



Atelier technique juin 2019

Mercredi 5 juin 2019

CDS-MIO/OSU Pytheas

Patrick Raimbault, Maurice LIBES

Maurice.libes@osupytheas.fr





Présentation générale du CDS

- MIO / OSU PYTHEAS
 - Not yet...



Qui est inclut dans le CDS ?

- SNO :
 - SOMLIT (données hydrologie et cytométrie « de labo » sur sites Somlit : envoyées à OASU Bordeaux BD officielle de SOMLIT)
 - MOOSE (base hébergée au SEDOO de OMP) : Mesures temps réel (glider, CTD) = Coriolis) – Données marines temps différé = SISMER – SEANOE
 - EMSO (mouillages = SISMER – SEANOE, EMSO France)
 - Coast HF = temps réel (Coriolis)
 - Phytobs = Base pelagos (Roscoff)



Ne concerne que les données
océanographiques

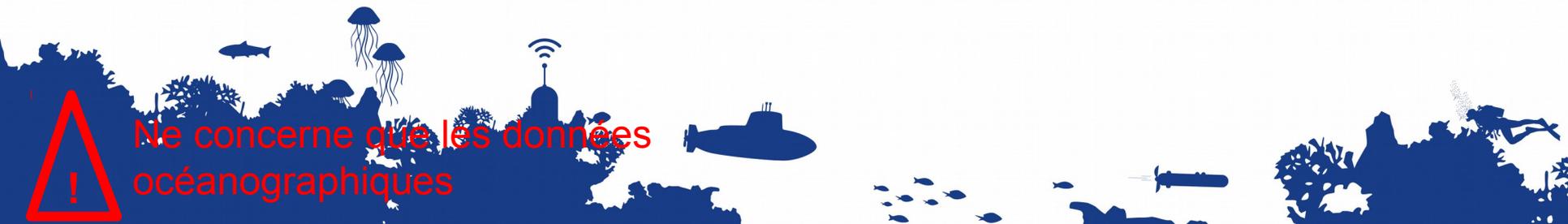


Qui est inclut dans le CDS ?

- **SO OSU** : activités d'observation et acquisition de données de *différentes UMR* de l'OSU Pytheas : (MIO pour l'Océano, IMBE et Cerege pour " Terre " et " Atmosphère ")
 - Acquisition via capteur environnementaux (ex : [htmnet](#), [emso](#), ocarina, cytomètre automatisé, [etc.](#))
 - Constitution d'un [portail d'accès](#) aux données environnementales de l'OSU Pytheas
- **SeadataCloud** : cytométrie en flux HF automatisée in situ, (capteurs autonomes) : résolution ensembles de classes du phyto et prise d'images : microphytoplancton en plus
- **SOERE** : ~~disparu ?~~



Ne concerne que les données
océanographiques



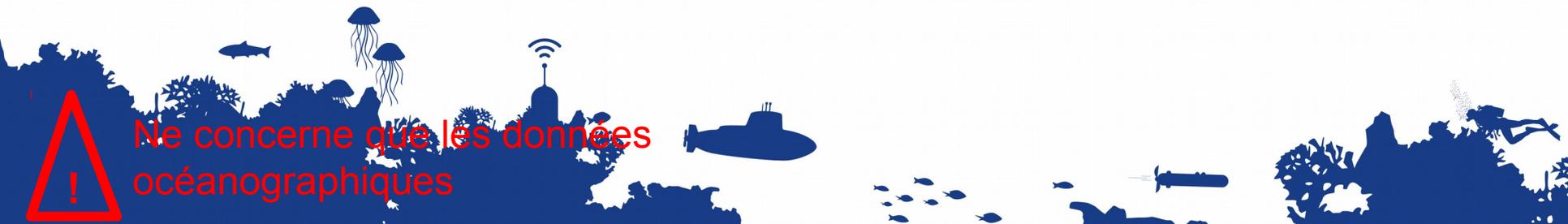
Qui est inclut dans le CDS ?

SeadataCloud : cytométrie en flux automatisée HF in situ (capteurs autonomes) : résolution ensemble de classes du phyto et prise d'images : microphytoplancton en plus + réflexion et mise en place d'un vocabulaire standard pour l'interopérabilité

- Ex : nom des groupes fonctionnels du phytoplancton résolu classes de tailles :pico-eucaryotes, nano-eucaryotes,
- Noms de genre : synéchococcus, prochlorococcus
- Identificateurs optiques : main total FLR caractéristiques optiques pour résoudre ces groupes fonctionnels



