

Réunion : Bilan des activités 2022-2023 sur l'action Data Sciences de l'IR ILICO

07 Avril 2023

ORDRE DU JOUR

Présentation.....	3
1. Plans de gestion de données (DMP)	3
2. Atelier FAIR Implementation Profile (FIP)	3
3. Mise à jour de la couche des stations ILICO dans Sextant.....	3
4. Mise à jour des fiches d'identité SNO sur le site web ILICO	3
5. Développement d'un portail vitrine	3
6. Communications externes.....	4
Synthèse des discussions.....	4
1. Intégration de données non labellisées SNO dans le portail vitrine	4
2. Mise à jour des métadonnées.....	4
3. Retour sur l'utilisation des données.....	5
4. Retour sur le portail.....	5
Conclusion	6

PARTICIPANTS

La réunion s'est tenue en visioconférence.

Durée de la réunion : 1h45

NOM	PRENOM	AFFILIATION	REDACTION DU COMPTE RENDU
THOMAS	Amandine	ILICO - Institut universitaire européen de la mer	X
LEHUEN	Amélie	Atelier NEO - Université de Caen	
AYACHE	Bruce	BRGM	
LEFEBVRE	Alain	ILICO - Ifremer	
TREGUER	Mickaël	ISI - Ifremer	
VERNET	Marine	ODATIS - Ifremer	
QUIMBERT	Erwann	ODATIS - Ifremer	
HOEBEKE	Mark	ODATIS / SNO PHYTOBS et BenthOBS - Station Biologique de Roscoff	
CORDIER	Emmanuel	OSU Réunion	
CAER	Gwenaël	SISMER - Ifremer	
CHARRIA	Guillaume	SNO COAST-HF - Ifremer	
BOSSE	Anthony	SNO MOOSE - Institut Méditerranéen d'Océanologie	
LEMOINE	Maud	SNO PHYTOBS - Ifremer	
EL HOUJJAJI	Fatima Ezzahraa	SNO PHYTOBS et BenthOBS - Station Biologique de Roscoff	
SAVOYE	Nicolas	SNO SOMLIT - Université de Bordeaux	
RIMMELIN MAURY	Peggy	SNO SOMLIT et COAST-HF - Institut universitaire européen de la mer	
TESTUT	Laurent	SNO SONEL - LIENSs	

PRESENTATION

1. PLANS DE GESTION DE DONNEES (DMP)

Plan de gestion de données (DMP – *Data Management Plan*) : document évolutif décrivant la collecte, le traitement, la documentation, le stockage, l'archivage, la conservation et le partage des données pendant et après un projet de recherche.

Le contenu des DMP a été transféré sur l'outil collaboratif en ligne DMP OPIDoR. Il s'agit d'un outil géré et hébergé par le CNRS proposant des modèles tels que le modèle « Science Europe : modèle structuré » choisi par l'IR ILICO. Ce modèle permet d'exporter les DMP sous différents formats dont un format json interrogeable par des machines.

Les **droits d'édition** des DMP dans DMP OPIDoR **vont être partagés avec chaque coordinateur** de SNO pour que des modifications puissent être apportées si besoin. Ces DMP **vont être rendus publics** sur DMP OPIDoR et déposés sur Archimer. L'**association de DOI** à chaque DMP est envisagée pour faciliter leur citation, les pérenniser et les rendre plus facilement trouvables.

2. ATELIER FAIR IMPLEMENTATION PROFILE (FIP)

L'IR ILICO a participé à un atelier FAIR Implementation Profile (FIP) animé par le Pôle National de Données de Biodiversité (PNDB) le 24/10/2022. Cela a permis de faire un **état des lieux des outils, des technologies et des standards relatifs aux données et métadonnées de l'infrastructure**. Le document obtenu a été mis à disposition dans la rubrique « Documents » du site internet ILICO et une [actualité](#) a été rédigée.

