

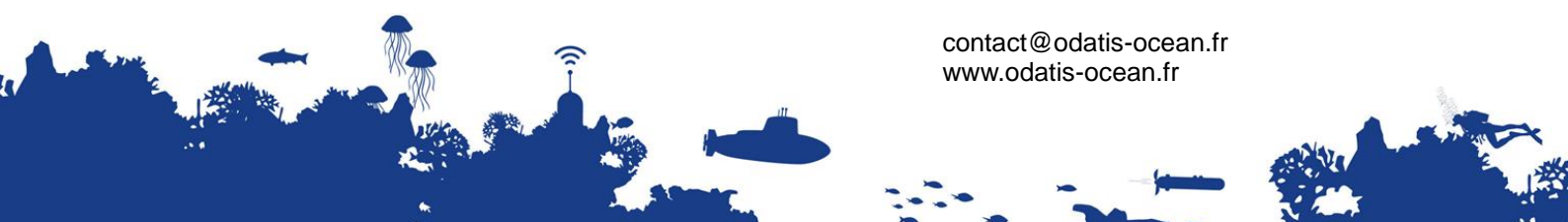


Compte Rendu de l'atelier technique ODATIS du 5 et 6 Juillet 2018

CR atelier technique 2018 #1



contact@odatis-ocean.fr
www.odatis-ocean.fr



Numéro du livrable	Titre court
	CR Atelier ODATIS 2018 #1
Titre long	
Compte rendu de l'atelier technique ODATIS du 5 et 6 Juillet 2018	
Description courte	
Auteur	Groupe de travail
Joël Sudre	
Dissémination	Copyright



Historique

Version	Auteurs	Date	Commentaires
0.1	Joël Sudre	09 Juillet 2018	Version initiale
0.2	Fabrice Mendes	15 Juillet 2018	Relecture et correction
0.3	Linn Sekund	23 juillet 2018	Relecture et correction

Table des matières

1. Accueil et tour de table des participants	4
2. Présentation générale sur ODATIS (Gilbert Maudire)	4
3. État des lieux des travaux du pôle (Linn Sekund)	6
4. Retour d'expérience depuis le 1 ^{er} atelier technique	7
4.1. SOMLIT (Fabrice Mendes)	7
4.2. LEFE CYBER (Catherine Schmechtig)	8
4.3. Cytométrie en flux (Michèle Fichaud – Soumaya Lhabib)	10
5. Les systèmes de visualisation	11
5.1. Outil de visualisation ERDDAP (Maurice Libes)	11
5.2. Outil de visualisation du SHOM (Valérie Cariou et Romain Lhullier)	12
5.3. Outil de visualisation Sextant (Mickael Tréguer)	13
5.4. Outil de visualisation CDS SAT-Brest (Jean-François Piollé)	15
5.5. Outil de visualisation Coriolis (Corentin Guyot)	15
6. Discussions	17
7. Conclusions et objectifs du prochain atelier	18

1. Accueil et tour de table des participants

Liste des participants à l'atelier ODATIS :

- ⑩ Aurélie Briand (Ifremer) – AB,
- ⑩ Valérie Cariou (SHOM) – VC,
- ⑩ Guillaume Charria (Ifremer – LOPS) – GC,
- ⑩ Gérald Dibarboure (CNES) – GD,
- ⑩ Michèle Fichaut (Ifremer) – MF,
- ⑩ Julie Gatti (Ifremer) – JG,
- ⑩ Corentin Guyot (Ifremer) – CG,
- ⑩ Mark Hoebeke (CNRS – SBR) – MH,
- ⑩ Romain Lhullier (SHOM) – RL,
- ⑩ Maurice Libes (CNRS - OSU Pytheas) – ML,
- ⑩ Gilbert Maudire (Ifremer) – GM,
- ⑩ Fabrice Mendes (CNRS – OASU) – FM,
- ⑩ Ornella Passafume (Ifremer) – OP,
- ⑩ Jean-François Piollé (ifremer – LOPS) – JFP,
- ⑩ Catherine Schmechtig (CNRS – OOV) – CS,
- ⑩ Linn Sekund (Ifremer) – LS,
- ⑩ Joël Sudre (CNRS -LEGOS) – JS,
- ⑩ Mickael Tréguer (Ifremer) – MT.



JS présente l'ordre du jour (*voir : 00_joel_sudre_agenda_atelier_odatis_2018_01.pdf*), en précisant que l'agenda a été modifié depuis son envoi lors du message d'invitation et que le sujet principal de ce second atelier technique ODATIS va porter sur les outils de visualisation utilisés et/ou développés dans les différents centres.

Tour de table des participants.