Ne concerne que les données océanographiques

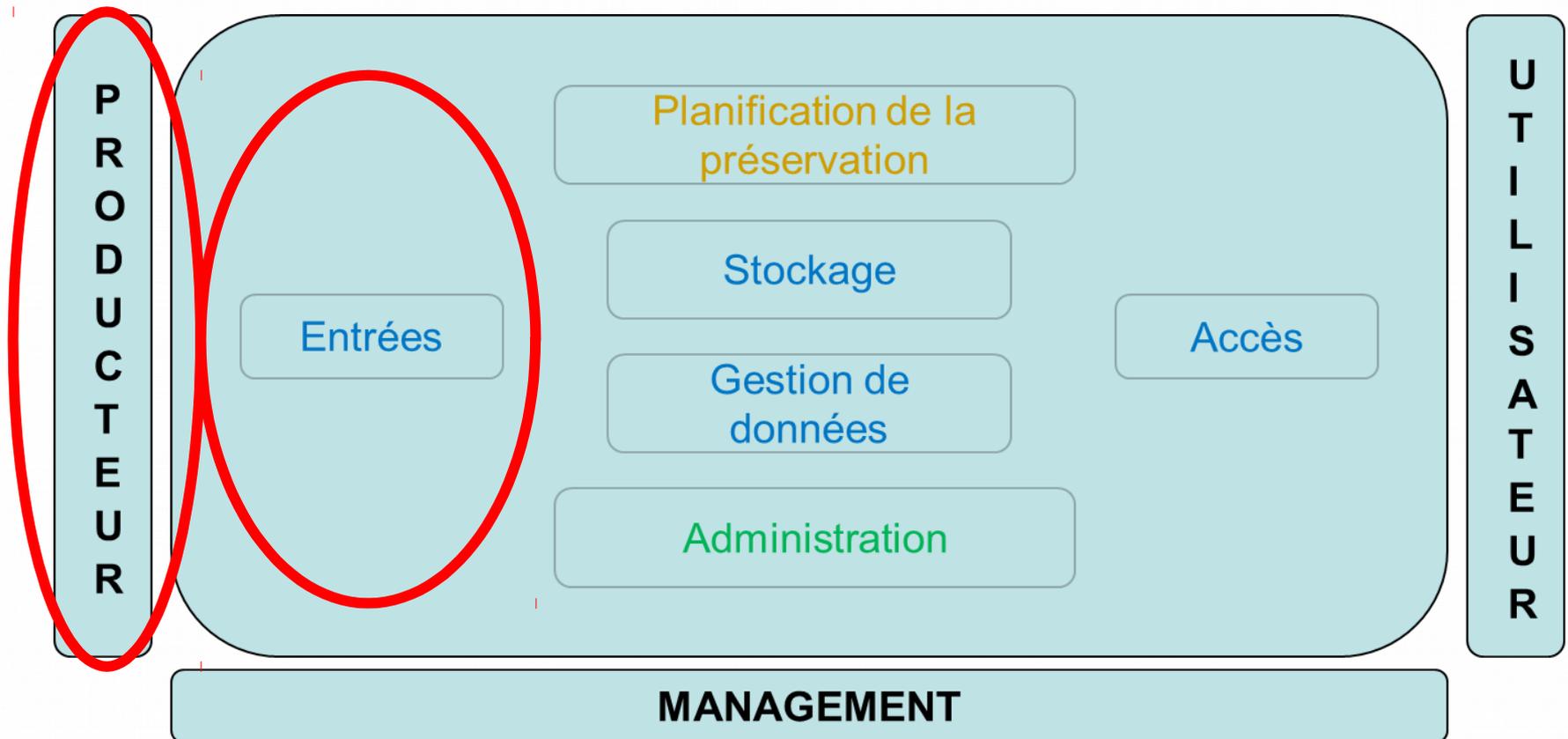




Modèle OAIS



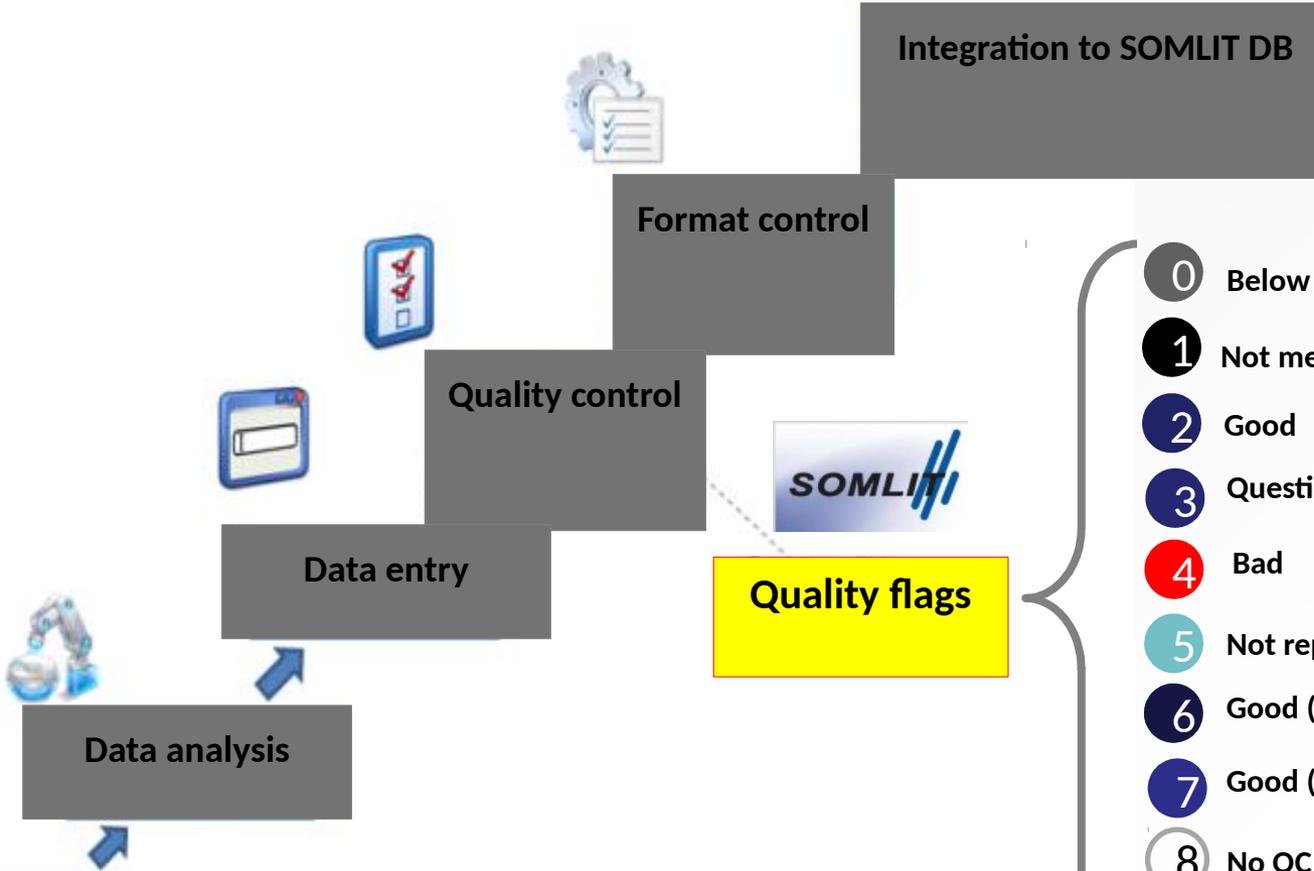
Modèle OAIS – Production & Entrées



Fonction « production » SNO

- Chaînes de traitement pour SNO somlit, moose, phytobs
 - Acquisition de terrain, collecte échantillons sur les sites [somlit marseillais](#), Analyse en labo, mise en fichiers
 - [Somlit HF](#) données brutes temps réel bouée SOMLIT Marseille
- Contrôle de cohérence des résultats
 - Le SNO fait un contrôle qualité (simer, sedoo). Seules les données qualifiées sont représentées
- Description des résultats : Sites WEB somlit, moose,
 - http://www.moose-network.fr/DATA_MOOSE/app/#/view1
 - <http://somlit.db.epoc.u-bordeaux1.fr/bdd.php>

SOMLIT Data management Workflow



- 0 Below detection limit
- 1 Not measured
- 2 Good
- 3 Questionable
- 4 Bad
- 5 Not reported
- 6 Good (mean of replicate measurements)
- 7 Good (not SOMLIT protocol)
- 8 No QC
- 9 Not sampled

Data Acquisition



Fonction « production » SO MIO

- Chaînes de traitement : équipes du M.I.O + service Obs OSU
