

Compte Rendu de l'atelier technique ODATIS du 17 et 18 Octobre 2018

CR atelier technique Oct 2018

Numéro du livrable	Titre court
	CR Atelier ODATIS Oct 2018
Titre long	
Compte rendu de l'atelier technique ODATIS du 17 et 18 Octobre 2018	
Description courte	
Auteur	Groupe de travail
Joël Sudre	
Dissémination	Copyright

Historique

Version	Auteurs	Date	Commentaires
0.1	Joël Sudre	25 Octobre 2018	Version initiale
0.2	Jérôme LLido	26 Octobre 2018	Relecture et correction
0.3	Fabrice Mendes	29 Octobre 2018	Relecture et précisions
0.4	Caroline Mercier	08 Novembre 2018	Insertion des hyperliens

0.5	Philippe Téchiné	08 Novembre 2018	Relecture et correction
0.6	Dimitry Khvorostyanov	12 Novembre 2018	Relecture et précisions

Table des matières

1. Accueil et tour de table des participants.....	5
2. Avancées et état des lieux du pôle ODATIS et groupe de travail inter-pôles (Gilbert Maudire).....	6
3. Gestion des données Océanographiques.....	7
3.1. Présentation du service de diffusion ODES : Online Data Extraction Service (Frédéric Briol).....	7
3.2. Présentation du projet Datalake du CNES (Pierre-Marie Brunet).....	9
3.3. Présentation de PANGEO (Guillaume Eynard).....	10
3.4. Gestion des données au CTOH (Fernando Niño).....	11
3.5. Gestion des données Océan au SEDOO (Guillaume Brissebrat).....	11
3.6. Présentation de la solution logicielle mise en place au LOCEAN : INSITUDE (Dimitry Khvorostyanov).....	12
3.7. Gestion des données des SNO SSS et ROSAME (Philippe Téchiné).....	14
4. Inventaire des fonctionnalités du site web.....	14
5. Discussions.....	16
5.1. Discussions autour de ERDDAP et du protocole DAP.....	16
5.2. Discussions autour du protocole SOS.....	17
6. Conclusions et objectifs du prochain atelier.....	17

1. Accueil et tour de table des participants

Liste des participants à l'atelier ODATIS :

- Frédéric Briol (CLS) – FB,
- Guillaume Brissebrat (SEDOO) – GB,
- Pierre-Marie Brunet (CNES) – PMB,
- Pascal Calvat (OASU) – PC,
- Gérald Dibarboure (CNES) – GD,
- Guillaume Eynard (CNES) – GE,
- Mark Hoebeke (CNRS – SBR) – MH,
- Dimitry Khvorostyanov (LOCEAN) – DK,
- Jérôme LLido (LEGOS - SNO PIRATA) – JLL,
- Gilbert Maudire (Ifremer) – GM,
- Fabrice Mendes (CNRS – OASU) – FM,
- Caroline Mercier (AKKA/CNES) – CM,
- Fernando Niño (LEGOS - SNO CTOH) – FN,
- Catherine Schmechtig (CNRS – OOV) – CS,
- Sabine Schmidt (EPOC) – SS,
- Laurent Soudarin (CLS) – LS,
- Joël Sudre (CNRS -LEGOS) – JS,
- Philippe Téchiné (LEGOS – SNO ROSAME et SSS) – PT,
- Jean-Michel Zigna (CLS) – JMZ.

Personnes excusées :

- Aurélie Briand (IFREMER) – AB,
- Maurice Libes (OSU Pytheas) – ML,
- Didier Mallarino (CNES) – DM,

JS présente l'ordre du jour (*voir : 00_joel_sudre_agenda_atelier_odatis_oct_2018.pdf*), en précisant que l'agenda a été profondément modifié depuis son envoi lors du message d'invitation. En effet, la partie ERDDAP (TP et Hackathon) ne peut pas être menée à bien car les 3 participants devant animer l'Hackathon (AB, ML, et DM) n'ont pas pu venir pour des raisons totalement justifiées.