2. Présentation générale sur ODATIS (Gilbert Maudire)

(*Voir : 01a_gilbert_maudire_presentation_odatis.pdf*)

GM rappelle la genèse des pôles de données et de services (AERIS, Form@ter, ODATIS, Théia) ainsi que celle de l'Infrastructure de Recherche (IR) « système Terre » présidée par Frédéric Huynh (IRD). Il est à noter qu'un cinquième pôle de données doit se joindre aux 4 précédents dans les années à venir et qu'il se nomme le Pôle National des Données de Biodiversité (PNDB). Ce pôle fait suite au pôle ECOSCOPE.

La mission, l'organisation et la structure des pôles et de l'IR sont ensuite mentionnées. GM présente ensuite le pôle ODATIS en détaillant :

- ⑩ son organisation,
- ⑩ ses centres de données et services (CDS),
- ⑩ son calendrier 2018,
- ⑩ sa convention et sa feuille de route,
- ⑩ son périmètre des données à gérer,
- ⑩ ses aspects techniques,
- ⑩ ses services hiérarchisés.

Discussion suite à la présentation de GM :

[?] : Est-ce que les données modèles sont incluses dans le pôle ODATIS ?

[GM] : Certaines oui via Copernicus, et les DIAS (Data Information Access Services).

[GC] : Il va falloir être prudent avec les données de modèles car elles sont très/trop nombreuses, de qualités et de validations très différentes. Il va être difficile de savoir quelles données seront pertinentes ou pas. Qui va décider ?

[ML] : Où les données doivent-elles être qualifiées ?

[GM] : Les données doivent-elles être qualifiées en amont du pôle dans les centres d'observations (SNO, système d'observation, etc.).

[OP] : quel est la stratégie pour les référentiels communs ?

[GM] : il est obligatoire d'harmoniser le plus possible avec les référentiels internationaux. Cependant des référentiels propres aux pôles semblent tout de même nécessaires. Il va être nécessaire aussi de travailler en concertation avec les autres pôles pour gérer toutes les interfaces entre les pôles (ex : limites terrestres comme le littoral en utilisant les fonds géographiques de IGN, du BRGM). Il y a un travail important et nécessaire (voire

obligatoire) qui doit être effectué sur les référentiels.

[GM] informe les participants que l'IGN et le BRGM vont se joindre au partenariat dans ODATIS, ce qui va permettre d'exploiter leurs ressources ainsi que de profiter de leurs expériences sur le géoportail (voir : <https://www.geoportail.gouv.fr/>).

3. État des lieux des travaux du pôle (Linn Sekund)

LS présente le catalogue qui est installé sur le portail web du pôle ODATIS (voir : *01b_linn_sekund_catalogue_odatis.pdf*). Ce catalogue des données et des données dérivées du pôle ODATIS permet de découvrir les jeux de données fournis par la communauté nationale ainsi que leurs liens avec les programmes internationaux.

Ce catalogue est construit avec l'outil Sextant de l'Ifremer qui propose plusieurs outils de services d'accès aux données :

- ⑩ recherche avec filtres de sélection,
- ⑩ description des données,
- ⑩ service de visualisation,
- ⑩ service de téléchargement direct,
- ⑩ service de traitement en ligne,
- ⑩ service de téléchargement via un portail local des producteurs partenaires.

Sextant permet le catalogage, le stockage et la diffusion de l'information géographique. Il utilise les standards de l'OGC (Open Geospatial Consortium) pour les services associés aux métadonnées (CSW, WMS, WFS, WPS – <http://www.opengeospatial.org/standards>).

Sextant utilise GeoNetwork comme interface web pour la recherche de données géospaialement distribuées à travers le catalogue ODATIS. Cet interface permet à la fois de faire de la recherche texte mais aussi de la recherche par emprise géographique ou au moyen de facettes par mots clés. L'ensemble des fonctionnalités de Sextant sont disponibles sur le site ODATIS (http://www.odatis-ocean.fr/fileadmin/images/acces_donnees/Manuel_d_utilisation_catalogue_ODATIS_2018.pdf). Cette API peut-être intégrée dans tous les sites et elle est « customisable ».

Actuellement les fiches du catalogue ODATIS (<http://www.odatis-ocean.fr/>) sont compatible INSPIRE, et utilise la norme ISO 19115 pour la description et la norme ISO 19139 pour le format d'échange. Le filtre « Variables » se base sur les « Essentials Ocean Variables » (EOV).

LS présente ensuite les fonctionnalités de découverte, de visualisation et de téléchargement via une démonstration en ligne du catalogue ODATIS.

En ce qui concerne le site web ODATIS, il est actuellement en français uniquement mais la traduction du site est en cours (outil sextant permettant le multilinguisme). Un état des lieux est aussi en cours pour déterminer les sites



distants qui permettent la visualisation et l'accès aux données. Certains onglets sur le site vont être réorganisés pour différencier les services aux producteurs des services aux utilisateurs. Enfin un formulaire de dépôt de métadonnées est en création actuellement pour permettre d'insérer de nouvelles fiches dans le catalogue.

