



**SIST 2016 : Séries Interopérables et Systèmes de Traitement**  
*29-30 sept. 2016 Montpellier (France)*



**CHROME**

# Gestion des données de Cytométrie en Flux

## *Flow Cytometry Data Management*

*Montpellier, le 29 septembre 2016*

Soumaya Lahbib (IE MIO)

Mathilde Dugenne (Doc. MIO), Maurice Libes(OSU), Melilotus Thyssen (Chercheur MIO), Chérif Sammari (Chercheur INSTM), Gérald Grégori (Chercheur MIO), et Pierre Marrec (Post-Doc MIO)



- 1. Présentation du projet A\*MIDEX CHROME**
- 2. Introduction au jeu de données de Cytométrie en Flux (CF) automatisée?**
- 3. Problématique de la gestion des données de CF**
- 4. Workflow de la gestion des données de CF**
- 5. Consolidation des données**
- 6. Conception, Intégration et développement**
- 7. Conclusion & perspectives**

# 1. Présentation du projet A\*MIDEX CHROME

Continuous and High Resolution Observation of the Mediterranean Sea



**CHROME**

Projet CHROME ▾

Equipements ▾

Données et résultats ▾

Contacts



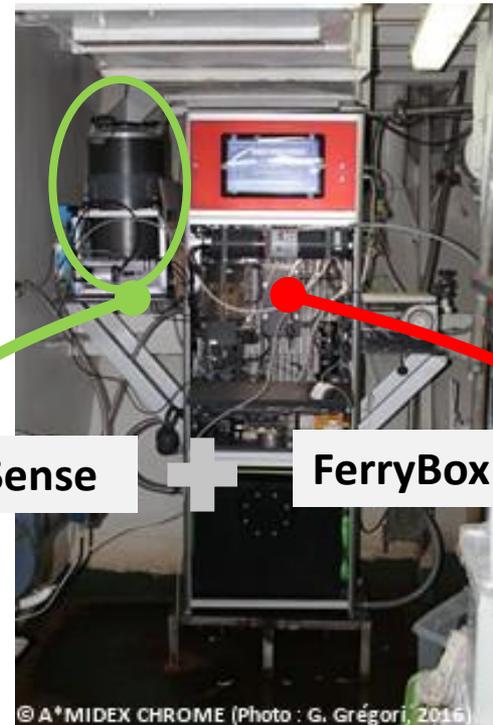
Actualités

Rechercher



# 1. Présentation du projet A\*MIDEX CHROME

## Pont 1 – C/F CARTHAGE



CytoSense

FerryBox



INSTM



mio  
Institut Méditerranéen  
d'Océanologie

./30min

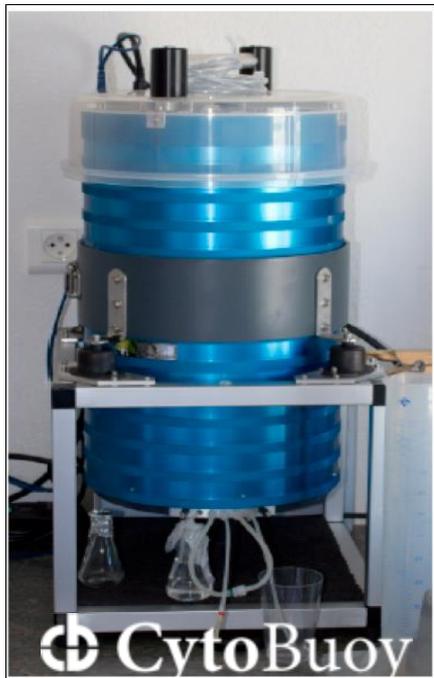
./1min

### C/F CARTHAGE trajectoires

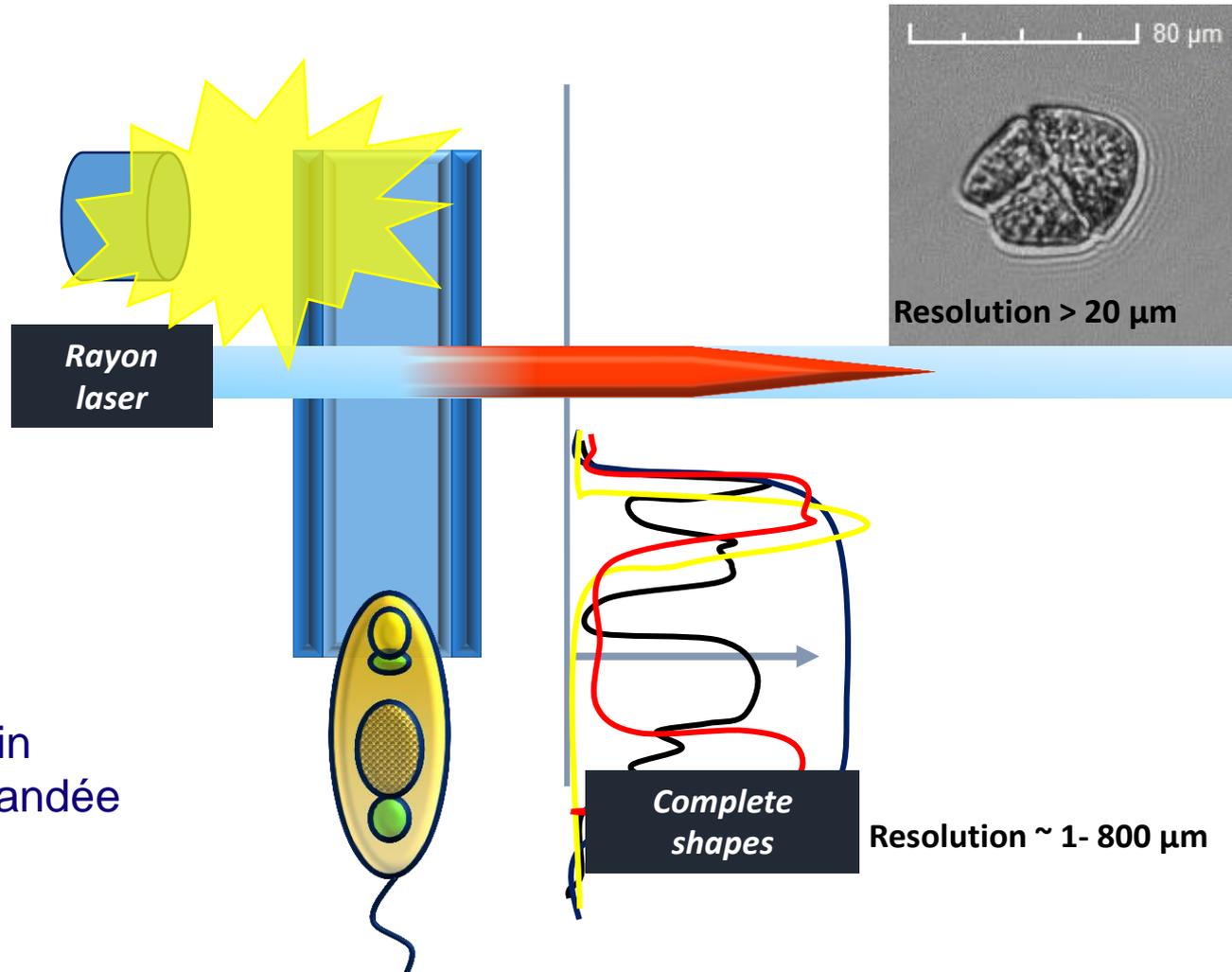
- Groupes fonctionnels phytoplanctoniques
- Abondance des ces groupes
- Fluorescences/dispersion par cellule
- Estimation de la taille
- Photos des cellules phytoplanctoniques (identification taxonomique >20 μm)

- Température
- Salinité
- Fluorescence
- Turbidité
- pH
- pCO<sub>2</sub>
- Oxygène

