



Avancement chaîne de traitement de cytométrie en flux

*M. Libes (MIO), M. Sourisseau (IFREMER)
(+ M. Thyssen & G. Gregori (MIO))*

Septembre 2020



Plan

- Etude comparative des workflow de cytométrie existants
 - *MIO/OSU Pytheas et*
 - *Laboratoire d'Ecologie Pélagique Ifremer*
 -
- Réécriture d'une chaîne de traitement de cytométrie complète en Python



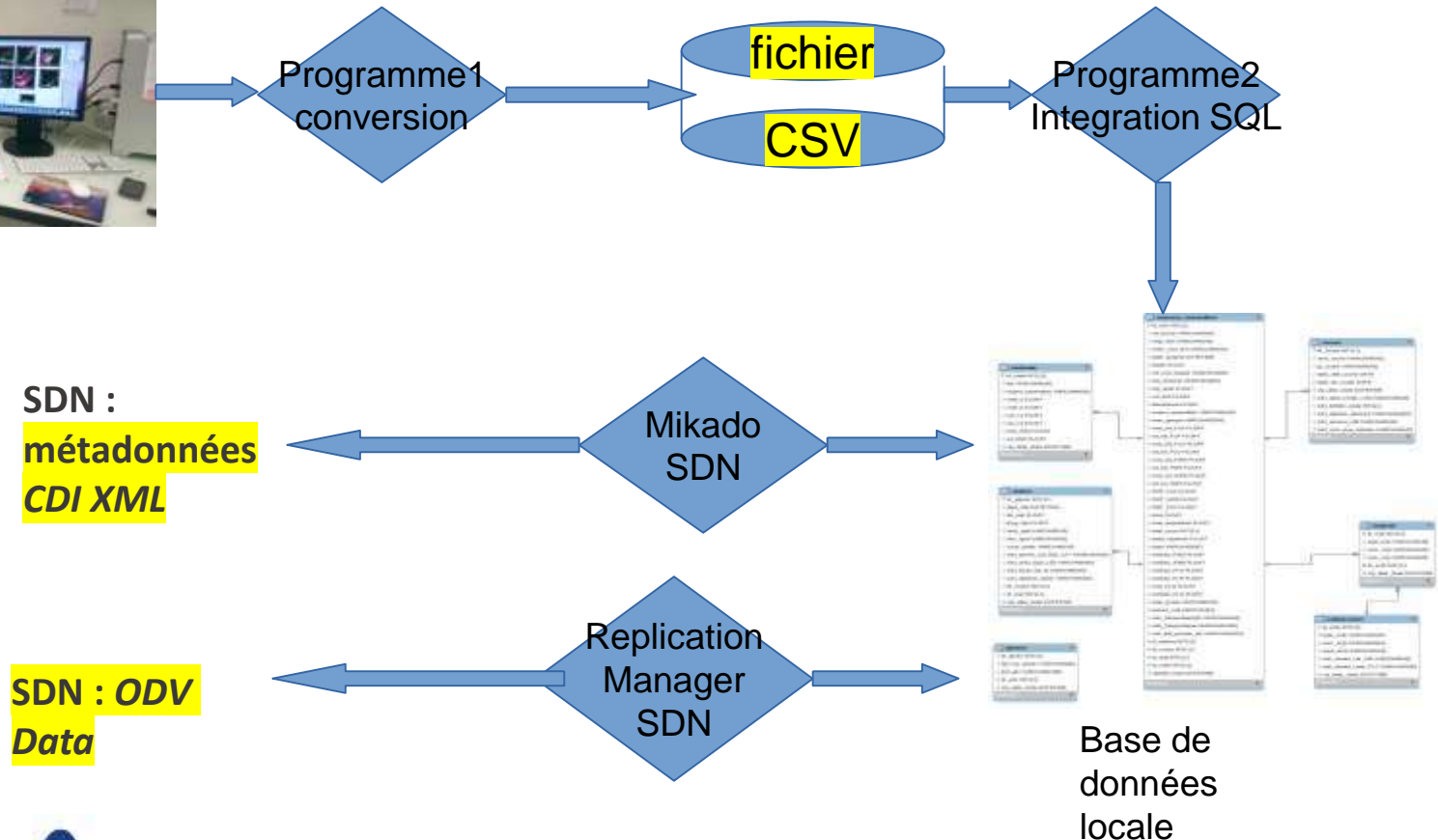
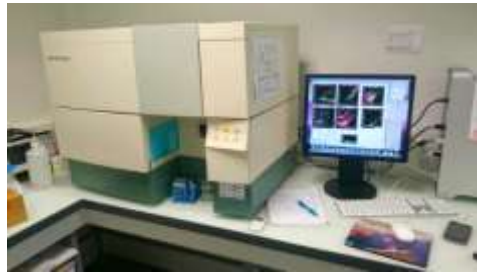
Etude comparative des workflow de cytométrie existants