4. Retour d'expérience depuis le 1^{er} atelier technique

Suite au premier atelier technique ODATIS, certaines actions à mener sur les projets SOMLIT, LEFE CYBER et Cytométrie en flux avaient été identifiées. Cette section est consacrée au retour d'expérience sur ces trois projets et permet d'appréhender les avancées, les écueils et les tâches restants à effectuer.

4.1. SOMLIT (Fabrice Mendes)

Suite au premier atelier ODATIS, FM avait pris comme action de mettre en conformité la base de données (bdd) SOMLIT avec le catalogue SeaDataCloud pour avoir une meilleure interconnexion avec les autres bdd incluses dans le pôle ODATIS (voir : *02_fabrice_mendes_somlit.pdf*).

FM rappelle le contexte et le contenu de la bdd SOMLIT. Afin de faire un démonstrateur du workflow nécessaire pour la mise en conformité de la base, la zone d'Arcachon, contenant 3 sites où des CTD sont effectuées, a été sélectionnée. La période prise comme démonstrateur est l'année 2015. Les différentes étapes du projet ont été les suivantes :

- ⑩ Comprendre les notions, les catalogues, les outils et le workflow, CDI (Common Data Index),
- ⑩ Mettre à jour les fiches (organismes, navires),
- ⑩ Choisir les listes de vocabulaire pour les paramètres,
- ⑩ Analyser les champs nécessaires,
- ⑩ Tester Mikado,
- ⑩ Modifier la BDD,
- ⑩ Générer les CDI,
- ⑩ Faire valider les CDI,
- ⑩ Installer et configurer le DLM (Download Manager) qui permet de connecter SeaDataNet (SDN) à la bdd.

La prochaine étape sera de passer le DLM en production après avoir vérifié et validé les fiches de toutes les stations (et navires), de déclarer les fiches dans EDIOS ou EDMED, d'automatiser la procédure et de reprendre la documentation pour décrire cette nouvelle procédure. Enfin il faudra traiter de la même façon les séries de surface de SOMLIT.

Les principales difficultés rencontrées pour cette mise en conformité sont surtout aux niveaux des notions différentes qui sont véhiculés avec des noms identiques entre la base locale (SOMLIT) et SDN, les emprises géographiques, le nom des champs qui est peu clair dans SDN et le workflow SDN qui n'est pas évident à prendre en main. FM précise que les interactions avec MF et Flavian Gheorghe ont été très précieuses pour l'avancée de ce projet.

Il est donc nécessaire d'avoir un accompagnement des personnes qui désirent mettre en conformité leurs bases



avec des experts qui ont l'habitude de manipuler les outils, le workflow et les variables de SDN.

Discussion suite de la présentation de FM :

MF précise que le DLM va être bientôt remplacé par un « replication manager » pour toutes les données en accès non restreintes avec un stockage sur un cloud (notion de cache si souhaité). Le SHOM a été sélectionné pour être un bêta-testeur de ce « replication manager ».

4.2. LEFE CYBER (Catherine Schmechtig)

CS présente les avancées qui ont été faites depuis le premier atelier sur la base Lefe Cyber au niveau de la mise en place des DOI, du catalogue, de la visualisation et du formatage (*voir : 03_catherine_schmechtig_lefe_cyber.pdf*). Pour cela un protocole sur un cas d'étude a été mis en place sur deux jeux de données de campagnes à la mer.

En ce qui concerne la mise en place des DOI, CS a travaillé en collaboration avec Frédéric Merceur de SEANOE et les PI des deux projets pris en cas d'étude (OUTPACE avec Karine Blanc et BIOSOPE avec Bernard Quéguiner). Bien que cette tâche est bien avancée, il reste cependant à préciser la granularité des DOI, le format des données finales (NetCDF, ODV-SDN) et où vont-être stockées les données.

Après installation et utilisation du logiciel MIKADO, environ 1/4 du dictionnaire Cyber a été transcodé dans la base SDN. Des tests sont en cours pour installer un Geonetwork afin de moissonner le catalogue de la base Lefe Cyber par l'IDS SEXTANT mis en place par le pôle ODATIS.

Pour les aspects visualisations, l'installation d'un serveur ERDDAP est en veille actuellement bien que des tests sur un PC perso aient été concluants car il est nécessaire de finaliser le formatage des données en amont et de le rendre aussi conforme au format accepté par ERDDAP.

Discussion suite de la présentation de CS :

[CS] Que fait-on des données ? Quel est le protocole à suivre pour arriver à un dépôt final après la mise en place du DOI ? Où vont-être stocké les jeux de données après le DOI ?