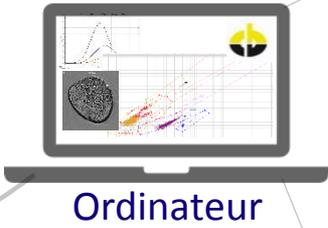
# Nouvelle technologie pour l'étude de la diversité fonctionnelle des groupes phytoplanctoniques à haute résolution spatio-temporelle



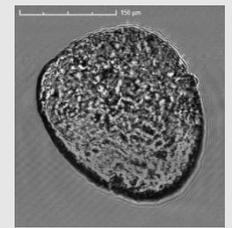
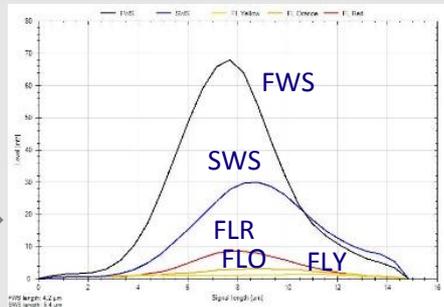
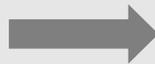
Une analyse (5 cm<sup>3</sup>)/15 min automatisée et télé-commandée



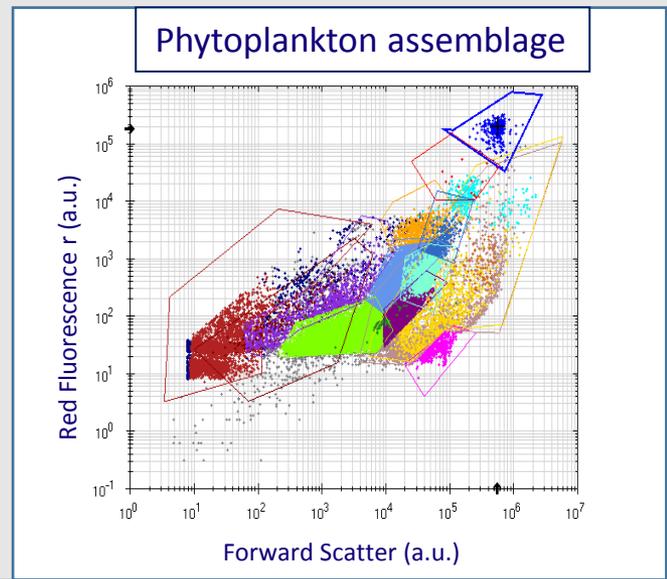
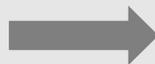
## 2. Introduction au jeu de données de CF automatisée?



1 cell



$1 \times 10^6$  cells



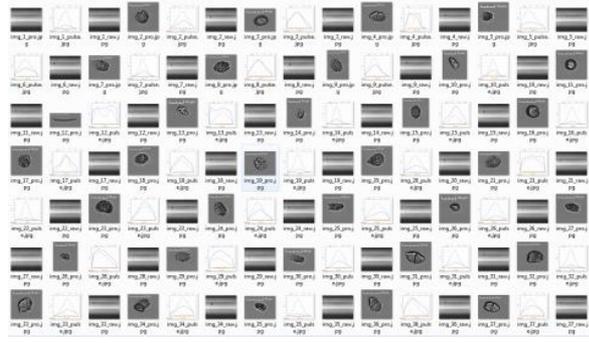
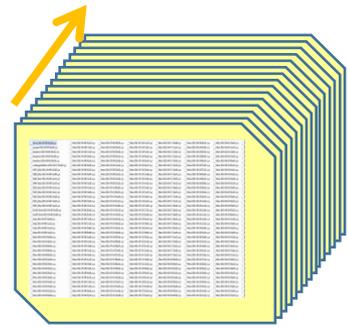
# 3. Problématique de la gestion des données de CF



