



DATA
TERRA



Compte rendu de l'atelier technique des 15 et 16 décembre 2020

CR atelier technique décembre 2020





Titre court	
CR_AT_dec2020	
Titre long	
Compte rendu de l'atelier technique du 15 et 16 décembre 2020	
Auteur	
Joël Sudre	
Dissémination	Copyright
	Pôle ODATIS

Historique

Version	Auteurs	Date	Commentaires
0.1	Joël Sudre	02 février 2021	Version initiale
0.2	Cécile Nys	04 février 2021	Relecture et corrections
0.3	Damien Boulanger	05 février 2021	Modifications mineures
0.4	Thierry Carval	05 février 2021	Modifications mineures
1.0	Cécile Nys Caroline Mercier	16 mars 2021	Version diffusable

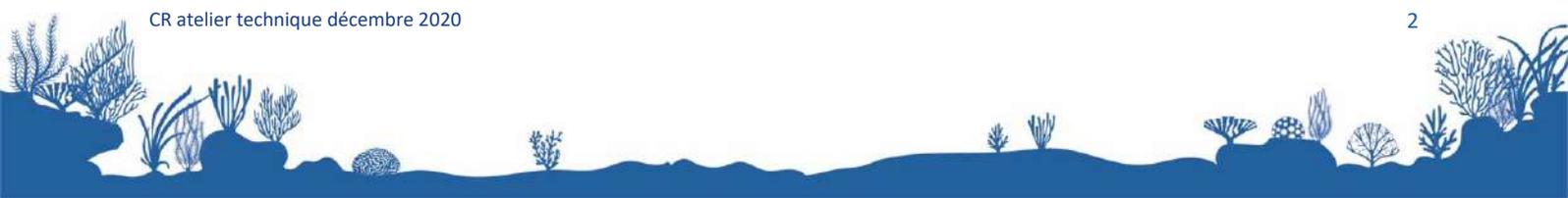




Table des matières

1. Accueil et tour de table des participants	5
2. Introduction – (Gilbert MAUDIRE)	6
3. Sessions FAIR	6
3.1. Principes FAIR – (Damien BOULANGER)	6
3.2. FAIR dans Data Terra – (Damien BOULANGER)	8
3.3. ENVRI-FAIR partie Marine Subdomain – (Damien BOULANGER, Valérie HARSCOAT & Thierry CARVAL)	10
3.3.1. Introduction générale du projet ENVRI-FAIR – (Damien BOULANGER)	10
3.3.2. Introduction du WP9 Marine sub-domain – (Valérie HARSCOAT).....	13
3.3.3. Améliorations des principes FAIR dans Euro-ARGO – (Thierry CARVAL)	15
3.4. Atmosphère - AERIS – (Damien BOULANGER)	16
3.5. Discussion à la suite de la Session FAIR	18
4. Sessions identifiants permanents	19
4.1. PID dans et pour ODATIS et COPiLOtE – (Gilbert MAUDIRE)	19
4.2. Principes, Guidelines pour différents jeux de données & bibliométrie – (Frédéric MERCEUR)	20
5. Synthèse des 2 jours des sessions de support par CDS	23
6. ANNEXES	24
6.1. Questions/Réponses sessions FAIR.....	24
6.2. Questions/Réponses sessions identifiants permanents	25





Table des illustrations

Figure 1. Les différents projets ENVRI	11
Figure 2. Principes d'implémentation du projet ENVRI-FAIR.....	11
Figure 3. Les WPs du projet ENVRI-FAIR	12
Figure 4. Analyse FAIRness, l'exemple de l'IR Euro-ARGO.....	14
Figure 5. Plan d'implémentation ENVRI-FAIR pour Euro-Argo. Les cases en orange représentent les futures améliorations à implémenter.	14
Figure 6. Schéma de l'ontologie ARGO	16
Tableau 1. Liste des participants à l'atelier ODATIS #8 des 15 et 16 décembre 2020	5





1. Accueil et tour de table des participants

Tableau 1. Liste des participants à l'atelier ODATIS #8 des 15 et 16 décembre 2020