[GM] : Il est nécessaire que les CDS, qui vont accueillir les données, soient certifiés RDA coretrustseal car cela devient une exigence obligatoire au niveau européen afin de garantir la pérennité des données. En effet il devient obligatoire dans les DMP (Data Management Plan) de demandes de financement européen de mentionner un centre certifié qui va in fine récupérer les données et les mettre à disposition. Il est de ce fait nécessaire qu'une action globale au niveau du pôle ODATIS soit effectuée pour cette certification et pour tous les CDS. Cette certification portera sur une organisation en CDS distribués ayant eu une certification par le pôle ODATIS et non pas sur un centre particulier. Cependant cette action risque de prendre du temps car les contraintes pour obtenir la certification sont nombreuses (*voir: <https://www.coretrustseal.org/why-certification/requirements/>*). Il est



donc nécessaire d'avoir une solution de transition le plus rapidement possible. Pour ce faire, le CDS Brest et SEANOE sont actuellement en train de demander cette certification. Ceci va permettre (tout du moins dans un premier temps) d'avoir l'assurance d'être certifié en demandant un DOI et en déposant les données à SEANOE.

Toujours au sujet de la certification des échanges sont aussi effectués entre le pôle ODATIS et le CINES pour insérer le CINES dans cette démarche.

En ce qui concerne le lien entre SEANOE et SDN (ou SDC - Sea Data Cloud), il va être nécessaire de définir aussi un protocole pour la DOIification et pour la mise en place de licence publique (Cco, CC-BY, CC-BY-NC), afin de ne pas avoir de redondance entre les DOI ou licences mises par SEANOE et par SDC.

État des bases avec le transcodage vers SDN :

- ⑩ Phytobs : Format ODV avec vocabulaire SDN,
- ⑩ Pelagos : tous les paramètres sont déjà présents dans SDN,
- ⑩ Quadrige : la base est toute transcodée vers SDN,
- ⑩ SOMLIT : en cours avec la nécessité de faire un effort sur les formats de sorties cependant toute la chimie du SOMLIT est déjà présent dans le vocabulaire SDN. Le véritable écueil est que le transcodage dans le vocabulaire SDN est très consommateur de temps,
- ⑩ Geotrace, le transcodage n'est pas évident car le vocabulaire n'est pas toujours présent dans SDN et il est nécessaire de mieux structurer les données.

Pour les données de génomique, le travail doit être entrepris en collaboration avec le BODC (British Oceanographic Data Centre) et le PNDB (Pôle National de Données de Biodiversité). Il est à noter que cet effort sera fait dans un ou deux ans.

Enfin en ce qui concerne les interfaces (Océan-Atmosphère par exemple), il va être nécessaire de mettre en place des GT (groupe de travail) au niveau d'Interpôles qui est le lieu le plus approprié pour ce type d'étude.

4.3. Cytométrie en flux (Michèle Fichaud – Soumaya Lhabib)

SL ayant terminé son contrat au MIO, MF s'est proposée pour rendre compte du travail accompli sur les données de cytométrie en flux (FCM) – (voir : *04_michele_fichaut_soumaya_lhabib_cytometrie_mio.pdf*). Ce travail collaboratif (MIO, OSU Pytheas, BODC, IFREMER, MARIS) a permis de mettre en place un système de gestion des données de FCM interopérable avec l'infrastructure Européenne SDN et en cohérence avec les standards et normes internationaux. Pour ce faire, un vocabulaire commun pour la FCM a été établi et mis en place dans SDN ainsi qu'un protocole de gestion de nouveau jeu de données qui permet d'intégrer les données FCM avec le système SDN.

Le vocabulaire a été établi par l'analyse de la bibliographie se rapportant à la FCM puis par l'envoi d'un questionnaire à 180 utilisateurs de la FCM de par le monde. Les retours au questionnaire ont permis intégrer dans SDN :



- ⑩ les définitions et les noms de groupes de la FCM,
- ⑩ les métadonnées de la FCM,
- ⑩ les métadonnées des méthodes d'échantillonnage de la FCM,
- ⑩ les données de FCM,
- ⑩ un nouveau format de fichier adapté à la FCM (importable dans ODV).

Au final le vocabulaire de SDN s'est enrichi de 26 termes et de 44 codes portant sur 5 listes différentes. La base de FCM basée au MIO a ensuite été interconnectée via Mikado et le DLM à SDN.

5. Les systèmes de visualisation

5.1. Outil de visualisation ERDDAP (Maurice Libes)

ML présente l'outil et les fonctionnalités de ERDDAP (Environmental Research Division Data Access Protocol) qui est un serveur de données permettant de mettre à disposition de manière assez simple et consistante des données scientifiques grillées ou tabulaires. ERDDAP permet aussi de créer des graphes et des cartes de ces données (*voir : 05_maurice_libes_erddap.pdf*). Ses objectifs sont de faciliter l'accès, la disponibilité et l'utilisation des données scientifiques dans des formats courants. ERDDAP est un projet libre avec un code source ouvert, développé par Bob Simons de la NOAA (National Oceanographic and Atmospheric Administration) aux États-Unis d'Amérique.