3. MISE A JOUR DE LA COUCHE DES STATIONS ILICO DANS SEXTANT

La couche des stations de l'IR ILICO diffusée par Sextant a été mise à jour pour **intégrer de nouveaux champs** (variables, fréquence d'acquisition, instruments d'acquisition, profondeur, producteurs de données, variables ODATIS). Ces nouveaux champs vont permettre d'effectuer des **filtres sur les stations affichées** dans l'interface cartographique de Sextant.

4. MISE A JOUR DES FICHES D'IDENTITE SNO SUR LE SITE WEB ILICO

Lors de la mise à jour du site internet ILICO, les pages contenant les fiches d'identité des SNO (Services Nationaux d'Observation) ont été mises à jour. Ces pages contiennent 4 onglets : présentation du SNO, l'intérêt scientifique, l'accès et la visualisation des données et les partenariats.

5. DEVELOPPEMENT D'UN PORTAIL VITRINE

Un prototype de portail d'exploration de données ILICO est en cours de développement. L'objectif de ce portail est de proposer une **interface homogène de visualisation et d'accès aux données des 9 SNO**. Ce portail est entièrement développé en python. Les données des SNO ont été copiées à un instant T et sont ensuite standardisées en **OGC SensorThings API**. Cette API normalisée repose sur le modèle Observations&Measurements.

Pour le moment, le portail permet d'accéder aux données PHYTOBS, BenthOBS et DYNALIT. Il se présente sous la forme d'un **tableau de bord** dans lequel l'utilisateur peut effectuer des requêtes attributaires (sur le réseau, la station, la variable, l'espèce, la période, le type de graphique) pour alimenter des graphiques. Chaque graphique est alimenté par des requêtes indépendantes les unes

des autres. La technologie **Sextant a été intégrée** dans le portail pour proposer une **interface cartographique** et accéder au **catalogue de données**.

Les données de certains SNO soulèvent des questions en termes de compréhension et/ou de représentation des données qui nécessitera des discussions avec les coordinateurs des SNO.

Le portail contient plusieurs onglets dont un onglet proposant des **cas d'usage à l'utilisateur** et un autre décrivant la **source des données**. Ces onglets sont en construction.

6. COMMUNICATIONS EXTERNES

Deux posters ont été présentés sur les données de l'IR ILICO entre 2022 et 2023 :

- Cocquempot Lucie, Thomas Amandine, Delacourt Christophe, Paillet Jerome, Burden Joanne, Charria Guillaume (2022). **A French research infrastructure for Coastal Ocean and Seashore Observations**. Sea Level Workshop "from Global to Coastal, from Past to Future". 1-3 June 2022, Brest. <https://archimer.ifremer.fr/doc/00775/88652/>
- Thomas Amandine, Hoebeke Mark, Quimbert Erwann, Treguer Mickael, Lefebvre Alain, Delacourt Christophe, Burden Joanne, Cocquempot Lucie (2023). **Portail vitrine de l'IR ILICO**. MerIGeo 2023 - quatrième édition du colloque national dédié à la géomatique appliquée au milieu marin. 14 au 16 mars 2023, Le Havre. <https://archimer.ifremer.fr/doc/00784/89609/>

SYNTHESE DES DISCUSSIONS

1. INTEGRATION DE DONNEES NON LABELLISEES SNO DANS LE PORTAIL VITRINE

Les données de certains observatoires régionaux ne sont pas visualisables dans le portail vitrine ILICO alors qu'ils proposent parfois des séries de données historiques sur 25 ans mises à disposition de manière interopérables.

A l'heure actuelle, le portail vitrine intègre uniquement les données labellisées SNO ce qui implique une labellisation par la Commission spécialisée Océan-Atmosphère (CSOA). Cependant, en parallèle du développement du portail, il existe des travaux sur les **Données Complémentaires d'Intérêt Collectif (DCIC)** correspondant à des jeux de données non labellisés mais pouvant présenter un intérêt à être croisées avec des données labellisées. Ces DCIC pourraient être par exemple citées ou référencées dans le portail pour établir un lien avec les données labellisées consultées.