Tour de table des participants.

Le compte rendu de l'atelier de Juillet 2018 a été approuvé et mis en ligne sur le site ODATIS avec un accès publique.

2. Avancées et état des lieux du pôle ODATIS et groupe de travail inter-pôles (Gilbert Maudire)

(Voir : [01_Gilbert_Maudire_IR_PoleOdatis.pdf](#))

GM informe des avancées de l'Infrastructure de Recherche (IR) « Système Terre » présidée par Frédéric Huynh (IRD) suite au séminaire de IR qui s'est tenu les 5 et 6 Octobre:

- appel à idée pour changer le nom IR
- discussion sur l'architecture technique de l'IR pour répondre à l'appel d'offre PIA
- Création d'une UMS en Janvier 2019. Cette UMS est une UMS de soutien dédiée à l'IR et aux Pôles. Cette UMS est composée du Directeur de l'IR (F. Huynh), du Directeur Technique (R. Moreno), des Directeurs de pôles à temps partiel. Cette structure va permettre de positionner du personnel « en transversal ».

Une nouvelle structure transversale aux pôles va aussi se créer (Dynamis) et elle va permettre de mettre à disposition un accès à l'imagerie satellite haute résolution en fédérant l'ensemble des contrats et des conventions avec les différentes missions pour obtenir un accès unique (et gratuit) à toute la communauté scientifique.

GM rapporte aussi qu'une évolution du Groupe technique inter-pôles va devoir se faire à la suite du départ courant 2019 de Françoise Genova. Une présidence tournante est évoquée. GM note qu'il y a une participation importante des membres du pôle ODATIS dans les groupes de travail de l'interpôles (animation des groupes : formats, catalogue,...).

GM fait un point sur le Pôle ODATIS suite à la tenue du Comité Directeur début septembre. Lors de celui-ci, il a été décidé que la présidence du CODIR sera tournante, que chaque CDS (9 à 12) y sera représenté. En ce qui concerne les Stations Marines de Sorbonne Université aucune décision a été prise et elles sont en attente d'organisation (OSU?).

En ce qui concerne les Consortiums d'Expertise Scientifique (CES) du pôle ODATIS, leur fonctionnement est encore à instruire.

Le Conseil Scientifique du Pôle s'est tenu au printemps sous la présidence de Sabine Schmidt et un

atelier conjoint avec RESOMAR s'est aussi tenu (organisé par JS et Guillaume Charria). Un rendu de cet atelier se fera lors du séminaire RESOMAR (13-15 novembre 2018) ainsi qu'une nouvelle session de l'atelier.

En ce qui concerne le groupe de travail Catalogue de l'inter-pôles, un modèle Pivot à deux niveaux est retenu (voir schéma dans le fichier de la présentation - un premier niveau avec les collections et un second avec les granules de données). Le niveau granule va devoir être implémenté rapidement dans certains cas.

Enfin GM évoque la participation d'ODATIS :

- dans le projet européen ENVRI FAIR (participation conjointe avec AERIS),
- dans la préparation du service Blue Cloud,
- dans un projet CEP de renforcement des e-infrastructures avec le CINES et EUDAT,
- dans SeadataCloud 2.

Discussion suite à la présentation de GM :

[JMZ] : En ce qui concerne le modèle Pivot, est-ce qu'un modèle intermédiaire de type dataset est envisagé ?

[GM] : Non pas pour le moment, mais dans tous les cas le modèle sera autour de la norme Inspire et d'O&M (Observe and Measurement) ?