 - Acquisition capteurs de terrain, collecte échantillons sur sites : [Drones marins ocarina](#), HF radars, [Stations météo htmnet](#), [Robot MII EMSO](#)
 - Analyse en labo, publication DOI
- Contrôle de cohérence des résultats : pas contrôle qualité systématique, datasets associés aux projets scientifiques
- Description des résultats
 - **Portail d'accès FAIR aux données env de l'OSU Pytheas**
 - Catalogage métadonnées [geonetwork](#)
 - Serveur de données cartographiques [geoserver](#)
 - Plateforme d'accès : [graphes et fichier ERDDAP](#)

Fonction « production »

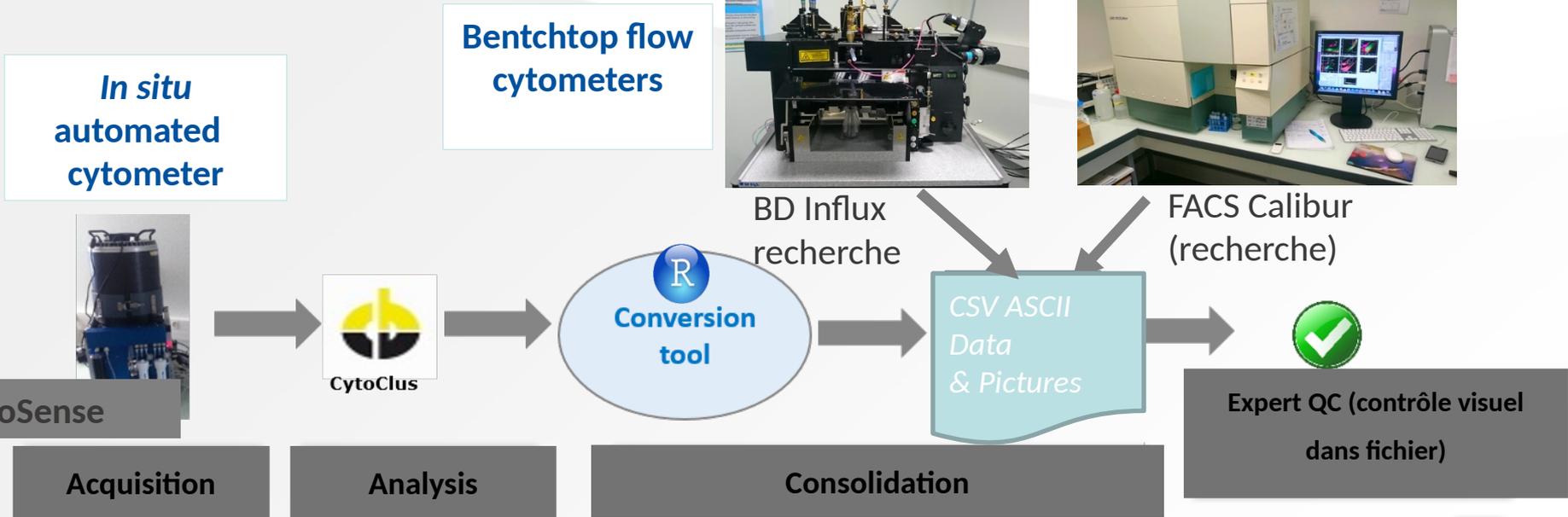
SeadataCloud



- **Chaînes de traitement**
- Cytosense, **Extraction R**, fichier CSV cyto, CQ, **Talend**, BD SQL cytobase, **extraction CDI XML**
- **Contrôle de cohérence des résultats**
 - Contrôle scientifique des données de cytométrie (tailles, propriétés optiques des cellules) du cytosense
 - Workflow SDC : CDI : Métadonnées décrivant chaque jeu de données avec les paramètres de mesures, position, date, point de contact, etc.
- **Description des résultats**
 - Portail SDC : <https://www.seadatanet.org/Metadata>
 - Localement **BD cytobase** <https://chrome.mio.univ-amu.fr/>