Liste des participants à l'Atelier Technique #8	
Damien Boulanger (SEDOO / CNRS)	Jean-François Le Roux (IFREMER)
Dominique Briand (IFREMER)	Romain Lhullier (SHOM)
Joanne Burden (IUEM / CNRS, Ilico)	Maurice Libes (OSU Pytheas / CNRS)
Pascal Calvat (OASU / CNRS)	Didier Mallarino (OSU Pytheas / CNRS)
Valérie Cariou (SHOM)	Gilbert Maudire (IFREMER)
Thierry Carval (IFREMER)	Fabrice Mendes (OASU / CNRS)
Victor Champonnois (LOCEAN / CNRS)	Frédéric Merceur (IFREMER)
Gérald Dibarboure (CNES)	Caroline Mercier (CNES/AKKA)
Ludovic Drouineau (IFREMER)	Cécile Nys (IFREMER/OceanScope)
Michèle Fichaut (IFREMER)	Nicolas Pouvreau (Shom)
Valérie Harscoat (IFREMER)	Clémence Rabévol (IFREMER, SAR/SIMM)
Fabrice Herpin (OASU / CNRS)	Armelle Rouyer (IFREMER, SAR/SIMM)
Mark Hoebeke (SBR / SU / CNRS)	Catherine Schmechtig (IMEV / SU / CNRS)
Céline Houbin (SBR / CNRS)	Laurent Soudarin (CNS/CLS)
Dimitry Khvorostyanov (LOCEAN / SU / CNRS)	Joël Sudre (CPST / CNRS)
Yvan Le Bras (MNHN, PNDB)	Stéphane Tarot (IFREMER)
Fanny Lecuy (Shom)	

2. Introduction – (Gilbert MAUDIRE)

GM présente l'ordre du jour (voir : [Agenda et accès aux présentations¹](#)). Cet atelier est dédié à l'ANR COPILOtE et se présente sous un format un peu différent de l'atelier technique ODATIS. En effet, cet atelier est constitué d'une demi-journée en session plénière avec l'ensemble des participants, suivie de sessions individuelles (~ 1h30/CDS) avec chacun des CDS² du pôle ODATIS. La session plénière va s'intéresser à explorer ce qui se cache derrière les mots FAIR (« Findable, Accessible, Interoperable, Reusable »), identifiants permanents (PID, « Digital Object Identifier » - DOI) pour que l'ensemble des centres soient au même niveau d'information. Ces aspects sont importants car dans le cadre d'ODATIS, lors d'ateliers précédents, un **cahier des charges** des centres (Centre de Distribution et de Services - CDS ou Centre de Collecte - CC) a été mis en place et ce cahier des charges impose que la gestion des données des centres se fassent **en accord avec les principes FAIR** et de **mettre des identifiants permanents (DOI) sur chaque jeu de données géré par les centres**. Il est aussi important que l'on harmonise les pratiques dans ODATIS autour de ces principes et de ces DOI (qui assigne le DOI et que doit-on mettre dans les « landing pages »).

3. Sessions FAIR

3.1. Principes FAIR – (Damien BOULANGER)

DB présente les principes FAIR (voir [202012_ODATIS_Atelier_Boulanger_PrincipesFair.pdf³](#)). Ces principes proviennent d'un article de *Wilkinson et al.* (*"The FAIR Guiding Principles for scientific data management and stewardship"*, 2016, scientific data, DOI: [10.1038/sdata.2016.18⁴](#)). Cet article qui formalise des principes déjà existants a rapidement été accepté par la communauté scientifique internationale. Ces principes permettent de rendre FAIR les données :

- F (Findable) – Facile à trouver. Les données sont faciles à trouver par des humains ou des machines :
 - Les données et les métadonnées ont un identificateur unique et persistant (DOI, handles, etc.),
 - Les données sont décrites par des métadonnées riches et standardisées (ISO 1915, Data Cite, Conventions, ...),
 - Les données et métadonnées sont enregistrées ou indexées dans un dispositif permettant de les rechercher (portail de données, pôles, etc.),
 - Les métadonnées contiennent l'identificateur unique et persistant des données.

¹ <https://www.odatis-ocean.fr/activites/ateliers-techniques/atelier-technique-decembre-2020>

² CDS : Centre(s) de Données et de Services

³ https://www.odatis-ocean.fr/fileadmin/documents/activites/ateliers/atelier_202012/202012_ODATIS_Atelier_Boulanger_PrincipesFair.pdf