3. Gestion des données Océanographiques

3.1. Présentation du service de diffusion ODES : Online Data Extraction Service (Frédéric Briol)

FB présente le portail web interactif pour la distribution des produits altimétriques de niveau 2 et 2P (le niveau 2+ correspondant au niveau 2 avec des corrections supplémentaires) du CNES (voir : [02_Frederic_Briol_ODES.pdf](#)). Le système adossé à ce portail doit permettre de gérer plus de 60 années de données altimétriques le long de la trace, ce qui représente un volume et une hétérogénéité des données importants. De plus, le serveur de données doit aussi permettre à un

utilisateur authentifié d'obtenir des produits à la carte de manière synchrone et de répondre à des requêtes effectuées simultanément. Lors de la création d'une requête il doit être possible de sélectionner un produit, choisir une mission, sélectionner une zone géographique, définir une période d'extraction, et de s'abonner (ce qui permet d'obtenir les dernières données mises à disposition). Un système de fichiers virtuels (VFS) a donc été mis en place afin de ne pas dupliquer la donnée au moment de l'extraction tout en permettant de faire des requêtes dynamiquement. Il est implémenté par le biais d'une bibliothèque qui permet à un service FTP et un service HTTP de fonctionner simplement. Ce système de fichier virtuel contient les informations nécessaires pour créer des requêtes sans avoir à accéder réellement aux données. Ce système utilise une base PostGIS permettant les requêtes géographiques (points, lignes, polygones) à partir de la base de données PostgreSQL, et un Mapserver (serveur WMS, WFS, WCS). L'architecture du service, de la base de données et le système de fichiers virtuels sont explicités dans la présentation aux moyens de graphes fonctionnels. Le « chef d'orchestre » permettant la gestion des extractions a été implémenté en python (80 % du serveur) et en C++ (20 % du serveur – optimisé pour la parallélisation), l'IHM quant à elle est en JavaScript et PHP. Le site web est consultable à l'adresse suivante : <http://odes.altimetry.cnes.fr>.

Discussion suite de la présentation de FB :

[MH] : Est-ce qu'il y a une mise en cache des données ?

[FB] : Ceci n'était pas possible car la mise en cache écroulait le serveur. Il y a donc un traitement à la volée de chaque arc au moment où le client clique sur le fichier virtuel. Il est à noter que le port ftp (21) a été déporté sur un autre port pour des raisons de sécurité.

[JS] : Est-ce que le code source est disponible pour la communauté ?

[FB] : CLS s'occupe de la maintenance mais n'est pas propriétaire du code source. Il faut demander au CNES si le code source est disponible pour une autre utilisation. Il est à noter que ce code a été écrit en 2012 et qu'il mériterait d'être mis à jour avec de nouvelles bibliothèques Python qui sont actuellement plus performantes.

[PC] : iRODS (<https://irods.org>) est un logiciel « sur étagère » qui permet aussi de faire le même type d'extraction au moyen de fichiers virtuels.

3.2. Présentation du projet Datalake du CNES (Pierre-Marie Brunet)

Actuellement l'ensemble des données du CNES est géré par projet. Les données sont donc pilotées par projet, ce qui ne simplifie pas l'accès pour les utilisateurs. PMB présente le projet Datalake ([02_PierreMarie_Brunet_Datalake_CNES.pdf](#)) qui est en cours de construction au CNES et qui va permettre d'ingérer, d'accéder, de produire et de valoriser l'ensemble des données satellitales gérées par le CNES par un accès unique. Ce nouveau type de stockage va être accessible via les deux HPC du CNES (Hal et Ktulu) qui sont dédiés aux simulations numériques mais aussi aux traitements des données satellitales.

Pour l'utilisateur, cela va permettre d'avoir un accès simple, standard, et direct à la donnée via un environnement de développement (voir présentation suivante) qui va favoriser l'adoption des nouveaux paradigmes de développement parallèle (Dask, Spark,...) et de travail des données en mémoire. Par le biais de cet environnement, l'utilisateur va pouvoir colocaliser des données multi-sources et multi-temporelles.