- **Objectifs**
- Connaître et comparer les chaînes de traitement des données de cytométrie établies au
 - *MIO/OSU Pytheas et*
 - *Laboratoire d'Ecologie Pélagique Ifremer*
- Etablir des recommandations pour profiler une chaîne de traitements commune qui pourrait être recommandée par le CES Odatis et utilisée dans un CDS cytométrie
- Étude détaillée accessible sur :
 - <https://docs.google.com/document/d/1MrRDDRzuA4thz3evZ0kyqsG4RKAjpOFk14EAXiqaM0M/edit?usp=sharing>



Résumé du workflow : *étapes communes importantes*



Conclusions sur fichier CSV

- En sortie des cytomètres il est nécessaire de créer un fichier de données intermédiaire, consolidé ayant fait l'objet d'un contrôle qualité, au format CSV (lignes, colonnes)
- Les entêtes de nos fichiers sont quasiment identiques
- Entête du fichier CSV à l'OSU Pytheas (en rouge les différences avec Laboratoire Ecologie Pélagique.)
- ***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.Nationality, Sampling.Date, Analysis.Date, Standardized.name, Selection.Set, File, Volume, Trigger.Channel, Trigger.level, SWS.amplification, FLO.amplification, FLR.amplification, Abundance, Mean.Total.FWS_varx1, SD.Total.FWS_varx2, Mean.Total.SWS, SD.Total.SWS, Mean.Total.FLR, SD.Total.FLR, Mean.Total.FLO, SD.Total.FLO, Mean.Max.FLR, SD.Max.FLR, Beta.0, Beta.1, Mean.Length, SD.Length***



Conclusions fichier CSV

- Différences
 - ordre des colonnes :
 - l'ordre des colonnes diffère : peu important d'un point de vue programmatique,
 - l'important est le libellé des colonnes : quasiment identique
Cependant pour une homogénéité totale on pourrait envisager d'avoir aussi le même ordre de colonnes
 - intitulés des colonnes :
 - Selection.Set (aka Expert_Nomination) absent dans le fichier CSV Ifremer
 - Mean.Max.FLR, SD.Max.FLR, parametres absents dans le fichier CSV Ifremer

•



Conclusions fichier CSV

Au laboratoire d'Ecologie pelagique

- Le fichier CSV créé comporte déjà la somme des abondances des groupes standardisés. Il correspond à la “vue” MIKADO_<le_nom_dela_campagne>. de la BD Cytobase de l'OSU Pytheas

Au MIO/OSU : on a un niveau de détail supplémentaire avec un fichier intermédiaire avant la somme des abondances des groupes:

- le 1er fichier CSV en sortie du cytomètre comporte une colonne “Selection Set” correspondant aux détails des groupes déterminés par l'expert cytométriste dans les analyses d'une date de prélèvement donnée
- Le 2eme fichier CSV fournit pour une date de prélèvement donnée “sampling_date”,
 - i) la somme des abondances des groupes expertisés par le cytométriste pour une date de prélèvement donnée
 - ii) Les moyennes pondérées des propriétés optiques de ces groupes pour une date de prélèvement donnée

Conclusions BD

- Les fichiers CSV produits sont intégrés dans une BD locale (optionnel)
- Parmi les différences on note la présence d'un peu plus de métadonnées dans les tables de la BD Cytobase OSU Pytheas, dont certaines utilisant le vocabulaire contrôlé du BODC pour l'intégration dans SeadataNet
- Peu importe le schéma relationnel de la BD des laboratoires. Ce qui importe ce sont :
 - les fichiers terminaux produits en sortie
 - Format interopérable ODV ou NetCDF demandé
 - Les métadonnées nécessaires
 - L'utilisation du vocabulaire contrôlé du BODC

Le vocabulaire interopérable pour la cytométrie est en cours de publication et issue d'un consensus entre 32 experts internationaux.



Conclusions : vocabulaire contrôlé

- Dans le cadre d'une nécessaire interopérabilité, il est important d'utiliser des tables du vocabulaire standard existantes dans SDN:
- C17_SDN_Embarcation_code.csv : nom et code des bateaux : 35LU "Le Suroit"
- L06_SDN_Embarcation_Type.csv : type du bateau : research vessel
- L22_SDN_cytometre_code.csv : le code des cytometres : TOOL1209,CytoBuoy CytoSense flow cytometer,
- et surtout la table des codes des groupes de cytométrie
 - F02_SDN_Cytometrie.csv : code des groupes de phytoplancton F0200003,Synechococcus,



