, `vol_ech`, `abundance`, `label_count`, `debit`, `laser`, `expert_nomination`, `nom_groupe`, `moy_soi_FLR`, `sd_soi_FLR`, `moy_soi_FLO`, `sd_soi_FLO`, `moy_soi_FWS`, `sd_soi_FWS`, `moy_soi_SWS`, `sd_soi_SWS`, `PMT_FLR`, `PMT_SWS`, `PMT_FLO`, `median_FWS`, `median_SWS`, and `median_FLO`.
- photos**: Attributes include `pk_id_photo` (Auto_increment), `file_cyz_photo`, and `lien_pic`.
- methode**: Attributes include `pk_id_meth` (Auto_increment), `file`, `expert_nomination`, `coef_a`, `coef_b`, `var_x1`, `var_x2`, `moy_ESD`, and `sd_ESD`.

Relationships are shown as follows:

- projet** (1,n) **a pour projet** (1,1) **mesures_echantillon**
- projet** (1,n) **mesurer par** (1,1) **methode**
- mesures_echantillon** (0,n) **appartient** (1,1) **photos**
- mesures_echantillon** (0,1) **a pour methode** (1,1) **methode**

A text box in the top left of the main window reads: 'Modélisation des données de cytométrie en flux Date: 19 juillet 2016 Version 8 outil utilisé: JMERISE'.

JMERISE
JMerise: logiciel de modélisation des MCD pour Merise
Merise

→ Why Jmerise:
- User-friendly DB modeling programme: based on entity-relation methods
- Generate Logical data model and Mysql script

CYTOBASE modeling

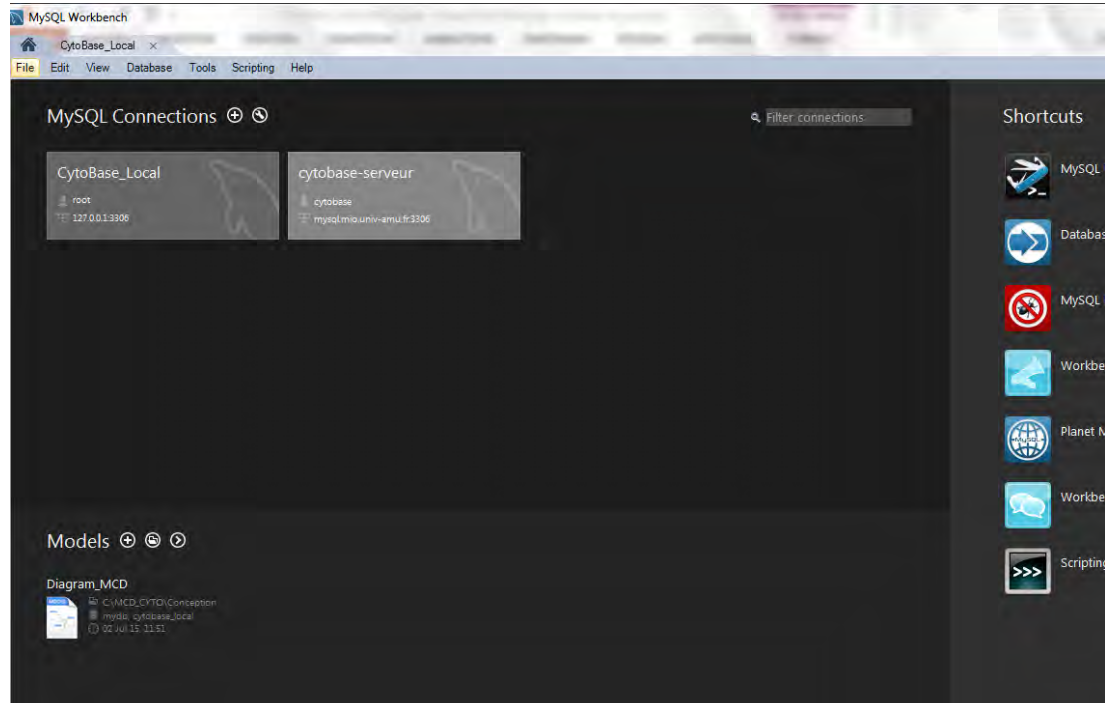


phpMyAdmin interface showing the database structure for 'cytobase' on localhost. The database contains 7 tables:

Table	Action	Lignes	Type	Interclassement	Taille	Perte
embarcation	Afficher Structure Rechercher Insérer Vider Supprimer	8	InnoDB	latin1_swedish_ci	16 Kio	-
materiel	Afficher Structure Rechercher Insérer Vider Supprimer	8	InnoDB	latin1_swedish_ci	16 Kio	-
mesures_echantillon	Afficher Structure Rechercher Insérer Vider Supprimer	17 900	InnoDB	latin1_swedish_ci	4,5 Mio	-
methode	Afficher Structure Rechercher Insérer Vider Supprimer	17 900	InnoDB	latin1_swedish_ci	2,5 Mio	-
photos	Afficher Structure Rechercher Insérer Vider Supprimer	8 801	InnoDB	latin1_swedish_ci	1,6 Mio	-
projet	Afficher Structure Rechercher Insérer Vider Supprimer	8	InnoDB	latin1_swedish_ci	16 Kio	-
station	Afficher Structure Rechercher Insérer Vider Supprimer	1 781	InnoDB	latin1_swedish_ci	240 Kio	-
7 tables	Somme	46 406	InnoDB	latin1_swedish_ci	8,9 Mio	0 0



MySQL Workbench



MySQL Workbench interface showing connections and a diagram. The connections list includes 'CytoBase_Local' and 'cytobase-serveur'. The diagram shows a model named 'Diagram_MCD'.