ERDDAP fournit aux utilisateurs une interface unique pour chercher des données dans des fichiers locaux ou distants. Il permet aussi d'établir des requêtes et d'obtenir des données ou des sous-ensembles de données dans divers formats. ERDDAP permet d'enrichir les métadonnées de chaque jeu de données, de les générer au format ISO 19115 et de standardiser les dates dans les fichiers.

ML présente l'ensemble des formats d'entrée puis les fonctionnalités d'extraction, de visualisation, d'agrégation de jeux de données, d'authentification que permet ERDDAP.

La liste ci-dessous reprend les avantages d'ERDDAP :

- ⑩ nombreuses fonctionnalités,
- ⑩ produit complet pour diffuser la donnée,
- ⑩ traite un très grand nombre de formats d'entrée et de sortie,
- ⑩ interface de requêtes,
- ⑩ graphiques et sorties WMS,
- ⑩ documentation très étoffée,
- ⑩ disponibilité de l'auteur,
- ⑩ Robuste à la charge (en comparaison avec Thredds).

Et les inconvénients :

- ⑩ interface pas très « glamour »,
- ⑩ en cas d'erreur les logs sont très fouillis et l'erreur est difficile à trouver,
- ⑩ nécessite de souvent adapter le XML à la main ce qui nécessite une bonne connaissance de l'outil,



- ⑩ la pérennité du logiciel et très dépendante de l'auteur.

ML fait ensuite une démonstration en ligne d'ERDDAP mis en place au MIO.

Discussion suite de la présentation de ML :

Cet outil est bien adapté à un déploiement dans certains CDS et SDN (pour le Virtuel Research Environment VRE) s'oriente aussi vers ERDDAP, il est donc intéressant d'investir du temps pour répondre à ces questions :

Est-ce qu'ERDDAP peut permettre l'authentification par une fédération d'identité ? (couplage avec shibboleth/OpenID etc.)

Est-ce qu'il y a une limitation de taille des jeux de données ?

Pour les gros volumes de données, quel est la tenue à la charge ?

Est-ce que l'extraction peut-être parallélisée ?

Est-ce que les données satellitales sont visualisables rapidement ? (tests sur le temps de réponse pour visualiser un jeu volumineux ou une donnée volumineuse).

Comment sont gérer les « flags » sur les données ?

5.2. Outil de visualisation du SHOM (Valérie Cariou et Romain Lhullier)

VC et RL présentent le portail de l'information géographique maritime et littorale qui permet d'accéder aux données de référence du SHOM (voir : *06_valerie_cariou_romain_lhullier_shom.pdf*).

Les thématiques abordées sur ce site sont la bathymétrie, la marée, la cartographie, la salinité, les courants, la température, etc. Le site permet la sélection de données, la visualisation multicouches, l'affichage d'océanogramme. On peut aussi accéder par script aux données.

Dans sa première version l'outil était basé sur SOS mais il a été observé des problèmes pour la gestion des historiques (données volumineuses, lenteur d'affichage voire impossibilité, etc.). Mais aussi de nombreux problèmes pour supprimer les données dans SOS. Les outils comme istSOS et 52°N ont donc été abandonnés par le SHOM au profit d'un développement d'un outil sur mesure. Le développement et la gestion du logiciel a donc été externalisé chez un sous-traitant qui a redéveloppé un code permettant de mettre en ligne l'historique. Ce projet est actuellement en cours. Cette externalisation pose tout de même des problèmes pour la mise en œuvre de changement dans le logiciel.

Les données peuvent être extraites dans différent formats : XML, txt, Json. La description des marégraphes est en SML (sensor model langage). Les DOI placés sur les données marégraphiques sont positionnées par année avec un fragment pour chaque station.

L'outil datashom permet l'interfaçage direct avec le site ODATIS au moyen de services (WMS), les métadonnées sont au format standard geoserver. Les capteurs sont eux décrits au format sensor ML ce qui offre une très bonne description des capteurs mais aussi de la description des événements du capteur au cours de son fonctionnement (description des anomalies, arrêt, etc.). Cette norme est bien pensée en termes de métadonnées et pour de petits



jeux de données. Cependant actuellement elle a un souci pour les gros jeux de données car elle multiplie par 10 la taille du jeu ce qui a pour conséquence de ralentir énormément la machine.

Le SHOM projette de mettre les données historiques en ligne, car actuellement seule les données les plus récentes sont mises en ligne avec une fenêtre temporelle de quelques jours (15 jours par exemple pour les données de RADAF HF qui sont actualisé toutes les 10 minutes). Cette limitation est actuellement voulue pour permettre une visualisation rapide du jeu de données considéré. De la même façon pour les visualisations des données temporelles le nombre de points affichés sur un graphe est limité à environ 1000 pts (ré-échantillonnage à la volée lors de la création du graphe en fonction de la période temporelle choisie par l'utilisateur).

RL termine la présentation par une démonstration en ligne de l'outil datashom.

Il est à noter que ce produit est sous-traité, ce qui peut entraîner des latences pour la mise en place de correctifs ou de modification de version.