Conclusions



- Quelques différences mineures dans le fichier CSV issu des Cytomètres :
- Dans le cadre d'un CDS :
 - Aller vers une standardisation des fichiers CSV issus des analyses de données cytométriques. Se mettre d'accords sur le format interne de ces fichiers et sur les métadonnées nécessaires (le format CSV "Mikado" étant déjà une bonne base initiale)..
 - Aller dans le sens de l'utilisation de vocabulaire contrôlé standardisé issus d'un consensus d'experts
 - Le format national de sortie de cytométrie en flux repose déjà sur le travail effectué dans le cadre de SDC, et est effectué en parallèle par un consortium européen dans le cadre du projet JERICOS3. Une méthode nationale n'a de valeur que si elle est internationale, sinon, il faut tout recommencer
- Quelques différences dans les BD :
 - Le modèle de données interne des BD locales est peu important pour autant que les processus de traitement logiciels fournisse les fichiers de sortie corrects avec les métadonnées attenantes.
 - Dans le cadre d'un CDS : fournir des fichiers de cytométrie complets au format ODV ou NetCDF pour Odatis
- Dans le cadre d'un CES Odatis, se mettre d'accord sur le format et contenu des fichiers de cytométrie interopérables produits (quels paramètres doivent être rendus publics?)
- Fournir des outils logiciels en Python aux laboratoires pour compiler leurs données de cytomètres de manière identiques : avoir une batterie d'outils logiciel s'adaptant aux différentes générations de cytomètres

Travail réalisé au MIO/OSU Pytheas

WP Chaîne de traitement des données de cytométrie

- Mars-Juillet 2020 : Réécriture en *python* de toute la chaîne de traitement avec un stagiaire (*L. Izard, M. Libes*) pour des les programmes *Cytoclus3* et *Cytoclus4*
- On traite les données brutes des modèles « *cytoclus4* »
- La chaîne effectue désormais un grand nombre de contrôles qualité sur les données brutes issues des cytomètres
- Intégration des campagnes *chrome*, *peacetime*, *fumseck*