- Easy communication with DB
- Management and administration of DB
- Editing and quick querying

CYTOBASE data integration



ETL



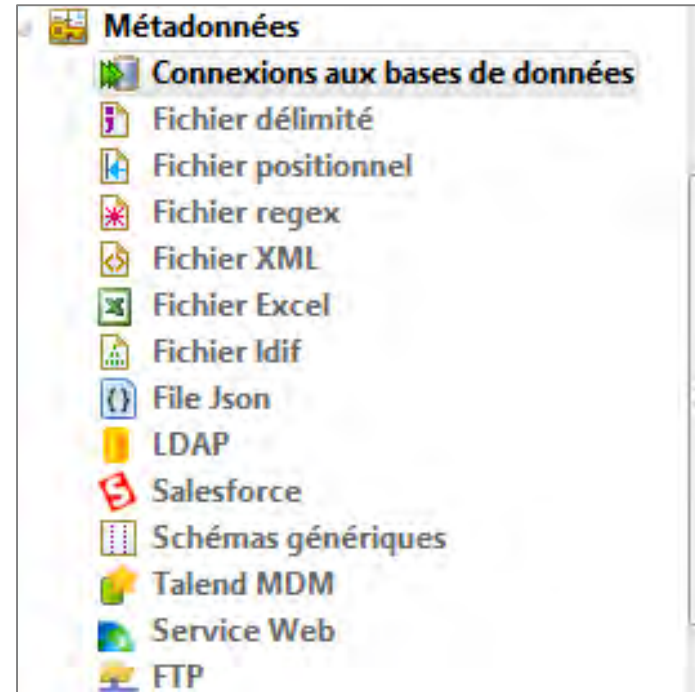
The screenshot displays the Talend Open Studio interface for a job named 'Job mesures_echantillon_v6 0.1'. The main workspace shows a data flow diagram with several components: 'Data_DEWEG_LEG2', 'tMap_1', 'mesures_v6 (Main)', 'station', 'materiel', and 'methode'. Data flow statistics are visible on the connections, such as '2687 rows in 0.31s' and '7833.82 rows/s (Lookup)'. The left sidebar shows a project tree with folders for 'Code', 'Modèles SQL', and 'Métadonnées', including a 'cytobase_v6 0.1' folder with sub-folders for 'Requêtes', 'Schémas des tables', and 'Schémas des vues'. The bottom panel shows the 'Job mesures_echantillon_v6' configuration and execution logs. The logs indicate the job started at 16:23 on 17/02/2016, connected to a socket on port 3853, and finished at 16:23 on the same date. The log text is as follows:

```
Job mesures_echantillon_v6
Démarage du Job mesures_echantillon_v6 à 16:23 17/02/2016.
[statistics] connecting to socket on port 3853
[statistics] connected
[statistics] disconnected
Job mesures_echantillon_v6 terminé à 16:23 17/02/2016. [Code
sortie=0]
```

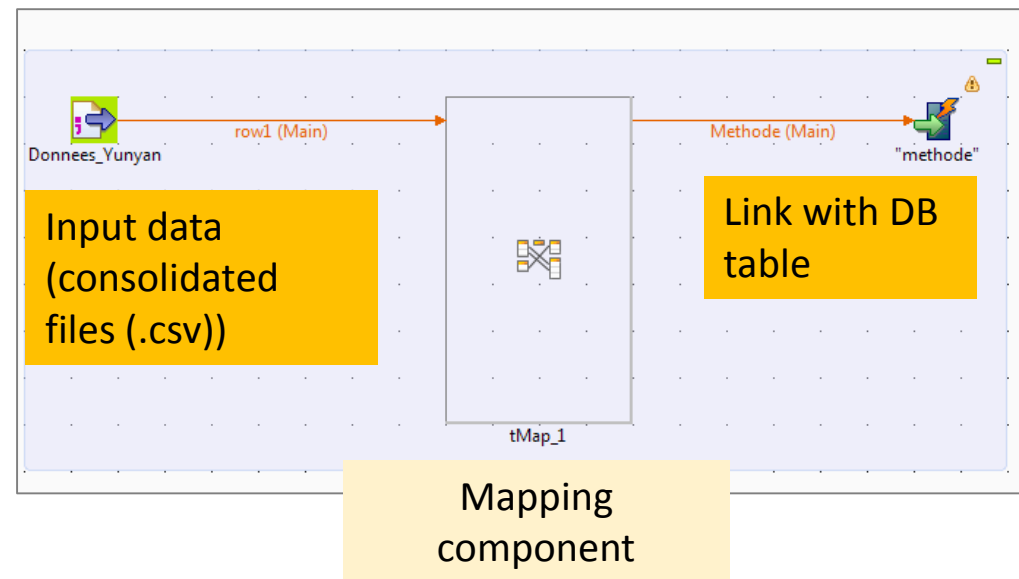
→ ETL (Extract, Transform, Load) processing development platform