⁴ <https://doi.org/10.1038/sdata.2016.18>

CR atelier technique décembre 2020



- A (Accessible) – Accessible. Une fois les données trouvées, l'utilisateur doit savoir comment y accéder et s'il a besoin d'être authentifié et/ou autorisé :
 - Les données et métadonnées sont accessibles par leur identifiant via un protocole standard de communication,
 - Ce protocole est ouvert, libre, et peut-être implémenté de manière universelle (REST, CSW, OAI-PMH, OPeNDAP, ...)
 - Ce protocole peut gérer des procédures d'authentification et d'autorisation si nécessaire (ORCID, EduGAIN, Oauth, ...)
 - Les métadonnées restent accessibles même si les données ne sont plus disponibles (DMP, ...).
- I (Interoperable) – Interopérable. Les données peuvent être intégrées avec d'autres données, interopérables avec des applications ou des workflows :
 - Les données et métadonnées utilisent un langage formel, accessible, partagé et largement applicable pour la représentation des connaissances (RDF, JSON-LD, ...),
 - Les données et métadonnées utilisent des vocabulaires qui eux-mêmes suivent les principes FAIR (GCMD, Standard names, SDN, CF, ...),
 - Les données et métadonnées contiennent des liens vers d'autres (méta)données (liens vers d'autres versions, articles, ...).
- R (Reusable) – Réutilisable. Les données doivent être réutilisables pour de futures recherches ou autres activités :
 - Les (méta)données doivent être richement décrites, par une pluralité d'attributs décrits et pertinents,
 - Elles sont fournies avec une licence d'utilisation claire et accessible (CC, ...),
 - Elles sont associées à leur provenance (standard W3C PROV, ...),
 - Elles suivent les standards des communautés concernées.

De plus en plus, on parle aussi de reproductibilité de la donnée qui permettrait de reproduire la donnée à partir des données brutes et de workflows associés (Machine actionnable). Cependant ceci n'est pas facile à mettre en œuvre et ce n'est pas la priorité qui doit être mise en avant pour le moment.

Il existe plusieurs projets et communautés qui sont liés et travaillent sur les principes FAIR. On peut citer entre autres : la Research Data Alliance (RDA – <https://www.rd-alliance.org/groups/rda-france>⁵), l'initiative européenne GO FAIR (<https://www.go-fair.org>⁶), ou la communauté de données pour la science et la technologie (CODATA – <https://codata.org>⁷).

Un point qui est de plus en plus important est d'arriver à évaluer la maturité FAIR des projets et des données pour cela RDA a mis en place un groupe de travail sur le FAIR Data Maturity Model. Ce groupe a pour objectif de définir des indicateurs communs correspondant aux principes FAIR et qui pourront être utilisables pour l'évaluer le degré de « FAIRness » (voir <https://www.rd->

⁵ <https://www.rd-alliance.org/groups/rda-france>

⁶ <https://www.go-fair.org>

⁷ <https://codata.org>

alliance.org/groups/fair-data-maturity-model-wg⁸) C'est un fichier Excel regroupant tous les indicateurs et permettant d'évaluer son degré de maturité FAIR.

Une autre initiative a aussi été initiée afin d'évaluer le degré de maturité FAIR par le projet GO FAIR. Les « FAIR Implementation Profiles (FIP) » sont des questionnaires en ligne permettant de déclarer les ressources, etc. Ces FIPs sont « machine actionnable », ils permettent d'être rempli par des communautés et de faire des matrices de convergence pour avoir une idée de ce qui est utilisé par ces communautés.

Enfin il existe aussi FAIRsFAIR qui est projet qui a une initiative de CoreTrustSeal (CTS) et FAIR, qui a pour objectif de lier les principes FAIR aux exigences CTS et de proposer une nouvelle version des exigences CTS. Il existe un document d'overview consultable (voir <https://zenodo.org/record/4003630#.YAGp6uhKhPa>⁹). Le document final est en cours de rédaction et proposera un mapping entre les principes FAIR et les exigences de CTS.

3.2. FAIR dans Data Terra – (Damien BOULANGER)

DB présente le FAIR dans Data Terra (DT - voir : [202012_ODATIS_Atelier_Boulangier_FairDataTerra.pdf](#)¹⁰). L'Infrastructure de Recherche Data Terra (IRDT) est au tout début de la mise en place de ces principes FAIR. Il existe des groupes de travail dans l'IR qui prennent en compte ces principes, en particulier :

- GT Catalogue
 - Mise en place d'un portail de données et de connaissances pour toute la communauté Data Terra,
 - Mise en place de format pivot basé sur l'ontologie SOSA (provient du projet PHIDIAS),
 - Inventaire des vocabulaires, avec un alignement de ces vocabulaires et une publication,
 - Evaluation de la maturité FAIR (le choix entre FAIR Data Maturity Model ou FIP n'est pas encore décidé) ;
- GT Authentification
 - Outil d'authentification centralisé à partir des solutions implémentées par AERIS et THEIA,
 - L'étude est en cours ;

⁸ <https://www.rd-alliance.org/groups/fair-data-maturity-model-wg>

⁹ <https://zenodo.org/record/4003630#.YAGp6uhKhPa>

¹⁰ https://www.odatis-ocean.fr/fileadmin/documents/activites/ateliers/atelier_202012/202012_ODATIS_Atelier_Boulangier_FairDataTerra.pdf
CR atelier technique décembre 2020

- GT ma-DMP
 - Profils type des projets,
 - Description des offres de services des CDS,
 - Intégration dans le futur appel à projet de Data Terra,
 - Etude en cours des outils : RDA DMP Common standard et DS Wizard.