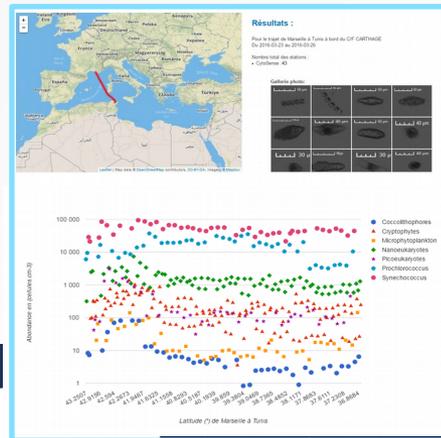


FCM Data management Workflow



SDN :
CDI XML

SDN :
ODV Data



TALEND
intégration

Accessability

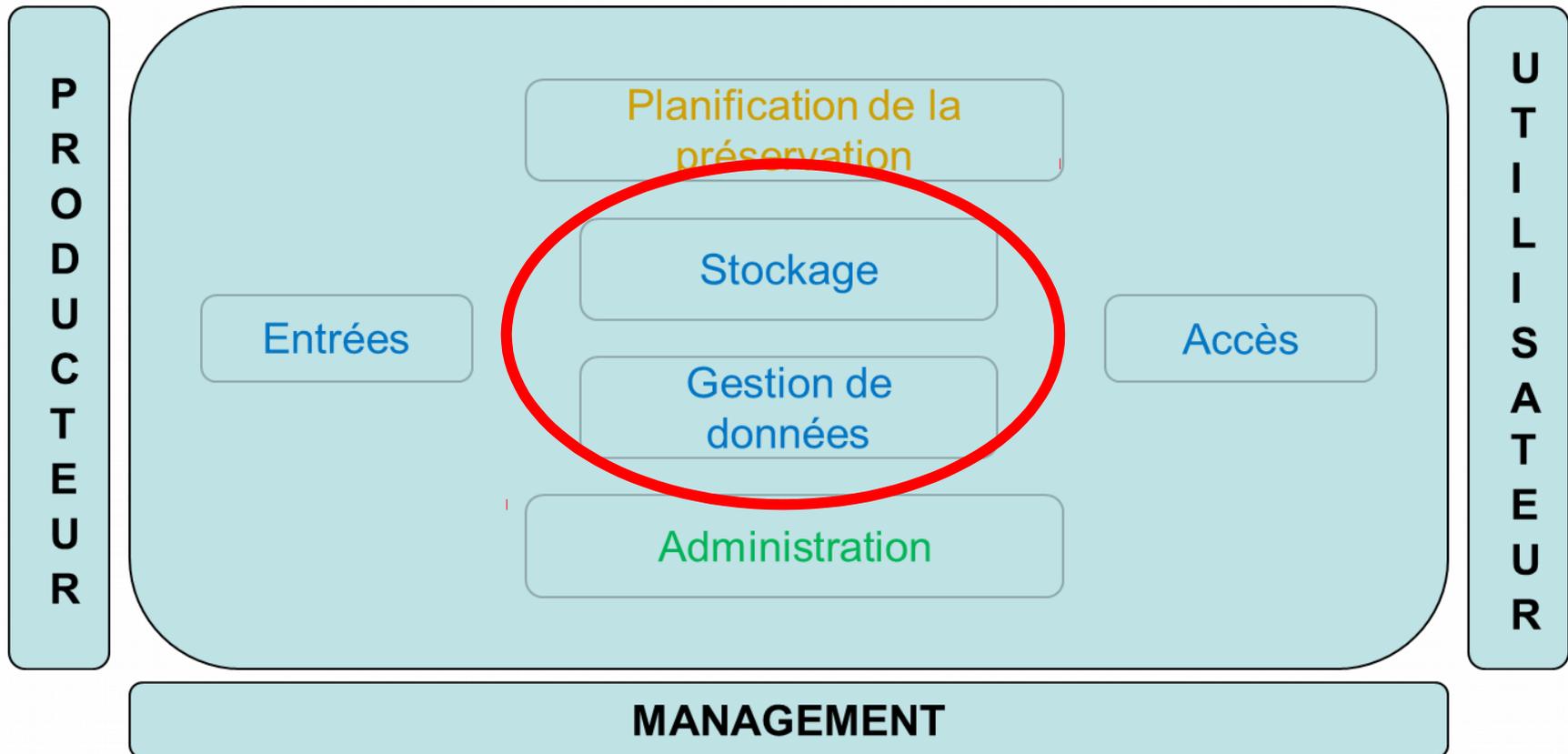


Integration

Fonction « entrées »

- **Harmonisation des données**
 - SNO : données " somlit ", " moose ", homogènes en BD unique
 - SO OSU : fichiers bruts, mise en format standard netCDF, ODV , CSV sur plateforme [ERDDAP](#)
 - SeadataNet :
 - Métadonnées produites (CDI XML) norme iso19115, données brutes au format ODV
- **Contrôle qualité**
 - SNO :
 - SDC cytométrie : contrôle visuel des résultats dans fichiers CSV effectuée par resp. scientifique de Cyto

Modèle OAIS – Stockage & Gestion des données



Fonction « stockage »

- **Conservation physique des données gérées**
 - SNO somlit, moose, Coast HF :
 - *externalisé à l'OASU (Somlit) et sedoo OMP (Moose)*
 - Cytométrie SDC et SO :
 - données d'observation locales
 - ***baie de disques netapp raid6 DP, redondées, sauvegardées***
 - ***Same serveur MIO/OSU Pytheas Bât. Oceanomed***