→ Connectivity with DB and multiple file types



→ Data integration optimized processing time



Input data

DB table

Talend Open Studio for Data Integration - tMap - tMap_1

Find :

Var

Expression	Type	Variable
TalendDate.parseDate("yyyy-M...	Date	<input type="checkbox"/> var1

row1

Column
Project
Project_starting_Date
Project_ending_Date
PI
Cytometer_ID
Station
Depth
Latitude
Longitude
Study_area
Samples_Operator
Standards_Reference
Clustering_Method
Observation_Type
Platform_Type
Platform_ID
Platform_tiolity
Sampling_Date
Alysis_Date
Standardized_me
Selection_Set

Methode

Expression	Column
row1.File	id_meth
row1.Selection_Set	file
row1.Beta_1	expert_nomination
row1.Beta_0	coef_a
row1.Mean_Total_FWS_var1	coef_b
row1.SD_Total_FWS_var2	var_x1
row1.Mean_Length	var_x2
row1.SD_Length	moy_ESD
Var.var1	sd_ESD
	Up_date_Date

Link between input data and target fields

Éditeur de Schéma Editeur d'expression

row1

Colonne	Clé	Type	<input type="checkbox"/>	N..	Modèle date (...)	Length	Precision	Défaut	Comm...
Project	<input type="checkbox"/>	String	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>		14	0		
Project_starting_Date	<input type="checkbox"/>	Date	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	"dd/MM/yyyy"	10	0		
Project_ending_Date	<input type="checkbox"/>	Date	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	"dd/MM/yyyy"	10	0		
PI	<input type="checkbox"/>	String	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>		12	0		
Cytometer ID	<input type="checkbox"/>	String	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>		10	0		

Methode

Colonne	Clé	Type	<input type="checkbox"/>	N..	Modèle date (...)	Length	Precision	Défaut	Comm...
id_meth	<input type="checkbox"/>	int	<input type="checkbox"/>	<input type="checkbox"/>		11	0		
file	<input type="checkbox"/>	String	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>		45	0		
expert_nomination	<input type="checkbox"/>	String	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>		45	0		
coef_a	<input type="checkbox"/>	Float	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>		12	31		
coef b	<input type="checkbox"/>	Float	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>		12	31		

Appliquer OK Annuler

CYTOBASE data accessibility



Javascript library



Google Developers

dygraphs

Javascript charts

CHROME +



<https://chrome.mio.univ-amu.fr/chrome-cytobase/>

(Free access)



<http://www.mio.univ-amu.fr/cytobase/>

(Access within the MIO)

CYTOBASE data accessibility

Continuous and High Resolution Observation of the Mediterranean Sea

CHROME



Map showing the location of Carthage (Tunisia) with coordinates: N38°48'47.27, E010°28'41.55 (38.8131, 010.4782). The map displays the coastline and surrounding waters. A pop-up window for 'CARTHAGE' provides details: Re-Re/Passenger Ship, État: Stopped, Vitesse / route: 0kn / 358°, Tirant d'eau: 6.5m, Réçu: 2016-10-10 21:13 UTC (AIS origine: 2063 (cometon)).

Zone d'étude (Area of Study):

Mediterranean Sea - Western basin

Trajet (Transect):

MRS-TUN : du 24-03-2016 au 25-03-2016

Date de début (Start Date)

23/03/2016

Date de fin (End Date)

26/03/2016

Note: Please refer to the date mentioned in "Transect"

Instruments :

- CytoSense
- Photos cellules phytoplanctoniques
- FerryBox (Données pas disponibles)

Rechercher



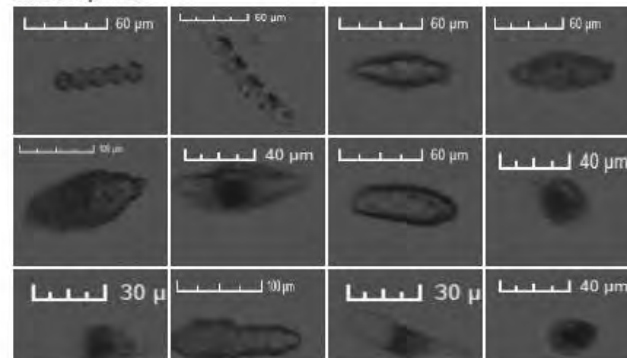
Résultats :

Pour le trajet de Marseille à Tunis à bord du C/F CARTHAGE
Du 2016-03-23 au 2016-03-26