2. MISE A JOUR DES METADONNEES

ILICO produit de la métadonnée à différents endroits (site web ILICO, Sextant – parfois connectés entre eux) avec différents niveaux d'information (ex : objectifs, missions, personnes impliquées, données, etc.). Mais **comment sont prévues les mises à jour ?** Manuellement ?

L'idée d'ajouter des métadonnées dans la couche SIG des stations était justement de pouvoir mettre à jour plus facilement les informations (via requêtes dans un outil SIG par exemple) que dans de nombreuses fiches métadonnées. L'inconvénient de cela, par contre, est que les services web sont des standards de données et non de métadonnées. Ces informations ne seront donc pas moissonnables par des catalogues de métadonnées (contrairement par exemple au standard ISO 19139).

La mise en place d'un outil automatisé de mise à jour des informations peut s'avérer très complexe car l'origine des informations est extrêmement diverse. De plus, imposer un standard commun à tous

les SNO serait compliqué car les données sont très différentes. Cependant, certaines métadonnées donnent beaucoup d'informations (ex : nom des personnes effectuant les relevés) qui deviennent rapidement obsolètes. Cela interroge sur le **niveau de précision à fournir** pour avoir le bon compromis entre **donner de l'information** et **maintenir cette information à jour**. Est-ce que donner un grand nombre d'informations n'est pas à court terme contre-productif ?

L'un des objectifs de l'IR est de faciliter la gestion de la donnée pour les SNO, et non de la complexifier. C'est aussi de permettre un accès global à toutes ces données de l'observation au niveau national français tout en les rendant compatibles avec des contraintes européennes (ex : lien ILICO-JERICO). Certains degrés de détails retrouvés dans les métadonnées peuvent anticiper les besoins à l'échelle européenne. **L'IR est donc sensible à l'automatisation des processus de mise à jour des données** dès que cela est possible.

La question sur la mise à jour des métadonnées pourrait peut-être se résoudre par la mise en place d'une **stratégie commune à l'échelle de l'IR pour la DOIsation des différentes informations**. Quand un objet est DOIisé (ex : SNO, jeu de données de SNO), une fiche de métadonnées est souvent associée à l'objet (= landing page). Est-ce qu'il ne faudrait donc pas faire l'inventaire de ce qui est DOIisé chaque année, à quelle échelle et qui le fait ? En admettant que toutes les informations peuvent s'intégrer dans la démarche de DOIisation ce qui n'est pas certain. Par exemple, le coordinateur du SNO DOIse la fiche métadonnée du SNO, il échange les métadonnées avec l'IR (uniformisation à l'échelle de l'IR) et l'IR DOIse un catalogue regroupant l'ensemble des fiches des SNO. A cela il faut aussi intégrer les démarches des stations d'observation qui saisissent et DOIent des métadonnées également. **Actuellement, chaque SNO a sa propre méthode d'attribution de DOI** (ex : à l'échelle du SNO ou des stations), il serait donc pertinent de proposer une stratégie commune (incluant notamment des vocabulaires communs).

3. RETOUR SUR L'UTILISATION DES DONNEES

Du point de vue de l'utilisateur des données, cela est très satisfaisant de pouvoir accéder facilement aux données via des portails. Cependant, parfois cela peut être compliqué de trouver :

- **La méthode de citation des sources** : l'utilisateur souhaite utiliser les données et les citer correctement mais la méthode de citation n'est pas toujours précisée. Lorsqu'il y a un DOI, la citation est proposée mais tous les jeux de données ne sont pas forcément rattachés à un DOI.
- **La description des champs** : lorsque l'utilisateur récupère les données, la définition des champs n'est pas toujours précisée et celle-ci n'est pas forcément intuitive. Cela pourrait par exemple être résolu par l'association d'un fichier de métadonnées aux données lors du téléchargement ou l'utilisation de vocabulaires contrôlés. Le vocabulaire contrôlé donne une définition des champs.