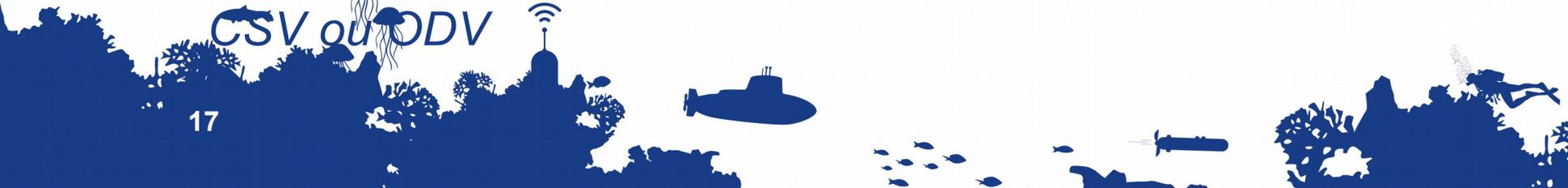
Fonction « gestion des données » ODATIS

- *Description des données gérées*

- **SDC cytométrie HF** : analyse qualitative et quantitative des cellules à l'échelle individuelle à haute fréquence (10000 cellules/sec, 1 ech /30mn : identification par propriétés optiques et tri physique taille, genre espèce phytoplancton, abondance
- **SO OSU** : données très variables, selon projet scientifique
 - https://erddap.osupytheas.fr/erddap/categorize/standard_name/index.html?page=1&itemsPerPage=1000
- **SNO MOOSE** : T° , $S\%$, O_2 , UV, nutriments NO_2 , NO_3 , NH_4 , PO_4 , $SiOH_4$; CO_2 , cytométrie flux, zooplancton, courants
- **SNO SOMLIT** : 
 - <http://sommelit-db.epoc.u-bordeaux1.fr/bdd.php>

Fonction « gestion des données »

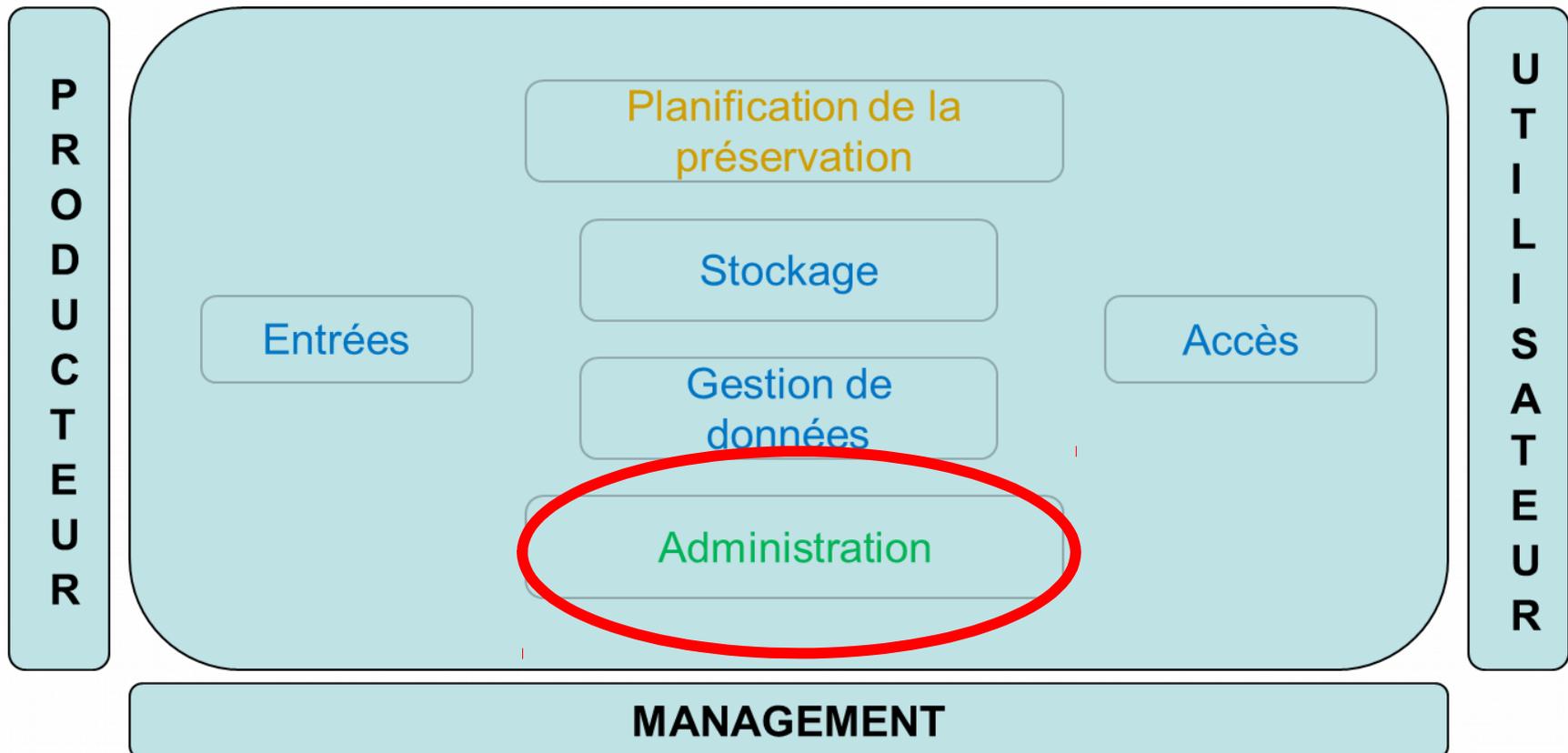
- **Description des résultats des chaînes de traitement**
 - **SNO** : cf somlit OASU et Moose Sedoo
 - **SDC Cytometrie** : cf plus haut
 - *Intégration des données cyto en BD SQL avec Talend (à refaire en python)*
 - *Génération de CDI en XML*
 - *Génération de data en ODV via le download manager*
- **SO OSU** :
 - *Traitement des fichiers de capteurs en python*
 - *Génération de fichiers de mesures propres en NetCDF, CSV ou ODV*