Discussion suite de la présentation de VC et RL :

La discussion s'est centrée sur l'utilisation de la norme SML pour la description des capteurs. Cette norme pourrait parfaitement convenir pour la description d'une campagne, d'un observatoire câblé, d'une ligne de mouillage etc. Cependant le principal défaut de cette norme est qu'elle multiplie la taille du jeu de donnée par des métadonnées de façon très conséquente. Cette norme en contrepartie gère parfaitement tout l'événementiel du ou des capteurs décrits.

5.3. Outil de visualisation Sextant (Mickael Tréguer)

MT présente l'infrastructure de donnée spatiale (IDS) SEXTANT qui est composée de divers modules (voir : *07_mickael_treguer_sextant.pdf*) :

- ⑩ un catalogue,
- ⑩ un visualiseur,
- ⑩ une API,
- ⑩ des services OGC,
- ⑩ un panier/extracteur de données,
- ⑩ un système d'administration.

L'IDS Sextant est basé sur l'utilisation des standards de l'OGC construits autour de la métadonnée (CSW) comme la visualisation avec WMTS, WMS, le traitement (WPS), le téléchargement (WCS, WFS), etc.

L'outil Sextant est construit à partir de logiciel libre en mode contributif (Geonetwork, mapserver, Qgis-server, GeowebCache, Thredds/ncWMS, Oceanotron). L'outil possède de nombreuses fonctionnalités qui sont décrites dans la présentation :

- ⑩ filtre sur les données attributaires,
- ⑩ moissonnage des données via le protocole WFS,
- ⑩ moteur d'indexation Elasticsearch,
- ⑩ requête WMS-FE,



- ⑩ affichage de cartes,
- ⑩ affichage de données multidimensionnelles,
- ⑩ traitement de données en ligne,
- ⑩ accès aux données d'observation sur la colonne d'eau (profils, trajectoires, série de points, etc.).

Le groupe de développement de l'IDS Sextant essaye de tirer parti des avantages de chaque norme pour améliorer l'interprétation et l'analyse des données par l'utilisateur. Il est important de noter, que l'IDS Sextant peut être téléchargée, déployée et utilisée sur d'autres sites que l'Ifremer.

Discussion suite de la présentation de MT :

La discussion a surtout porté sur les DIAS et le traitement à distance de données de gros volumes via des technologies de cloud, notebook etc. Actuellement toutes ces technologies sont à l'étude et il est nécessaire d'attendre encore 1 ou 2 ans pour voir apparaître des solutions pouvant être utilisées pour les données satellitaires. Il est à noter qu'il faut bien séparer les données de types observations in-situ et les données de type satellitaires (ou de modèle synthétiques) car elles ont des contraintes et des traitements très différents. L'effort actuel des ateliers doit donc actuellement porter sur les données in-situ.

5.4. Outil de visualisation CDS SAT-Brest (Jean-François Piollé)

JFP présente plusieurs logiciels qui ont été développés pour les données satellitaires au CDS SAT Brest (CERSAT - voir : [08_jean_francois_piolle_cds_sat_brest.pdf](#)).

Cette collection de logiciels comprend :

- ⑩ un logiciel de visualisation: Syntool (<http://syntool.ifremer.fr>); Syntool permet de faire de l'analyse visuelle, de la découverte sur des données satellitaires de niveau 1 et 2, de la colocalisation « basique ». Il travaille sur des images statiques pré-générées et des données in-situ formaté en Json. Ce logiciel est développé et maintenu par OceanDataLab. Cependant le processus d'ingestion des données est fastidieux et nécessite un stockage important. De plus il est vite limitant en termes d'analyse scientifique.
- ⑩ un logiciel permettant de faire du travail à distance sur les données satellitaires : Jupyter (<http://jupyter.org>). Jupyter est développé en Python et permet l'analyse de données, leurs combinaisons, le partage de résultats, et d'ingérer du code pour traiter les données dans 40 langages différents. Jupyter est complété par Jupyterhub qui est un serveur de notebook permettant d'avoir une interface personnalisable pour l'utilisateur.
- ⑩ un middleware en python : Cerbere (<http://cerbere.readthedocs.io/en/latest/>). Cerbere est une API python dédiée à la lecture, l'interprétation et l'écriture de données satellitaires.
- ⑩ un logiciel permettant de faire des match-up et de qualifier la donnée satellitaire : Felyx
- ⑩ un logiciel permettant de faire de la détection et de l'analyse croisée entre des données satellitaires tuilées : Naiad (<http://naiad.readthedocs.io/en/develop/>). Ce logiciel repose sur Elasticsearch et permet de traiter des données de gros volumes (SST à 1 km de résolution, etc.). Il permet aussi de faire de l'analyse de ces



données.

- ⑩ un logiciel permettant de faire de l'analyse sur du « big data » : NEXUS

Discussions sur HADOOP et les technologies du «big-data» qui peuvent découper, dupliquer, distribuer et traiter les données en lots (« chunks »).