Pour le CNES, Datalake va permettre d'optimiser les infrastructures et le support associé, mais aussi de rationaliser l'offre de service, d'améliorer les performances de calculs et l'interopérabilité entre le cloud et le centre de données. Les données vont être organisées sous forme de collection avec un quota spécifique, une politique de placement de purge et/ou de migration. Des Access Control List (ACL) spécifiques seront placés sur chaque collection de données.

Ce projet va démarrer avec les données Sentinel qui sont actuellement au catalogue et il sera ouvert à d'autres données au fil de l'eau.

Discussion suite de la présentation de FB :

[JMZ] : Pourquoi aller sur les « Data and Information Access Services » (DIAS) qui vont être payants plutôt que sur le Datalake ?

[PMB] : Attention le Datalake et l'atelier de valorisation n'ont pas pour vocation à faire de l'opérationnel. L'opérationnel devra être reporté dans les DIAS.

[PC] : Est-ce que le Gitlab est ouvert à tout le monde ?

[PMB] : Seules les personnes utilisant les données du CNES auront accès au Gitlab. Pour cela, il faut avoir un valideur CNES qui donne l'autorisation d'accès. A noter que le CNES a vraiment une

politique d'ouverture pour mettre à disposition des outils qui vont éviter de redévelopper les mêmes outils pour différentes données et de dupliquer les données (ce qui représente plusieurs centaines de To).

3.3. Présentation de PANGEO (Guillaume Eynard)

GE présente le projet Pangeo ([04_Guillaume_Eynard_Pangeo.pdf](#)) qui est une plateforme open source de développement dédiée aux traitements de gros volumes de données en géoscience. (<https://github.com/pangeo-data/pangeo>)

Pangeo est basé autour du langage Python et utilise des packages Python considérés comme essentiels pour le projet (Xarray, Iris, Pandas, Scipy, Matplotlib, Numpy, Bokeh, Cython, Numba, Dask, Sm, Scikit-image, Scikit-learn, Jupyter).

Cette plateforme est déployée sur les HPC du CNES et peut être accessible via le navigateur Web. Il est à noter qu'elle peut aussi être déployée sur une machine standard ou un cluster de calcul ce qui peut faciliter le développement des codes en local puis exporter ces codes sur un cluster, un HPC ou faire du calcul via de l'informatique dématérialisée ou en nuage (cloud computing). L'intérêt est de traiter la donnée là où elle est stockée et de ne plus la dupliquer.

La mission de la communauté PANGEO est de cultiver un écosystème dans lequel il est possible de générer, de distribuer et de maintenir des outils pour l'analyse de données en géoscience, et ce, quel que soit le volume de ces données.

Il est à noter que pour les technologies en nuage le Netcdf n'est pas adapté car chaque fichier est « lourd » et très riche en métadonnées. Pour palier à cela, il est nécessaire de convertir les fichiers Netcdf en Zarr, format qui est beaucoup plus adapté à la parallélisation des entrées/sorties avec des technologies en nuage.

La suite logicielle Pangeo associée à l'infrastructure du HPC CNES est une piste sérieuse qui pourrait servir de modèle à une infrastructure à mettre en place pour l'IR « système Terre » et les pôles. Il y a donc un appel à la communauté scientifique de la part du CNES à venir tester cette plateforme et à l'utiliser pour des projets nécessitant à la fois de la donnée satellite, *in-situ*, synthétique afin de voir si elle répond aux exigences d'un travail scientifique avec des outils implantés à distance.

3.4. Gestion des données au CTOH (Fernando Niño)

FN présente le Service National d'Observation Centre de Topographie des Océans et de l'Hydrosphère (CTOH – [05_Fernando_Nino_CTOH.pdf](#)) dédié à l'altimétrie satellitale. Ce service d'observation de l'INSU est sous la responsabilité de Florence Birol et existe depuis 1999.