Fonction « gestion des données »

- *Description des résultats des chaînes de traitement*
- *Gestion des droits d'accès et des conditions d'accès*
 - SDC Cytométrie : accès aux données sur acceptation validation du P.I via le portail SDC
 - http://seadatanet.maris2.nl/v_cdi_v3/search.asp
 - SO OSU : plate forme Erddap accès public aux données
 - SNO MOOSE : ouvert sur demande
 - SNO SOMLIT : en BD accès via portail web sur demande

Modèle OAIS – Administration



Fonction « Administration » : SDC + SO



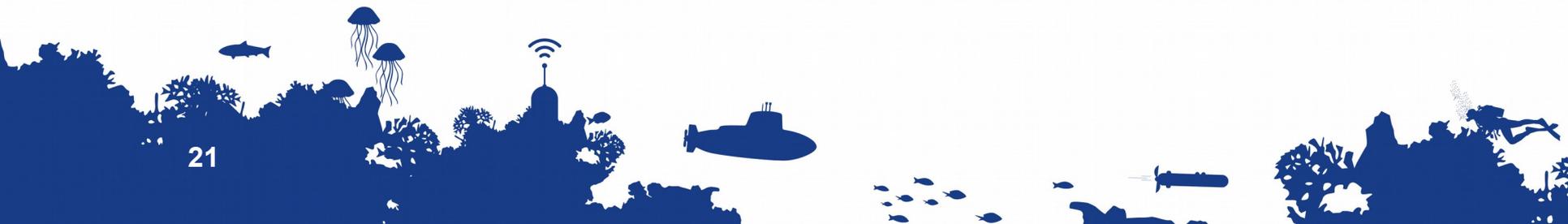
- **Moyens techniques informatiques**
 - Salle serveurs 40m2 du MIO Bât. Oceanomed Luminy
 - 2 climatisations, Onduleur 40kva
 - Baies de disques 100T redondées, sauvegardées
- **Maintien en conditions opérationnelles**
 - Equipe informatique stable de l'OSU Pytheas (SIP)
- **Maintien en configuration**
 - Service Observation OSU Pytheas :
 - ML 100 %; DM 50 %, PR 10 %
- **Passage en exploitation des logiciels** : ML 100 %