5.5. Outil de visualisation Coriolis (Corentin Guyot)

CG présente les outils de visualisation du portail Coriolis (*voir : 09_corentin_guyot_coriolis.pdf*) dont l'objectif initial est d'être une solution permettant la découverte, la visualisation et la diffusion de donnée « big data ». Cet outil doit gérer toutes les données de la base Coriolis. Son processus d'importation est basé sur ElasticSearch (data et metadata). La configuration de l'API se fait via une base de données PostgreSQL (<http://data.coriolis-cotier.org>). Ce site permet de visualiser la position des données, de les visualiser sous forme de séries temporelles (courbe avec 1000 points d'affichage maximum). Ce site a été décliné aussi pour Euro Argo (<http://co-discovery-demo.ifremer.fr/euro-argo/>) et pour Coriolis Global (<http://co-discovery-demo.ifremer.fr/coriolis/>). Au vue du grand nombre d'observation pour Coriolis Global (5 milliards d'observations), des évolutions sont en développement et en test pour passer d'une base Oracle à une base Cassandra. Cette nouvelle base a en effet plus de robustesse, gère une quantité massive de données sur plusieurs centres. À noter que d'autres types de base sont aussi à l'étude (Impala, Hbase, Etc.). Une évolution sur l'ergonomie et l'architecture de la configuration de l'API est aussi en cours.

Discussion suite de la présentation de CG :

[MH] : Comment sont gérées les droits des fichiers d'extraction ?

[CG] : Les fichiers sont extraits et déposé sur un espace dédié puis les droits sont changés avec le login du client qui a effectué la requête.

Précisions de GM :

En ce qui concerne les bases, elles sont mises à jour tous les 24 h. Actuellement il y a de nombreuses duplications des bases (Oracle, ElasticSearch, etc.).

Pour Euro-Argo, la chaîne opérationnelle est la suivante :

- ⑩ récupération des trames Argos de tous les flotteurs,
- ⑩ reconstruction des profils (bdd Oracle qui resynchronise les profils, les données structurées en objet métier, contrôle de la validité du profil et des données manquantes, retour vers les PI),
- ⑩ extraction des profils en fichier netCDF et mise en place du DOI pour le Temps Réel
- ⑩ fichier NetCDF déposé sur le ftp
- ⑩ synchronisation avec les autres centres (pour obtenir les trames manquantes etc.),
- ⑩ au bout de 24h le fichier doit-être commun entre tous les centres,
- ⑩ diffusion et mise à disposition du fichier,
- ⑩ calcul de résidus a posteriori,



- ⑩ mise en place d'un deuxième DOI pour le Temps Différé (mensuel), DOI sur la série temporelle complète, DOI avec fragment mensuel.

La mise en place d'un DOI sur les requêtes est en cours d'étude mais pas encore au point (historisation des fichiers, lourd en terme de stockage, etc.). Les différentes versions des fichiers doit-être conservées car les données sont utilisées comme référence par les autorités.

6. Discussions

Discussion sur les séries temporelles :

[ML] : Les implémentations de serveurs de données de type OGC SOS comme istSOS (<http://www.istsos.org/index.html>) ou 52°North (<https://52north.org/>) devrait être étudiées pour les séries temporelles.

[RL] : Au SHOM, nous avons testé ce type de serveurs est les temps de réponse étaient trop longs dès lors que le nombre de données était élevé.

[MH] : En ce qui concerne sensor ML pour la biologie et les taxons, il n'y a pas de possibilité pour visualiser les occurrences

[GM] : Il y a deux niveaux de visualisation à considérer :

- ⑩ la géométrie des observations,
- ⑩ les résultats de l'observation (profils, séries temporelles, indice d'abondance, etc.).

Pour ODATIS, le catalogue commun pour découvrir la donnée est quasiment acquis. Il a nécessité beaucoup d'effort et d'investissement mais il est en très bonne voie. L'idée est maintenant de connecter des services de visualisation pour mieux appréhender les données (sous forme de quicklook?).

À l'ordre zéro nous devons avoir un minimum d'harmonisation avec un même niveau d'exigence pour cela il faut travailler à la fois pour optimiser la visualisation de jeux de données centralisée mais aussi sur des jeux de données décentralisés avec la mise en place de protocoles en arrière-plan voire des caches techniques. Il faut en effet dans un premier temps bâtir les services de bases, mettre ensuite en place des protocoles communs entre les divers CDS, puis bâtir des services de hauts niveaux.

Dans ce cadre, il semble qu'un déploiement massif d'ERDDAP dans les CDS soit une solution optimale pour factoriser les efforts. Il faudra tout de même avoir une attention particulière sur les contraintes de sécurité dans les CDS.