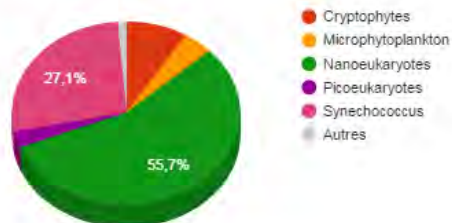
Nombre total des stations :

- CytoSense :43

Galerie photo:

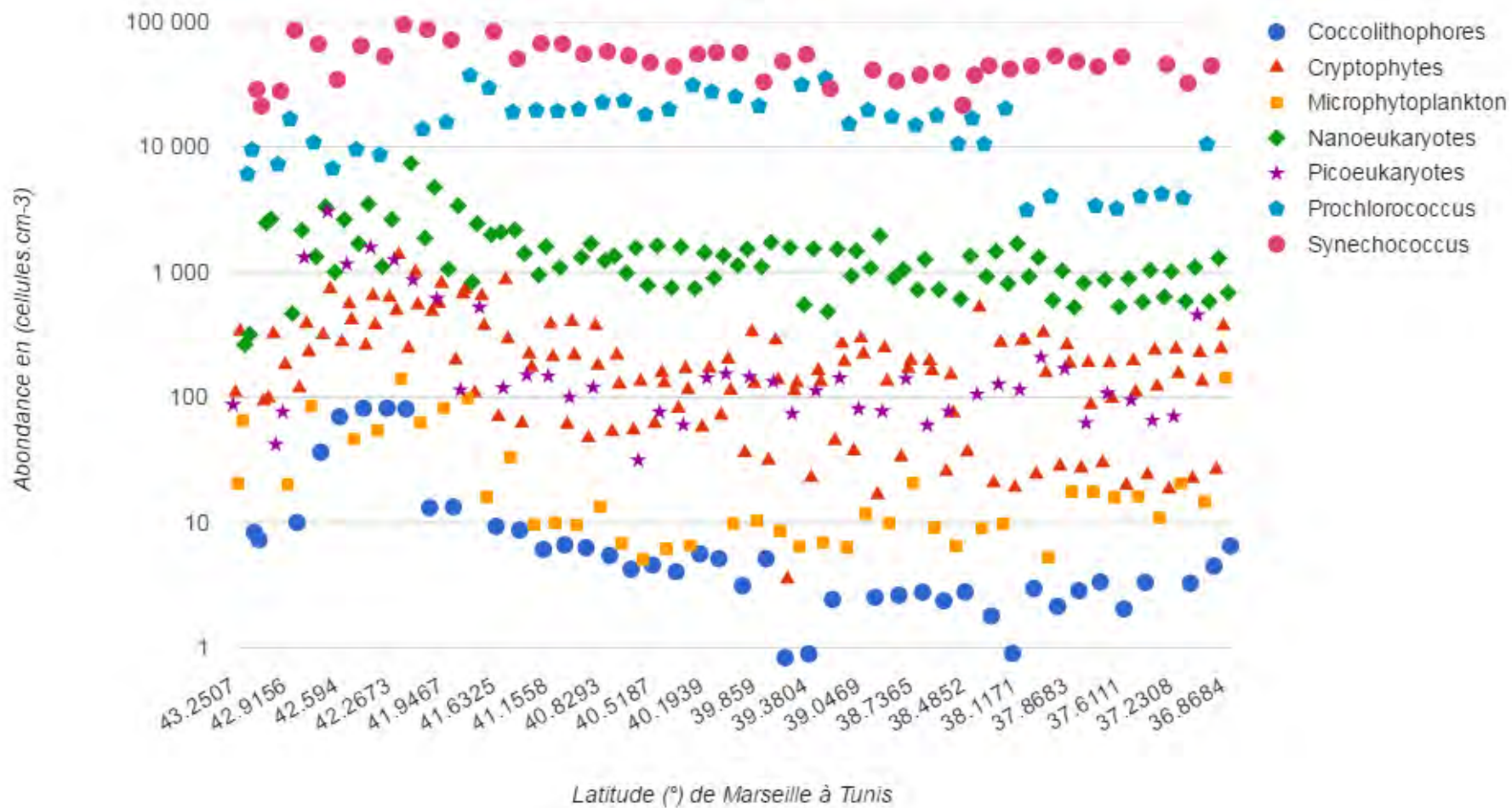


Total Fluorescence Rouge (u.a.cm⁻³)