Mesures par le CytoSense



Classification manuelle



## Des fichiers statistiques (CSV) :

- Les valeurs Moyenne des propriétés optiques
- Comptage des cellules
- Photos



1. Grande quantité de données
2. Données mémoirophages
3. Absence de base de données de CFA



La BD CYTOBASE

# 4. Workflow de la gestion des données de CF

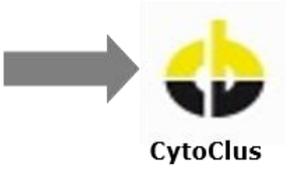


Table Données  
Table photo  
Collection  
photos

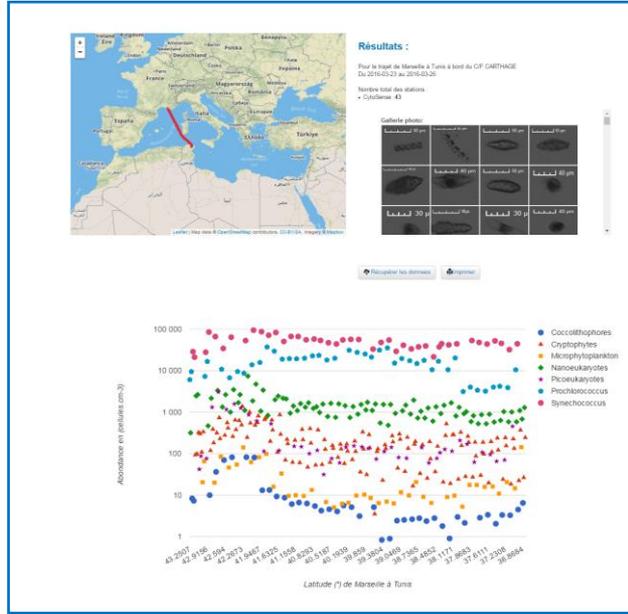


**Acquisition**

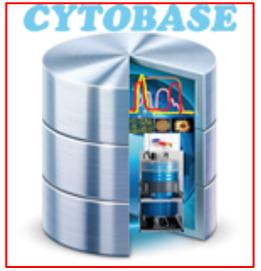
**Analyses**

**Consolidation**

**Validation**



**Restitution**



**Intégration**

## 5. Consolidation des données

Fichiers de sorties	Format	Nombre	Paramètres
<b>Moyenne</b> avgprops_process_xx-xx-xxxx avgprops_process_xx-xx-xxxx avgprops_process_xx-xx-xxxx etc..	csv	1 ou +	Selection set, File name, Date, Time, Total Analyzed Volume [mL], Number of Particles Min. TOF, Max. TOF, Mean TOF, etc...
<b>Comptage</b> counts_process_xx-xx-xxxx counts_process_xx-xx-xxxx counts_process_xx-xx-xxxx counts_process_xx-xx-xxxx etc...	csv	1 ou +	File name, Volume [mL], Trigger chn, Level [mV], Date Particles, Beads 2 mu - Number, Beads 2 mu - pct/Tot Beads 2 mu - N/mL, Microphytoplankton - Number, etc...

# 5. Consolidation : Cytobase Input Processor (M. Dugenne, 2015)

## Cytobase data

Parcourir... 11 fichiers sélectionnés.  
Upload complete

**Warning:**

Upload successful



Create Inputs folder

Métadonnées

Project and samples context   Raw data   Size conversion

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

Données

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: Beads 2 mu

Trigger

Channel/Level: FL Red 10

PMT's amplification

SWS: 50

Standardized name

Cluster: Standard beads

Associate

# 5. Consolidation : Cytobase Input Processor (M. Dugenne, 2015)

Photos

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

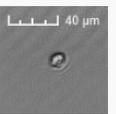
**Add samples pictures**

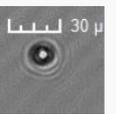
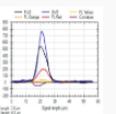
Select file

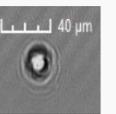
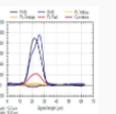
Add pictures  
 15 fichiers

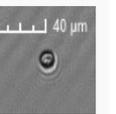
Check all

Show  entries   Search:



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

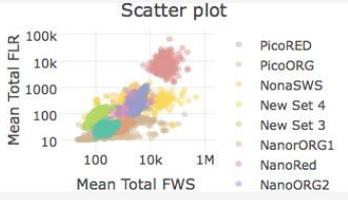
Validation

Select station

X axis

Y axis

**Scatter plot**

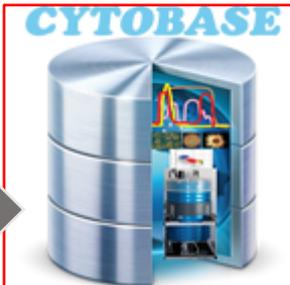


- PicoRED
- PicoORG
- NonaSWS
- New Set 4
- New Set 3
- NanorORG1
- NanoRed
- NanoORG2

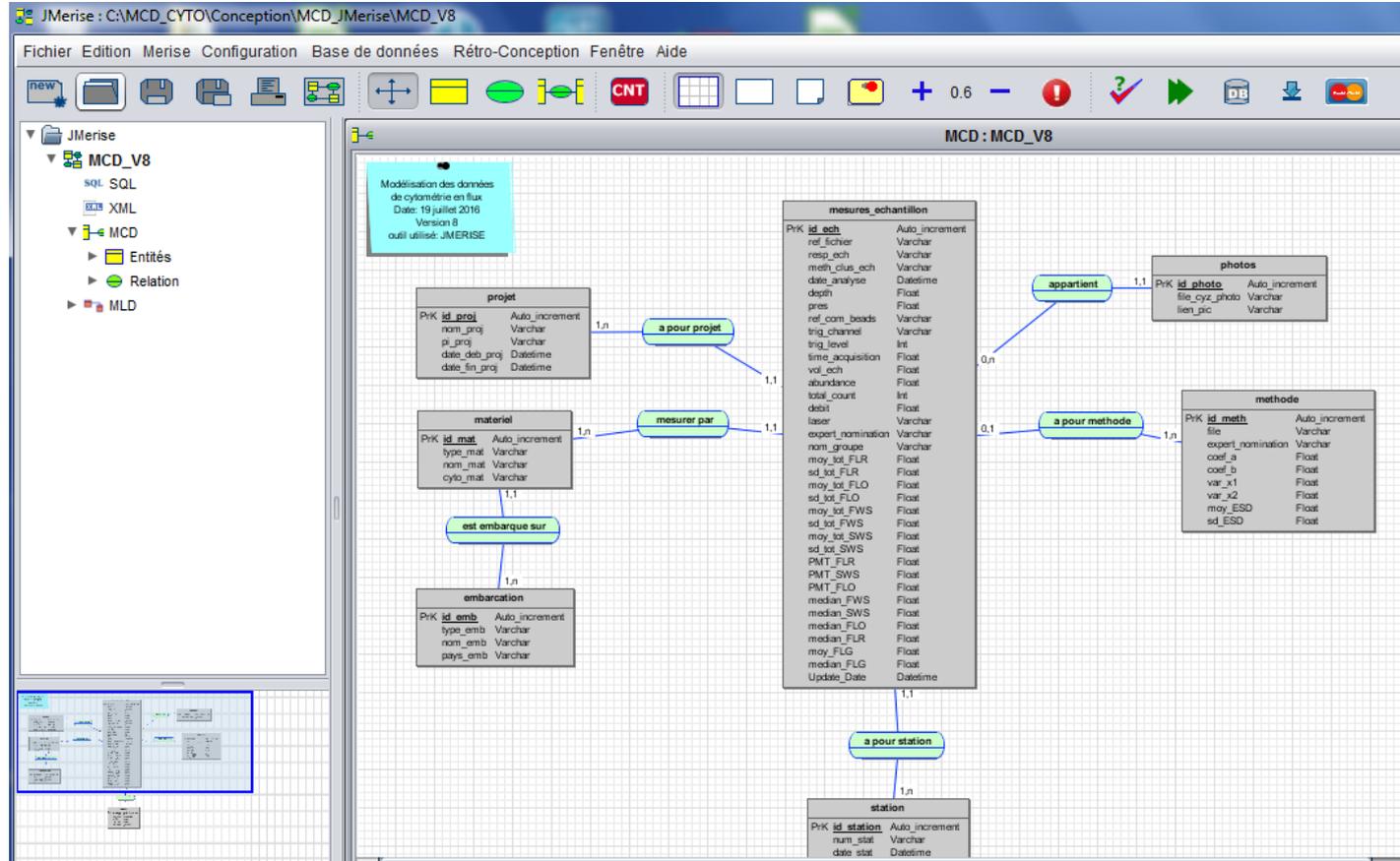


Download Table

Table Données  
Table Photo



**JMERISE**  
Logiciel de modélisation  
des MCD pour Merise



# 6. Conception, Intégration et développement

phpMyAdmin

Structure SQL Rechercher Requête Exporter Importer Opérations plus

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	-
<b>7 tables</b>	<b>Somme</b>	<b>46 406</b>	<b>InnoDB</b>	<b>latin1_swedish_ci</b>	<b>8,9 Mio</b>	<b>0 0</b>



MySQL Workbench

- Communication avec la base
- Gestion et administration
- Modification et interrogation plus rapide

MySQL Workbench

CytoBase\_Local

MySQL Connections

CytoBase\_Local (root, 127.0.0.1:3306)

cytobase-serveur (cytobase, mysqlmio.univ-amu.fr:3306)

Models

Diagram\_MCD

C:\MCD\_CYTO\Conception

mydb\_cytobase\_local

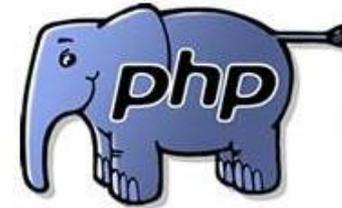
02 Jul 15 11:51



The screenshot displays the Talend Open Studio interface. The main workspace shows a job design with several components: 'Data\_DEWEG\_LEG2', 'tMap\_1', 'mesures\_v6 (Main)', 'mesures\_echantillon', 'station', 'materiel', and 'methode'. The job is titled 'Job mesures\_echantillon\_v6 0.1'. The left sidebar shows the 'Repository' with a tree view of 'Connexions aux bases de données' and 'Schémas des tables'. The bottom panel shows the 'Job mesures\_echantillon\_v6' execution logs, including a message: 'Message 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 scartie=0]'. The right sidebar contains a 'Tr...' panel with various integration components like 'Applicati...', 'Bases de...', 'Big Data', 'Business...', 'Cloud', 'Code Util...', 'Divers', 'DotNET', 'ESB', 'Fichier', 'Ecriture', 'Internet', 'Logs...', 'tAssert', 'Orchestr...', 'Qualit...', 'Matchin', 'Système', 'Talend M...', 'Trans...', and 'Champs'.

- Plateforme de développement des traitements ETL
- Connectivité avec les BD, types de fichiers, etc.
- Générateur de code en langage Java ou Perl
- Optimisation du temps d'intégration des données