Les objectifs du CTOH sont de documenter sur le long terme la formation, l'évolution et la variabilité de l'Océan, de soutenir le développement et l'utilisation des nouvelles applications altimétriques. Le CTOH fait aussi du support aux utilisateurs sur les données et produits altimétriques et assure le transfert de compétence vers les centres opérationnels (CNES,ESA).

Le CTOH collabore avec des partenaires nationaux et internationaux pour développer des produits altimétriques non-standards (altimétrie côtières, cryosphère, hydrologie continentale, fine échelle) à partir des données GDR des différentes missions. Le CTOH introduit dans les produits standards L2 (le long de la trace) des paramètres non-standards (données auxiliaires tels que marée, troposphère, ...) et les met à disposition pour la communauté scientifique. Il génère aussi des produits à valeur ajoutée (côtier, épaisseur de la banquise).

L'ensemble des données (toutes missions altimétriques confondues) représente environ 130 To qui est accédé par 2 serveurs de calcul. Le traitement des données SAR s'effectue sur le cluster de l'ESA et les tests sur les simulations des données SWOT sur le cluster du CNES.

Les jeux de données délivrés par le CTOH possèdent un DOI obtenu via le SEDOO. La donnée est diffusée soit par le site du LEGOS via un formulaire ou une demande de mail, soit via AVISO+ par ODES (voir [03_Frederic_Briol_ODES.pdf](#)) ou par ftp.

3.5. Gestion des données Océan au SEDOO (Guillaume Brissebrat)

GB présente la gestion des données Océan au Service de données de l'OMP (SEDOO – [06_Guillaume_Brissebrat_SEDOO.pdf](#)). Ce service a pour mission de favoriser l'échange de données scientifiques via le développement d'applications informatiques (gestion, traitement et archivage des données, interfaces web, production de produits à valeur ajoutée). Il a en charge les données de missions internationales (AMMA et MISTRALS), certaines données nationales pour les pôles de données ou les SNO et des données locales propres aux projets des laboratoires de l'OMP

et du Centre National de Recherches Météorologiques (CNRM). Ceci représente 30 applications en production et 60 sites web institutionnels.

Le SEDOO s'appuie sur le SI de l'UMS831 et sur le SI du LA (Laboratoire d'Aérodologie de l'OMP) ainsi que sur le site de sauvegarde et d'archivage de l'OMP.

Les données de l'atmosphère représente 75 % des données gérées par le SEDOO, les autres 25 % sont réparties sur les trois autres pôles de données (Theia, Odatis, [Form@ter](#)).

LE SEDOO possède une « usine à sites web en wordpress » ainsi que différents kits pour mutualiser les sites web. Il génère des sites de support aux campagnes en mettant en ligne les données exploitées pendant les campagnes (affichage de produits de modèles, satellite, *in-situ* etc). Il gère les données SSS, SOLWARA, etc. (voir la présentation).

Le SEDOO peut attribuer des DOI car il a une convention avec l'INIST. Il attribue des DOI pour les données qui sont gérées au SEDOO et a mis en place le site dataverse pour *Æris* afin d'attribuer des DOI sur les données d'*Æris* . Il gère le préfixe des données du pôle *Æris* pour les 4 CDS. (<https://dataverse.org>)

Discussion suite à la présentation de GB :

[SS] : Est-ce que le SEDOO produit les données qui sont sur le site de support aux campagnes ?

[GB] : Non, les images, les produits qui sont mis en ligne sur ces sites sont envoyés par des scientifiques ou des personnes tierces pour la campagne.

3.6. Présentation de la solution logicielle mise en place au LOCEAN : INSITUDE (Dimitry Khvorostyanov)

DK a présenté la gestion des acquisitions Iridium/Argos de l'équipe DITM et des projets associés du LOCEAN. La plateforme s'appuie sur la technologie Python (framework Django) couplé à une base de données (actuellement sqlite3 mais pouvant être PostGIS ou autre), ce qui donne satisfaction sur la gestion des métadonnées et des données acquises (voir le résumé ci-dessous).