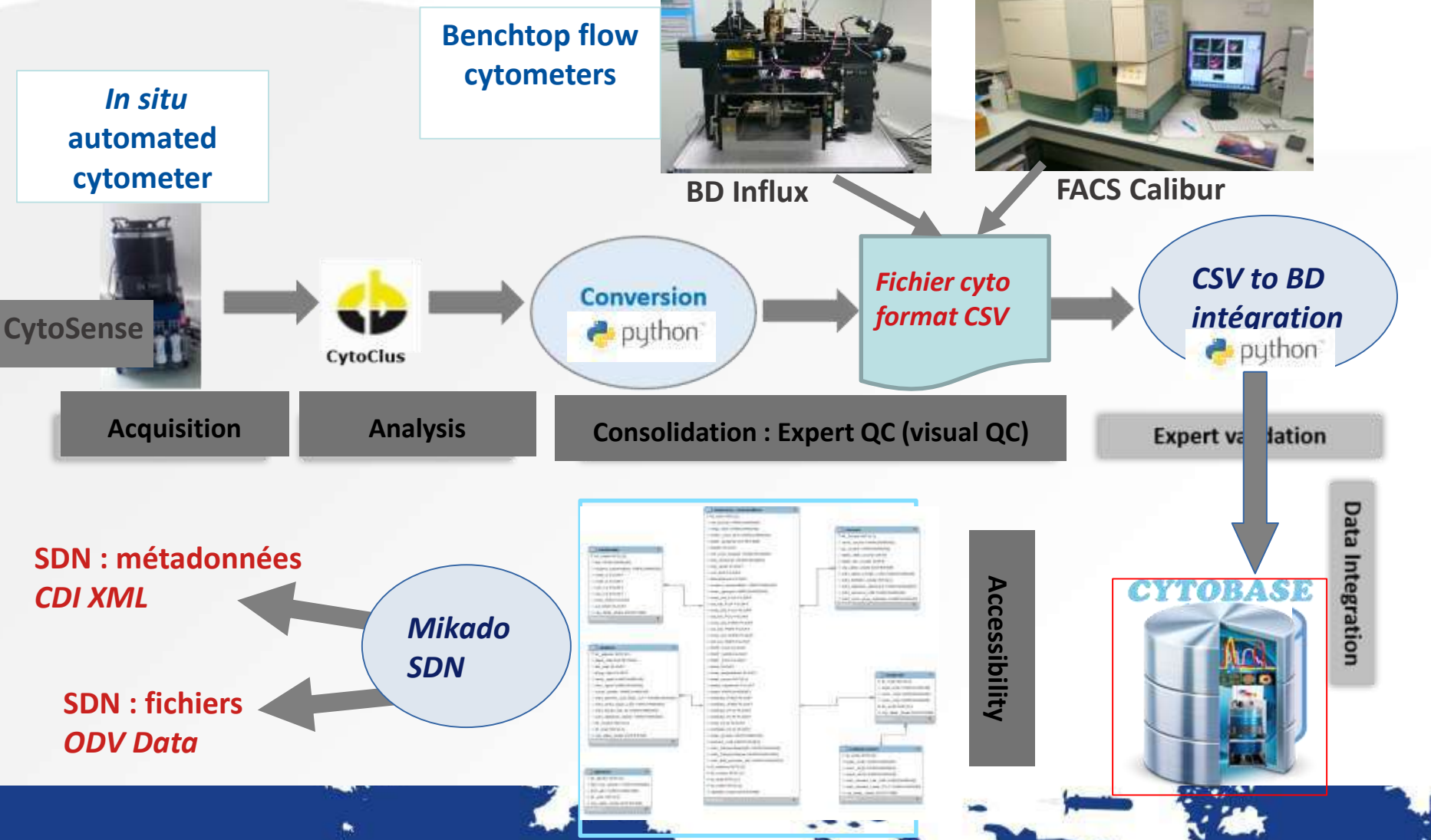


SeaDataNet

PAN-EUROPEAN INFRASTRUCTURE FOR OCEAN & MARINE DATA MANAGEMENT



FCM Data management Workflow



Travaux réalisés et futurs à DYNECO/PEL

Chaîne de traitement des données de cytométrie

Juillet 2020 – Mars 2021

- Intégration des images associées à l'imagerie en flux dans la BD locale (fini)
- Intégration des données de 2 suivis temporels (Octobre)
- Mise en place de procédure d'annotation (Novembre).
- Mise en place de procédure d'export de ces images vers des plateformes d'imagerie (Janvier).





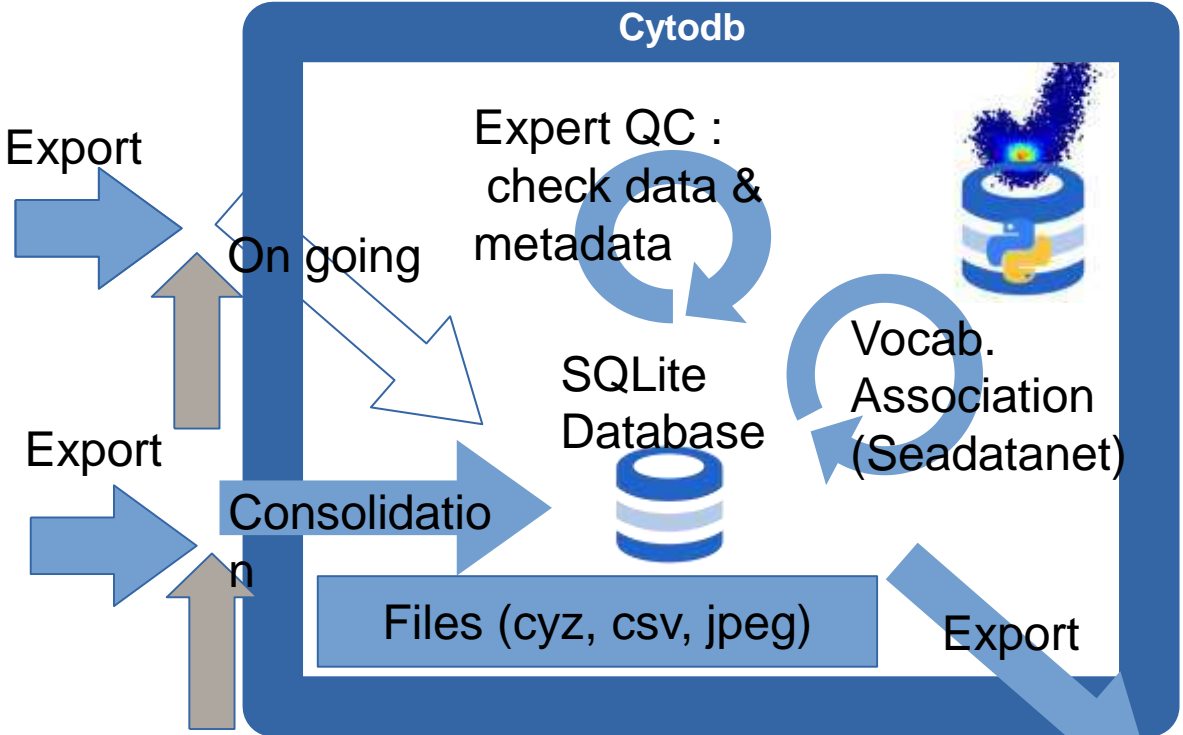
NovoCyt



Data acquisition (Lab or autonomous)



Manual Analysis



Inputs :
 Cruise, positions,
 ...
 Strains, Lineage ...
 Voc. Ass. clusters