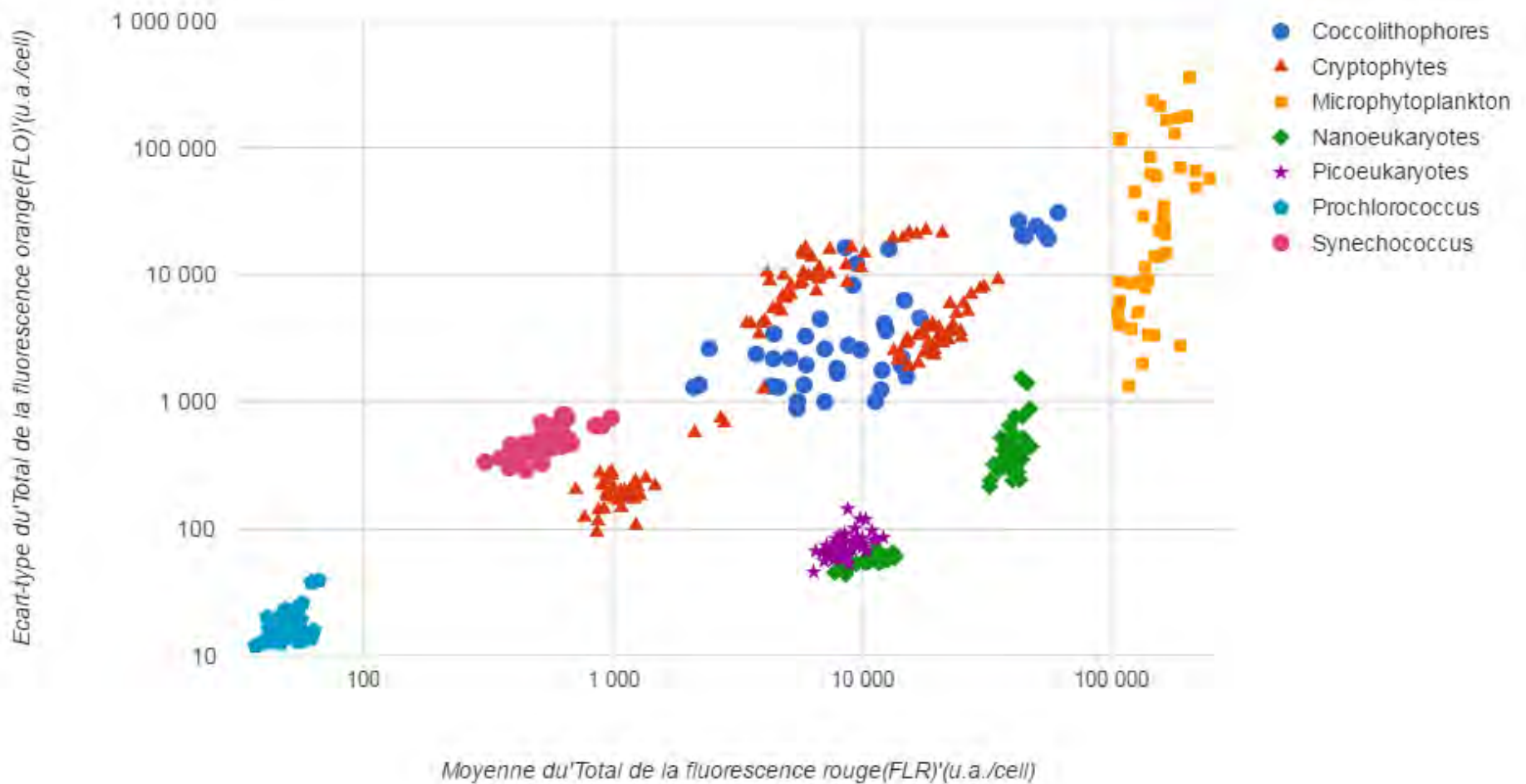


Total Abundance (cell.cm⁻³)





Cytometric distribution parameters per group



CYTOBASE data accessibility (Access within the MIO)



FLOW CYTOMETRY DATABASE



Zone d'étude:

Nom et date du Projet:

Date de début:

Date de fin:

Flow cytometer :

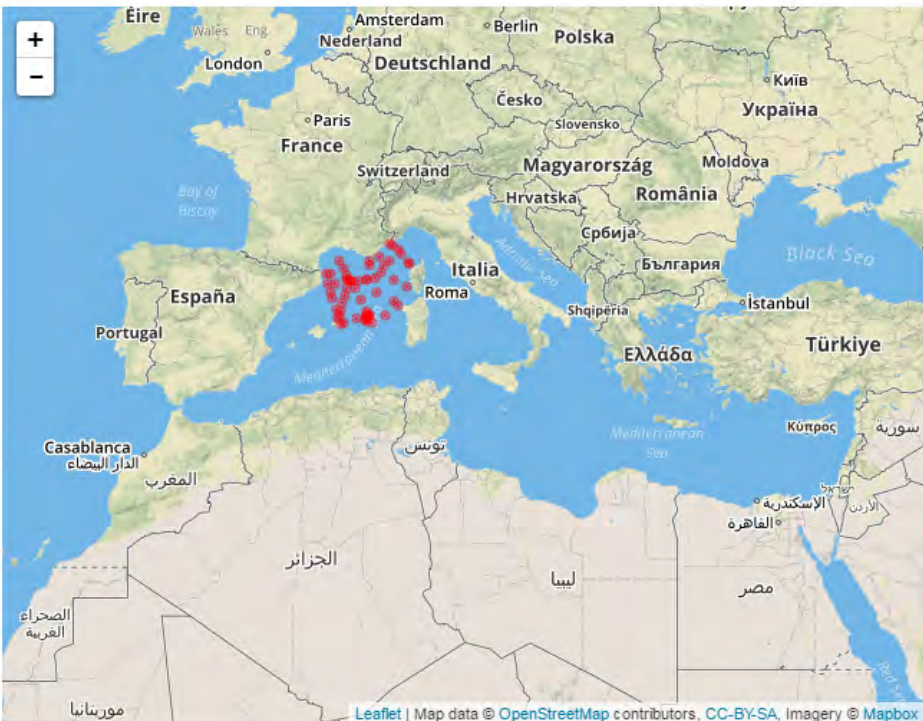
- CytoSense
- FACS Calibur
- BD Influx
- All instruments
- Phytoplankton groups