La visualisation tire profit des avancés en terme de performances des technologies WebGL via CesiumJS (<https://cesiumjs.org/>), ainsi que de la librairie Highcharts.

LOCEAN lab has an instrumental development team that designs and deploys various buoys, such as drifters or floats. LOCEAN is also a principal investigator in a number of projects involving deployment buoys in various parts of the globe, acquiring and managing their measurements data. The INSITUDE platform is developed at LOCEAN to manage the acquisitions, follow in the near-real time, and interactively explore the data related to these projects.

The software solution employed by the platform consists of five components: (1) Django application for the meta-data management, (2) Python processing script; (3) node.js/expressJS for the server-side interaction with the database, (4) Highcharts for interactive visualization of time series and vertical profiles, (5) CesiumJS for interactive visualization of the buoy trajectories.

The workflow involves the following three steps:

(1) The user specifies the relevant meta-data using the web interface of the Django application; the meta-data database is thus updated;

(2) The processing script is launched automatically at regular times during a day: it reads the meta-data from the database, queries the mailboxes and/or external web services for the data wanted, receives the data, processes it, and fills the measurement database. It also generates a few data files for each processed dataset, some of which are directly accessible via a dedicated web page and others can be used for processing with external user programs (e.g. matlab or python scripts);

(3) The data stored in the measurement database can be interactively explored using the DataView application, allowing zoomable views of time series, vertical profiles, and trajectories shown on the virtual globe for one or more variables selected by the user via the web interface. The advantages of the system include centralized and intuitive access to the acquisition control, to the acquired data and related meta-data (projects, campaigns, buoys, people, etc.), and interactive visualizations performed with a unique tool accessing a unique database and allowing simultaneous exploration of data for different variables and from different sources.

Il s'en est suivi une discussion riche sur la montée en puissance des technologies NodeJS qui ont été conçues dans un objectifs de simplicité et efficacité (même langage côté serveur et côté client = javascript ; des communications non-bloquantes avec la base de données, efficacité dans certains cas même avec un seul fil d'exécution). Il es difficile d'évaluer s'il s'agit d'un effet de mode ou si le phénomène va perdurer et jusqu'à quel point il va pouvoir se positionner sur le traitement de la donnée.

3.7. Gestion des données des SNO SSS et ROSAME (Philippe Téchiné)

PT présente le SNO ROSAME dirigé par Laurent Testut et le SNO SSS dirigé par Gaël Alory (07_Philippe_Téchine_Rosame_SSS.pdf). PT est en charge de la gestion, du traitement et de la diffusion de l'ensemble des données de ces deux SNO.

Le Réseau d'Observation Subantarctique et Antactique du niveau de la MER (ROSAME) est une collaboration de divers instituts (LEGOS, DT/INSU, IPEV, TAAF, IGN, LIENSs). C'est un réseau de données *in-situ* de marégraphes dans les TAAF créé en 1991 possédant 4 sites côtiers et plusieurs stations de plateau. Ce service est intégré depuis 2015 au SNO SONEL et il est l'une des composantes du programme international GLOSS. ROSAME contribue par la distribution de ces données marégraphiques à diverses applications scientifiques multi-thématiques allant de l'étude des marées océaniques à la validation des observations satellitales.

Le réseau Sea Surface Salinity (SSS) quant à lui est une collaboration entre le LEGOS, l'US IMAGO de l'IRD, le LOCEAN, l'IPEV et le CSIRO. Ce réseau *in-situ* de thermosalinographes embarqués sur des navires de commerce sur tous les océans du globe a été créé en 2002. Ce service est labellisé SNO par l'INSU et IRD-Sud et il fait partie du SOERE CTDO2. Il est aussi partie intégrante du programme international GOSUD.