Un DOI fixe un jeu de données et doit permettre de faire un rejeu lors d'extraction ultérieure du jeu à l'identique (ce qui implique qu'il faut garantir que les données n'aient pas changé surtout pour le temps différé). De plus le producteur de DOI se porte garant de la pérennité des données. Cependant dans certains cas, comme Sextant, l'attribution se fait en faisant confiance aux producteurs de données. Il y a aussi la responsabilité inverse où il faut



garantir la disponibilité de la donnée. Il y a donc une responsabilité partagée qui doit être établie par convention entre producteur et distributeur.

Actuellement SEANOE permet de mettre un DOI uniquement si le jeu de données est téléversé sur le site. D'ici la fin de l'année, il sera possible d'obtenir un DOI à SEANOE avec la possibilité de téléverser les données dans un CDS (qui prendra en charge la garantie de pérennité). À noter que chaque CDS devra au préalable avoir l'aval d'ODATIS pour avoir la qualification.

ODATIS doit statuer sur les organismes permettant de mettre en place les DOI via son CODIR. ODATIS doit aussi mettre en place une convention de pôle pour obtenir la certification Coretrustseal pour cela il doit trouver pour les données in-situ un deuxième centre pouvant garantir la sécurisation et la pérennisation des données. Le CNES pourrait être un bon candidat en attendant que le CINES puisse répondre positivement. Il manque actuellement de directives institutionnelles claires (qui appuieraient les recommandations de l'interpôles). Ces recommandations ont été remontées au niveau de l'IR « système terre » pour cela.

Actuellement les fournisseurs de DOI sont :

- ⑩ SEANOE au coup par coup pour les produits dérivés sur lesquels les scientifiques désirent placer un DOI,
- ⑩ SEANOE pour chaque campagne à la mer car cela est obligatoire pour les besoins d'évaluation de la TGIR Flotte,
- ⑩ Les produits de nature cartographique dans sextant peuvent aussi obtenir un DOI.

Pour ces sites de l'IFREMER où l'on peut obtenir un DOI il y a un lien entre.

Des DOI peuvent aussi être obtenus par le SEDOO.

7. Conclusions et objectifs du prochain atelier

GM fait la synthèse des objectifs de ce second atelier ODATIS.

Nous avons plusieurs objectifs :

- ⑩ offrir à court terme des services de découverte de la donnée avec une prévisualisation ; Il est donc nécessaire de compléter les métadonnées du portail,
- ⑩ mettre au-dessus des CDS une couche d'interopérabilité qui permet de faire de la visualisation de la métadonnée, du téléchargement etc. par la mise en place de protocoles normalisés et potentiellement de visualiser la donnée,
- ⑩ permettre de faire de la covisualisation multi-sources, etc. (services avancés)

Il est nécessaire de dissocier l'aspect satellite des aspects in situ. Pour l'aspect in situ, plusieurs protocoles sont candidats, cependant ERDDAP étant un candidat sérieux, il semble opportun d'approfondir cet aspect (accès aux données, organisations des données, etc.). La visualisation de séries temporelles ou d'objet métier nécessite aussi d'approfondir les protocoles SOS et les outils tels qu'istSOS, 52°North qui répondent aussi à ce besoin.



Pour ce faire le prochain atelier va se concentrer sur ces deux aspects :

- ⑩ 1^{er} journée sur ERDDAP,
- ⑩ 2^{ème} journée sur les protocoles SOS (outils istSOS, 52°North,...).

Afin de préparer ce troisième atelier il est nécessaire de constituer une base de données commune qui va être utilisée par les outils que l'on désire tester et évaluer. Pour ce faire, il est nécessaire d'extraire par les participants des fichiers de leurs bases (quelques fichiers exemples) et/ou pour les bases de données postGRES un sous ensemble de la base.

L'atelier permettra de dérouler des opérations à faire et de mettre une procédure à suivre pour les CDS désirant de mettre en place ce service.

A noter que la mise en place d'ERDDAP, nécessite l'administration d'un serveur apache Tomcat. La phase de test et de mise en place d'un serveur ERDDAP vont nécessiter de redémarrer à plusieurs reprise le serveur apache. Pour cela, il est fortement conseillée que les administrateurs systèmes soit aussi conviés à l'atelier car dans certains CDS, l'administration système et la gestion des données scientifiques et des outils associés sont complètement dissociés).

Pour les données satellitaires le même exercice est à faire (AVISO, CDS couleur ?, LOPS (AB et JFP)).

La base de fichier commun doit contenir à minima :

- ⑩ un sous-ensemble d'une-base en MySQL,
- ⑩ des fichiers netCDF grillés,
- ⑩ des fichiers netCDF non grillés,
- ⑩ des fichiers ODV,
- ⑩ des fichiers ODV bio.

Une partie de l'atelier sera aussi consacrée à faire l'inventaire des fonctionnalités que l'on désire avoir sur le site ?

XXX VOIR ou l'on dépose les fichiers pour la constitution de la base avec Gilbert XXX

Le prochain atelier se fera les 17 et 18 Octobre à Toulouse.