## 6. Conception, Intégration et développement

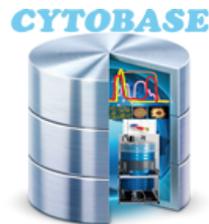


Google

Developers

dygraphs

**CHROME** +

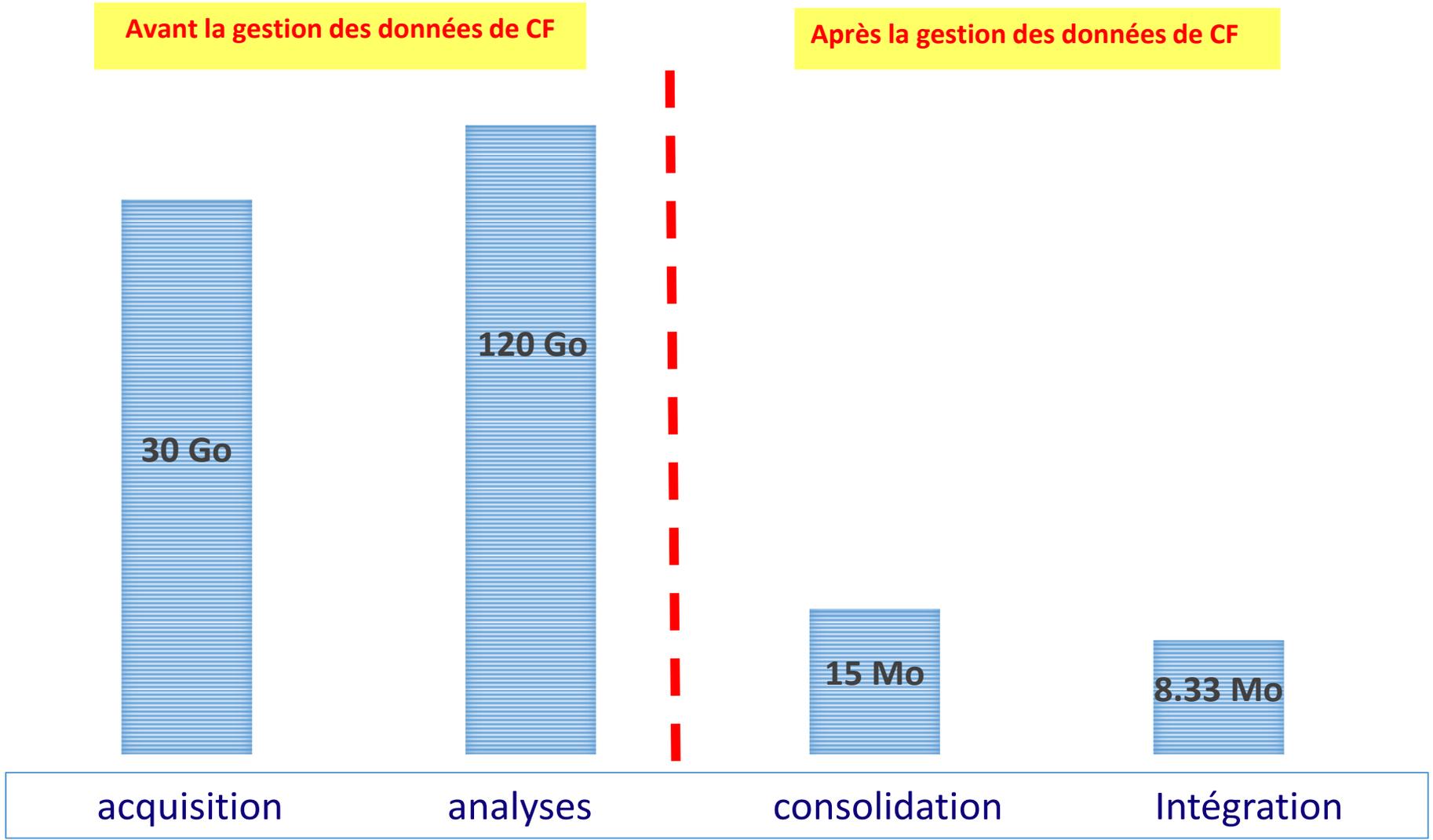


<https://chrome.mio.univ-amu.fr/chrome-cytobase/>  
(Accès libre)

<http://www.mio.univ-amu.fr/cytobase/>  
(Accès restreint au MIO)

## 7. Conclusion & perspectives

- Cas d'une campagne de 9 jours: 1 échantillonnage/ 20 min



## 7. Conclusion & perspectives

---

- Un système de mesures automatisées à haute résolution : la FerryBox et le Cytomètre en Flux de type CytoSense.
- Mettre en place une nouvelle méthode de gestion des données de CF (automatisée) et classique.
- La prochaine étape consiste à rendre cette base de données interopérable avec le réseau national (SOMLIT) et international (selon les normes INSPIRE)
- Participation dans le projet EU (*SeaDataCloud*) pour travailler sur le vocabulaire commun, les standards et les normes à développer pour la CF.

# Remerciements

---

- Le réseau régional des développeurs de Provence ProDev
- Les journées JDEV organisées par le CNRS
- Equipe Osupytheas du M.IO