Pour ces deux services d'observation, les chaînes de traitement de l'ensemble des données intégrant des contrôles de qualité sont implémentées au LEGOS. Ces deux SNO s'occupent à la fois de la réception, du traitement, de la qualification et de la distribution des données marégraphiques et de salinité de surface de l'océan. Les sites web (<http://www.legos.obs-mip.fr/observations/rosame> et <http://www.legos.obs-mip.fr/observations/sss>) permettent de consulter, visualiser en temps réel les données. Les données en temps différées sont aussi intégrées dans différents centres de distribution nationaux et internationaux (voir présentation). Un travail conséquent a été fourni sur des indicateurs de qualité et de répartitions spatiales des données (cartes et chronogrammes).

4. Inventaire des fonctionnalités du site web

CM présente le site web du pôle ODATIS (<http://www.odatis-ocean.fr>).

Il reste à mettre à jour les pages de certains CDS.

Le site web est en partie traduit en anglais et la version anglaise du site reste à finaliser. À noter qu'il est parfois assez délicat pour un non expert du domaine de traduire certaines pages sans altérer le contenu scientifique. Il est donc demandé à chaque producteur de données de vérifier

leurs pages web dédiées et de faire remonter les corrections si nécessaire. L'idéal est que les producteurs fournissent à la fois une version française et anglaise de leurs textes. Il est rappelé que les directives du ministère au sujet des langues sont très claires et que le site web des pôles doit être en français prioritairement (les versions anglaise et autres étant optionnelles).

Les pages de l'onglet « actualités » ont été modifiées et sont maintenant centrées sur les nouveaux produits intégrés dans le pôle.

En ce qui concerne le catalogue, deux niveaux vont être considérés :

- un niveau collection (jeux de données),
- un niveau granule (paramètres).

Ce catalogue doit suivre la norme ISO 19115. Il est donc nécessaire de rajouter manuellement quelques mots clés après moissonnage pour être conforme à l'utilisation de Geonetwork pour le site ODATIS. L'utilisation du Geonetwork d'Odatis permettra d'être orienté vers les Thésaurus d'Odatis.

Le CDS-Sat de Brest a demandé un passage de la norme ISO 19115-2 à la norme ISO 19115-3 car il y a eu un enrichissement des métadonnées par l'ESA et la NASA pour plusieurs produits satellitaires à partir du catalogue des EOVS (Essential Ocean Variables). Il est donc nécessaire de reprendre certaines fiches du catalogue pour les passer à cette nouvelle norme.

Un serveur de vocabulaire est envisagé pour IR « Système Terre » mais il se pose certaines questions pour son enrichissement, sa mise à jour etc. Au niveau du pôle THEIA, Véronique Chaffard étudie la possibilité de mettre en place un serveur de vocabulaire en français avec une synchronisation avec le serveur du PODC. Le serveur du PODC sera enrichi en retour (enrichissement par paquet avec une validation par les personnes qualifiées au PODC)

Afin de rendre plus visible et d'avoir un meilleur référencement, le site web ODATIS doit aussi être accessible par le https. Pour cela, il est nécessaire que l'Ifremer (détenteur du nom de domaine) fasse le certificat et le transmette ensuite au CNES (CLS) qui développe le site web.

Enfin il est nécessaire de mettre en conformité le site web avec la RGPD (avertir que le site utilise des cookies).

5. Discussions

5.1. Discussions autour de ERDDAP et du protocole DAP

L'hackathon sur ERDDAP n'ayant pas pu se dérouler au cours de cet atelier du fait de l'absence des animateurs, une discussion sur l'utilisation d'ERDDAP au sein d'Odatis s'est engagée.

En ce qui concerne les formats de données, il a été confirmé que 2 formats sont possibles pour les données gérées dans le pôle :

- un format Netcdf avec la norme CF (voir <http://cfconventions.org/>),
- un format ASCII en colonne type CSV (ODV spreadsheet file format - voir https://www.bodc.ac.uk/resources/delivery_formats/odv_format/).