Heterotrophic Bacteria

Search



FLOW CYTOMETRY DATABASE



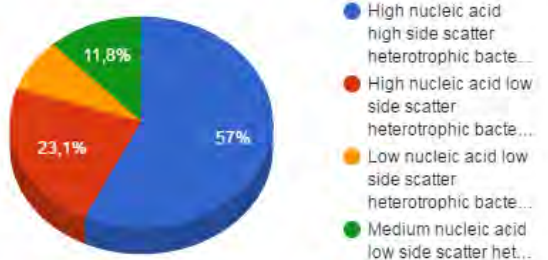
Results :

For the period from 2013-03-31 to 2013-04-30
Project name **DEWEX LEG2**
Flow Cytometer : **FACS Calibur**
Station number : **59**

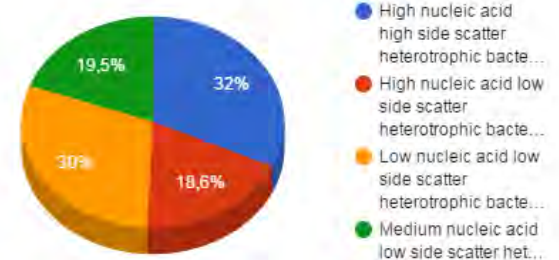
No pictures

[Download Data](#) [Print](#)

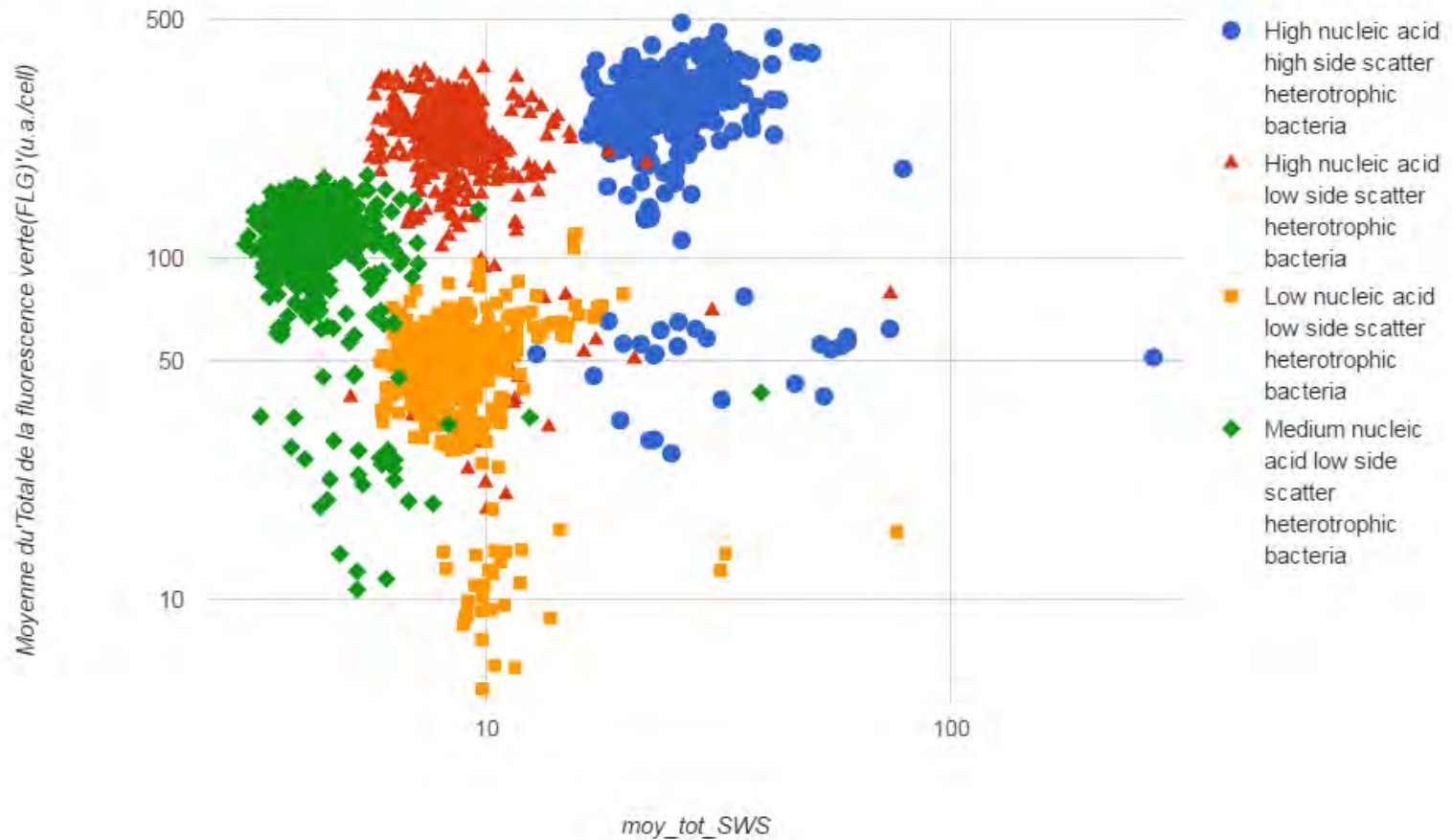
Contribution to Green Fluorescence (u.a..cm⁻³)