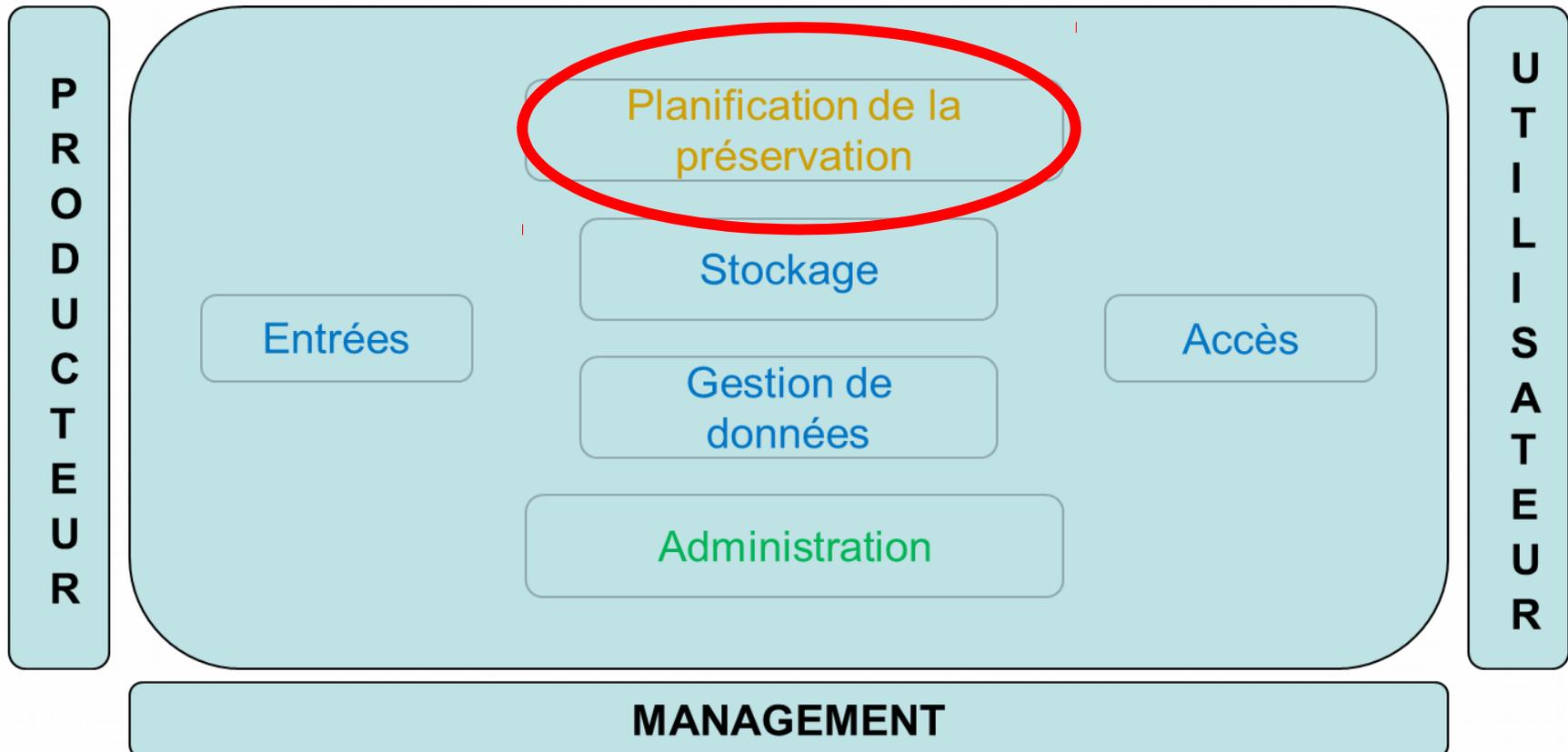
Fonction « Administration » : SDC + SO



- *Moyens techniques informatiques*
 - Plusieurs Serveurs de virtualisation proxmox
 - Machines virtuelles sauvegardées chaque jour
 - Reprise à " tiède " en cas de problème
 - 15mn le temps de restorer une sauvegarde de VM
- Présence de Helpdesk et assistance informatique opérationnels



Modèle OAIS – Planification de la pérennisation



Fonction « planification de la pérennisation »

- **Données et produits non reproductibles :**
 - SNO, SO, SDC : Toutes Données de capteur de terrain non reproductibles, acquises une fois. Nécessitent d'être stockées sauvegardées
- **Métadonnées**
 - SO OSU : Métadonnées Normalisées ISO19115 et Utilisation des « standards_name » de la convention NetCDF CF pour les fichiers de données
 - SDC Cyto : CDI XML basé sur ISO19115, INSPIRE
- **Description des formats, vocabulaires**
 - SDC Cyto : Vocabulaire contrôlé et en cours de définition pour la Cytométrie, pour créer des « standards_name »
 - SO OSU : Utilisation des « standards_name » de la convention NetCDF CF pour les fichiers de donnée

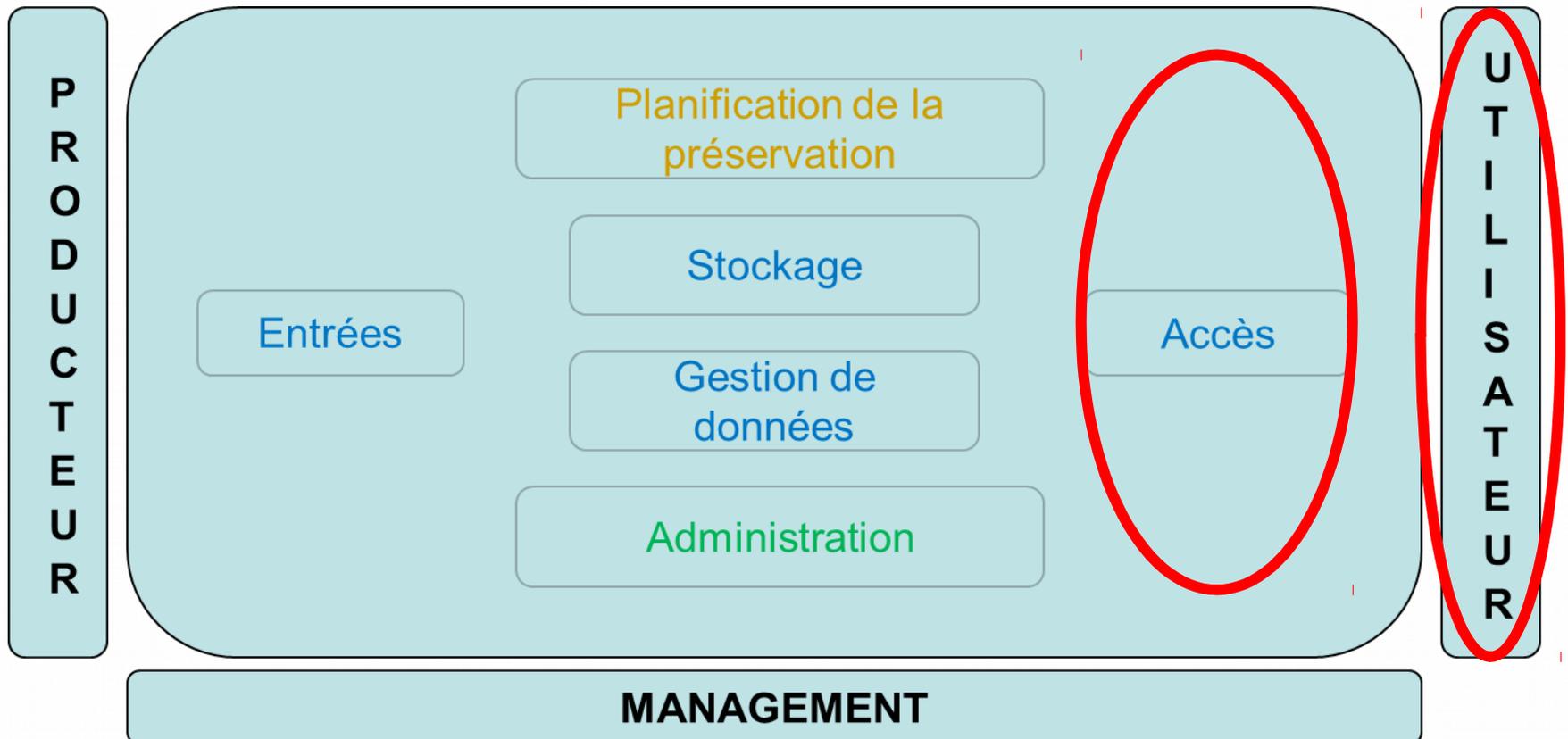
Fonction « planification de la pérennisation »



- **Logiciels :**
- SO OSU :
 - logiciels interopérables sur standards OGC : geonetwork, geoserver, erddap
 - Développements en LAMP : htmnet
- SDC Cyto : Ecosystème Logiciel Européen :
 - DM, NEMO, MIKADO
- SNO : développements spécifiques LAMP ? BD SQL ?