Au moins l'un de ces deux formats doit être adopté par l'ensemble des CDS pour homogénéiser l'ensemble des données. Un travail de normalisation de la donnée restera toujours nécessaire pour chaque CDS.

Au delà du simple format d'échange, il est nécessaire au sein d'Odatis que chaque CDS soit capable de fournir une certaine collection de services web pour permettre des échanges et de l'interopérabilité entre les CDS et vers le site ODATIS. Il est donc nécessaire au sein du pôle de mettre en place des composants techniques permettant de regrouper les données afin d'offrir des services web avancés. Le principal écueil est de synchroniser les dépôts des CDS avec le portail. Pour cela il est envisagé de mettre en place un « cache technique national » sur les données pour les centraliser afin qu'elles soient toutes disponibles en un temps acceptable par les différents services que va offrir le site Odatis (la synchronisation des données se faisant une fois par jour). À noter que ce « cache technique national » doit-être considéré comme un accélérateur pour permettre d'avoir des temps de réponse acceptables et non comme la vitrine ultime du pôle. Ce cache ne sera d'ailleurs absolument pas accessible, ni par les fournisseurs de données, ni par les utilisateurs et son utilisation se fera de manière transparente via le site du pôle.

Afin de synchroniser les dépôts des CDS plusieurs solutions sont envisagées utilisant le protocole DAP (ou OpeNDAP) pour fournir des web services (web services qui restent à définir pour le pôle) soit:

- chaque CDS a son ERDDAP,
- un ERDDAP central avec une récupération des données sur un ftp,

- un mélange des deux solutions précédentes.

ERDDAP est un bon candidat pour une mise en place dans les CDS, cependant sa faiblesse réside au niveau RH car cet outil est développé par une seule personne. Il est donc aussi envisagé d'étudier des solutions alternatives à ERDDAP. Alternatives qui peuvent être des logiciels prêts à l'emploi (Hyrax, Oceanotron, Dapper,...) ou l'utilisation de librairie à utiliser par des développeurs permettant de faire un outil « maison » (PyDAP, libdap,etc).

Il est nécessaire de pousser plus en amont cette discussion qui va lourdement peser sur l'architecture technique du pôle. Il a donc été convenu que le prochain atelier serait consacré à la mise en place d'un cahier des charges et à l'architecture globale du pôle Odatis.

5.2. Discussions autour du protocole SOS

Le protocole SOS est aussi à explorer car il permet d'avoir un état des capteurs à chaque instant (contrôleur d'observatoire, configuration de système complexe de capteur). Cependant ce protocole est coûteux en terme de flux transféré et il semble plus adéquat à une utilisation en amont des CDS par les producteurs de données. Ce protocole sera donc envisagé dans un second temps au niveau du pôle (pôle qui a un rôle de distribution de la donnée et de création de nouveaux produits à valeur ajoutée en aval des producteurs). Une étude coordonnée entre le pôle ODATIS (et AERIS) et le réseau SIST est proposée pour animer un atelier dédié permettant de faire venir des intervenants spécialistes de ce protocole (52°North, istSOS).

6. Conclusions et objectifs du prochain atelier

L'hackathon ERDDAP n'ayant pas pu se faire au cours de cet atelier, il a été reporté à l'atelier de novembre.

Au vue du calendrier et de l'ouverture imminente de l'appel d'offre pour le PIA3 (Programme d'Investissement d'Avenir) auquel le pôle Odatis doit répondre au travers de la proposition de l'IR « Système Terre », il est nécessaire d'établir un cahier des charges et de réfléchir à l'architecture globale du pôle. Pour ce faire, l'atelier de novembre aura comme sujet principal l'écriture du cahier des charges et de schématiser l'architecture du pôle.

Le [prochain atelier](#) se fera les 20 et 21 Novembre à Paris (LOCEAN - Jussieu).