Relative Abundance (cell.cm⁻³)

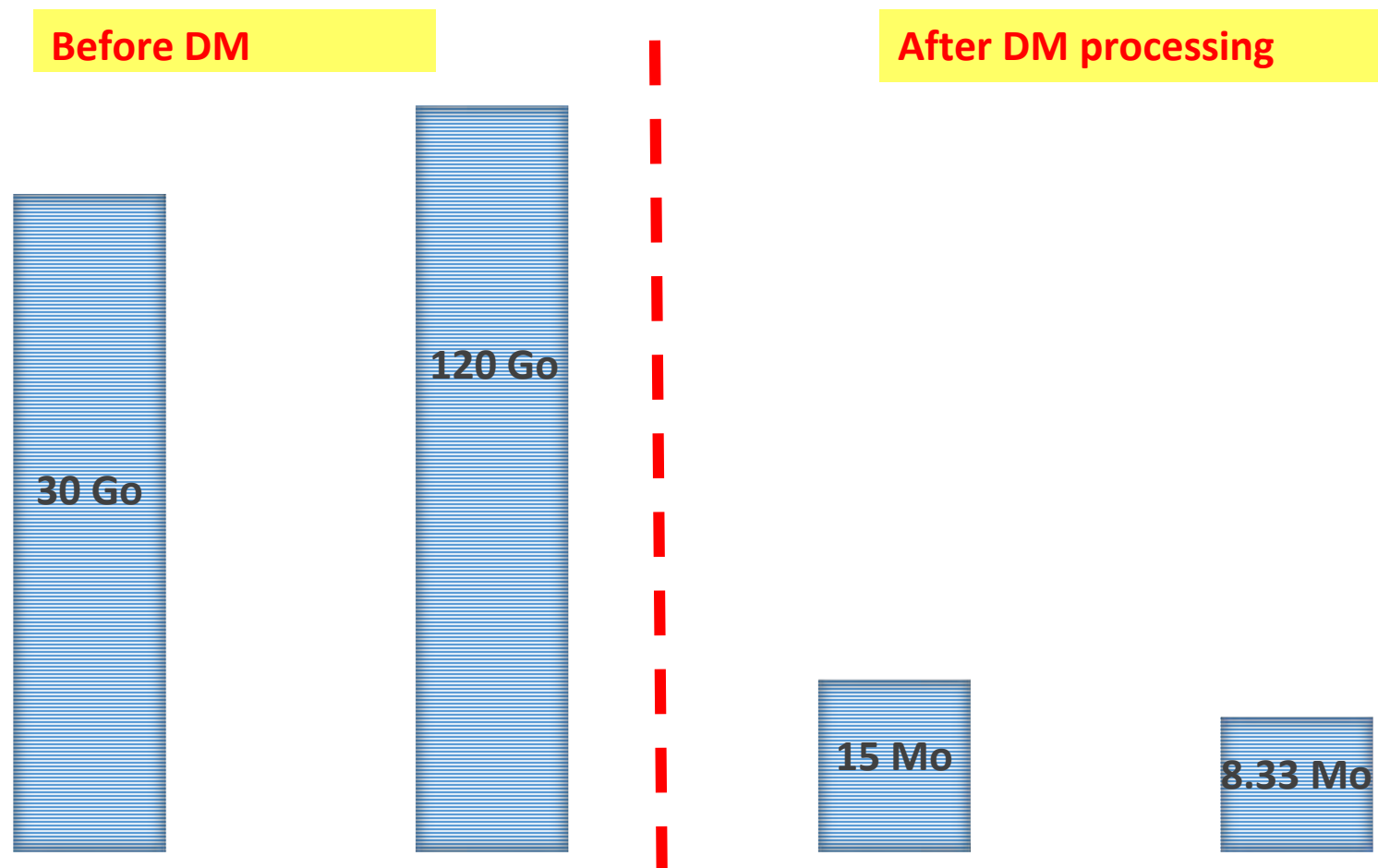


Cytometric distribution parameters per group



CYTOBASE data management method of the MIO

- Case of a 9 days Cruise : 1 sample / 20 min



Data acquisition

data analysis

data consolidation

Cytobase

Next Step!