4. RETOUR SUR LE PORTAIL

L'IR ILICO propose un prototype de portail pour voir si cela sollicite de l'intérêt. Cette réunion est également l'occasion d'avoir un retour des SNO sur deux questions :

- Est-ce que le prototype présenté apporte une valeur ajoutée ? Si c'est le cas, il sera possible de prendre en compte de nouvelles variables et de nouvelles contraintes en intégrant notamment de l'automatisation.
- En termes de figure de synthèse, tableau de synthèse, etc. qu'est-ce qui est considéré comme le minimum à proposer dans le portail ?

Une plus-value du portail pourrait être de croiser les données avec des données non labellisées, par exemple des données de climatologie.

La colocalisation des stations est un enjeu important pour la valeur ajoutée du portail car on ne peut pas comparer des données qui n'ont pas été acquises au même endroit. Il serait également intéressant de pouvoir proposer des agrégations (ex : à l'échelle régionale), de la visualisation 3D.

Concernant les gros volumes de données (ex : données MOOSE), il sera nécessaire d'identifier les jeux de données à afficher en amont. Il peut aussi être envisagé d'agréger les données ou de proposer une carte de position des profils et gliders intégrée dans le portail dans laquelle chaque position renverrait vers les DOI des données.

Concernant le partage des données, la création de nouveaux outils n'est pas forcément possible pour certains SNO pour des questions de moyens. Est-ce qu'il y a la possibilité d'accompagner les SNO les moins avancés pour faciliter la mise à disposition de leurs données (notamment en termes de ressource humaine) ?

Cela s'inscrit dans les missions du pôle océan ODATIS par l'intermédiaire des Centre de Données et Service (CDS). La plupart des SNO (sauf CORAIL et ReefTEMPS) sont associés à un **CDS qui a notamment pour mission d'accompagner les SNO sur le partage de leurs données.**

CONCLUSION

Les droits d'édition des DMP sur **DMP OPIDoR vont être partagés avec les coordinateurs** des SNO. Leur contenu va être publié sur Achimer et un DOI leur sera associé.

L'IR ILICO a produit de nombreuses métadonnées sur l'interface Sextant et sur son site internet. Cependant, se pose la question de la mise à jour de ces **métadonnées parfois très détaillées ce qui peut les rendre rapidement obsolètes.** Il est nécessaire de trouver le bon compromis entre la diffusion de l'information et la capacité à la maintenir à jour. Une première proposition faite durant la réunion serait de **réduire le niveau d'information.** Une seconde proposition est de mettre en place une **stratégie de DOIisation au sein de l'IR ILICO** par exemple en associant un DOI à chaque SNO et chaque station (la métadonnée serait alors présente dans la landing page du DOI). Cependant, la méthode de DOIisation est différente d'un SNO à un autre à l'heure actuelle, une harmonisation des pratiques serait donc nécessaire.

Un portail vitrine est en cours de développement, il intègre pour le moment les données PHYTOBS, BenthOBS et DYNALIT. Des discussions seront nécessaires avec les coordinateurs de certains SNO pour échanger sur les données et la manière de les représenter. Ce portail a vocation pour le moment d'afficher uniquement des données labellisées SNO. Cependant, des travaux sont en cours pour inventorier les données d'autres **réseaux non labellisés**, appelées « **Données Complémentaires d'Intérêt Collectif (DCIC)** », qu'il serait intéressant de croiser avec des données labellisées. Plusieurs **enjeux** ont été soulevés pour ce portail durant la réunion, notamment la **colocalisation des stations** pour pouvoir comparer des données acquises aux mêmes endroits et **l'affichage de gros volumes de données** (ex : données de haute fréquence). Pour cela, des **agrégations** peuvent être envisagées ou des **renvois vers les DOI** des jeux de données.

Du point de vue de l'utilisateur des données, il a été remonté la **difficulté dans certains cas de savoir comment citer correctement les sources des données** (facilité si un DOI est mis en

place, car il y a une proposition de citation, mais cela n'est pas toujours le cas). De plus, lors du téléchargement des données, la **définition des champs n'est pas toujours explicitée**. L'association d'un **fichier de métadonnées** au téléchargement des données ou l'utilisation de **vocabulaires contrôlés** (avec la mention du vocabulaire en question) facilitent leur compréhension.

Contact : data@ir-ilico.fr