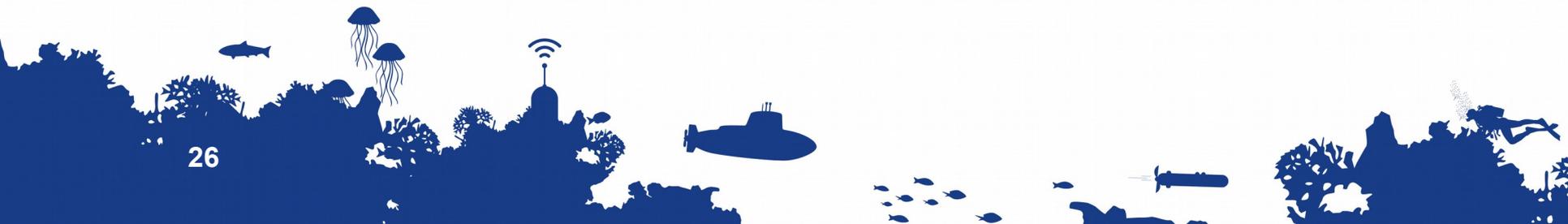
Modèle OAIS – Accès & Utilisateurs



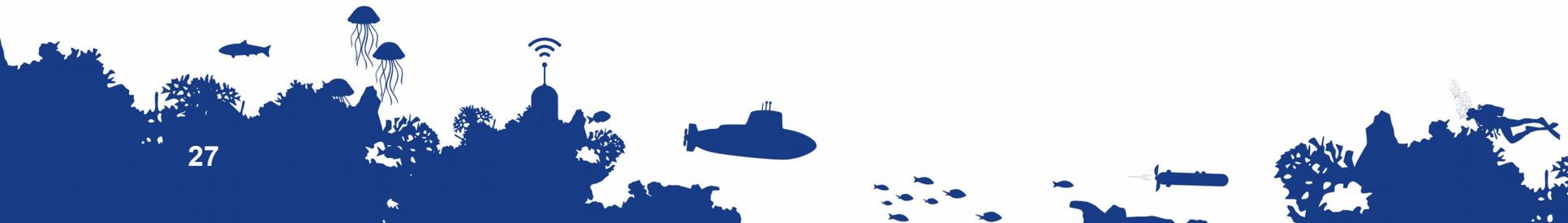
Fonction « Accès »

- **Contrôle d'accès**

- SDC Cyto : accès aux données sur demande et validation par le P.I via le portail SDC
http://seadatanet.maris2.nl/v_cdi_v3/search.asp
- SO OSU : actuellement accès public via portail
<http://erddap.osupytheas.fr/erddap>
 - Possibilité de protéger les dataset par login/passwd
- SNO : MOOSE, SOMLIT : cf procédures OASU et Sedoo



Fonction « Utilisateurs »





Besoins

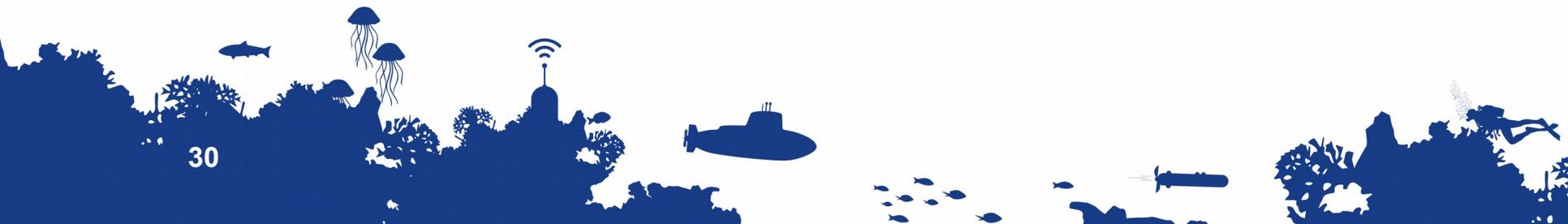


Besoins techniques

- ***Plan Matériel*** : À définir selon cahier des charges et objectifs projets : situation actuelle correcte
 - Pérennisation Baies de stockage ? Renouvellement régulier
 - Redondance serveurs ? Renouvellement régulier
- ***Plan Logiciel – chaîne de traitement***
 - Continuer l'élaboration d'un portail d'accès OSU Pytheas et de diffusion des données Cyto vers SDC
 - Amélioration de la chaîne de traitement de cytométrie en flux vers SDC ou vers Odatis (à préciser)
 - Recodage intégration données Cyto en BD SQL
 - Utilisation de NEMO/mikado Ifremer adapté à la cyto

Besoins techniques

- Les plus gros besoins sont sur le " manpower "
- Ouverture sur les évolutions des capteurs innovants du milieu marin (biodiversité, contaminants, métaux traces...)
- Veille technologique sur les capteurs nouvelles générations (cytométrie, hydro, biogéochimiques...)
- Se préparer à une arrivée et quantité de données HF in situ exponentielle





Merci de votre attention !