The screenshot shows the NEMO software interface. The main window displays a data table with columns for Project, Date, Station, and various parameters. A 'Parameters selection' dialog box is open, showing a list of parameters with their units. The 'Data Description' panel on the right shows options for validating steps, resetting, and selecting parameters (P09, P01 via P09, P01 via P02). It also includes sections for 'Measured' (below sea surface, below sea bed, above sea level), 'Vertical References' (depth below sea surface, depth below sea bed, pressure, height above sea level), and 'fall rate'.

Parameters selection P09

Type in name to filter

CODE - NAME	Unit
ABCP - ALPHA BETA CAROTENES	milligram/...
AG63 - Ag<63um IN DRY WEIGHT...	milligram/...
AGSX - Ag IN DRY WEIGHT SEDIM...	milligram/...
AL63 - Al<63um IN DRY WEIGHT ...	milligram/...
ALKW - ALKALINITY	micromole...
ALKY - ALKALINITY	millimole/...
ALTS - HEIGHT ABOVE MEAN SEA...	meter
AMIS - SEDIMENT AMINO-ACIDS	microgra...
AMON - AMMONIUM (NH4-N) C...	millimole/...
AMOP - AMONIUM IN SEDIMENT...	millimole/...
AMOW - AMMONIUM (NH4-N) C...	micromole...

Data Description

Validate step

Reset

Parameters list

P09

P01 via P09

P01 via P02

Measured

below sea surface

below sea bed

above sea level

Vertical References

depth below sea surface

depth below sea bed

pressure

height above sea level

fall rate

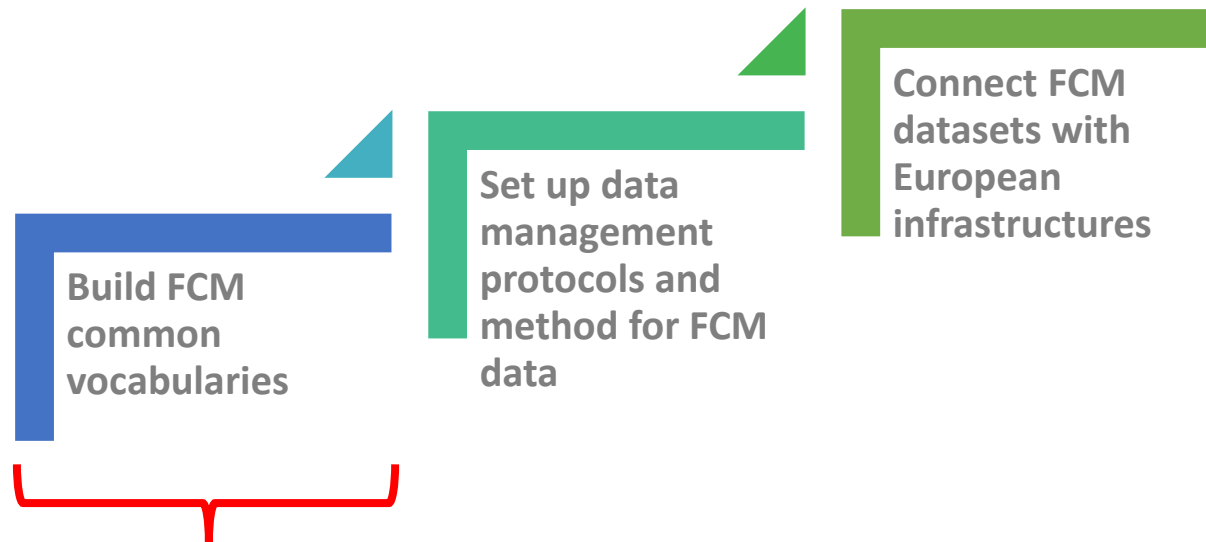
select

Not enough and accurate vocabulary for flow cytometry in standard lists

Next Step!

➔ WP9.5.2 of SEADATACLOUD (VLIZ, MIO, NERC-BODC and ICES):

Ingesting, validating, long-term storage and access of Flow Cytometry data



🔑 We need to work with the automated and *in situ* FCM Community:
Task 3.1 of JERICO Next

Conclusion

- **Smart storage and sustainability of FCM data with net decrease on file size**
- **CYTOBASE is a dynamic and user friendly web-based interface**
- **The next step, we are working on FCM data interoperability with French national DB (SOMLIT) and international DB (SeaDataNet) (in compliance with INSPIRE Directive)**
- **It will be possible to link with task 5.4 (JERICO) and WP9 of SeaDataCloud so as to work with large FCM community.**
- **By working on FCM common vocabulary, we can adopt the best practice tools and data management method used in MIO (CYTOBASE) and in SeaDataNet (SeaDataCloud).**

Thank you for your attention