L'ensemble de ces groupes de travail s'intègrent dans le futur projet du PIA3/Equipex+ GAIA DATA qui a été accepté fin décembre. Ce projet va permettre de mettre en place l'infrastructure distribuée de données et services pour l'observation, la modélisation et la compréhension du système terre, de la biodiversité et de l'environnement. Ce projet est piloté par Data Terra et est une collaboration de cet IR avec le PNDB (pour la partie biodiversité) ainsi que l'IR CLIMERI (pour la partie modélisation). Ce projet va développer plusieurs services pour la communauté scientifique :

- Services de découverte, d'accès et de gestion des données
 - Catalogue (métadonnées, vocabulaires, ontologie), système d'accès et de recherche transdisciplinaire,
 - Archive à long terme, entrepôts, DOI, services avancés de visualisation,
 - Aide à la collecte des données des observatoires ;
- Services transversaux pour faciliter les travaux transdisciplinaires
 - Grille de données, cloud, standardisation de la production, portail de connaissances,
 - Support utilisateur,
 - Formations ;
- Services – Virtual Analysis Platform (VAP)
 - Travail collaborative,
 - Bac à sable, développement ;
- Services d'analyse des données à la demande et Virtual Research Environnement (VRE)
 - Interface interactive,
 - Exécution de scripts à la volée,
 - Définition et exécution de workflows de traitement spécifiques des domaines sur les 8 centres structurants.



Pour reprendre les principes FAIR :

- F1 : IRDT va permettre de mettre des DOI sur les jeux de données,
- F2 : Format de métadonnées pivot DT en cours de construction,
- F3 : Portail de données et de connaissances et Entrepôt de données prévus dans GAIA DATA,
- A1.1 : Format pivot DT et serveurs THREDD prévus dans plusieurs centres de DT,
- A1.2 : GT authentification,
- A2 : GT ma-DMP,
- I1 : Mapping du format pivot sur plusieurs standards,
- I2 : les vocabulaires sont en train d'être standardisés dans le GT catalogue,
- I3 : Prévus dans le modèle pivot,
- R1.1 : GT Interpôle sur les licences mais il est à noter qu'il ne va pas être possible d'avoir une licence identique pour tous les jeux de données,
- R1.2 : le sujet de tracer la provenance n'a pas encore été discuté (application du standard PROV à discuter),
- R1.3 : plusieurs mapping sont prévus entre ISO, DataCite, Etc.

3.3. ENVRI-FAIR partie Marine Subdomain – (Damien BOULANGER, Valérie HARSCOAT & Thierry CARVAL)

3.3.1. Introduction générale du projet ENVRI-FAIR – (Damien BOULANGER)

DB présente le projet ENVironnemental Research Infrastructures building Fair services Accessible for society, Innovation and Research (ENVRI-FAIR - voir [202012_ODATIS_Atelier_Boulanger_EnvriFair.pdf](https://www.odatis-ocean.fr/fileadmin/documents/activites/ateliers/atelier_202012/202012_ODATIS_Atelier_Boulanger_EnvriFair.pdf)¹¹) qui regroupe toutes les Infrastructures de Recherche (IR / *En anglais : Research Infrastructure – RI*) européennes dédiées au système terre. Sa mission est de partager des expériences et de trouver des solutions communes pour le partage l'accessibilité, l'interopérabilité de données, mais aussi de développer des services communs afin de les exposer comme services de l'European Open Science Cloud (EOSC). Enfin ENVRI-FAIR doit préparer ENVRI-Hub qui sera un Hub connecté à EOSC.

Historiquement ENVRI-FAIR succède à deux autres projets européens (Figure 1) que sont :

- ENVRI (qui a permis de construire la communauté et de définir un cadre commun aux infrastructures et de développer un ENVRI reference Model),
- ENVRI-Plus (qui a permis de mettre en place les premiers développements communs avec des solutions partagées et un service portfolio qui a été développé).

¹¹https://www.odatis-ocean.fr/fileadmin/documents/activites/ateliers/atelier_202012/202012_ODATIS_Atelier_Boulanger_EnvriFair.pdf
CR atelier technique décembre 2020



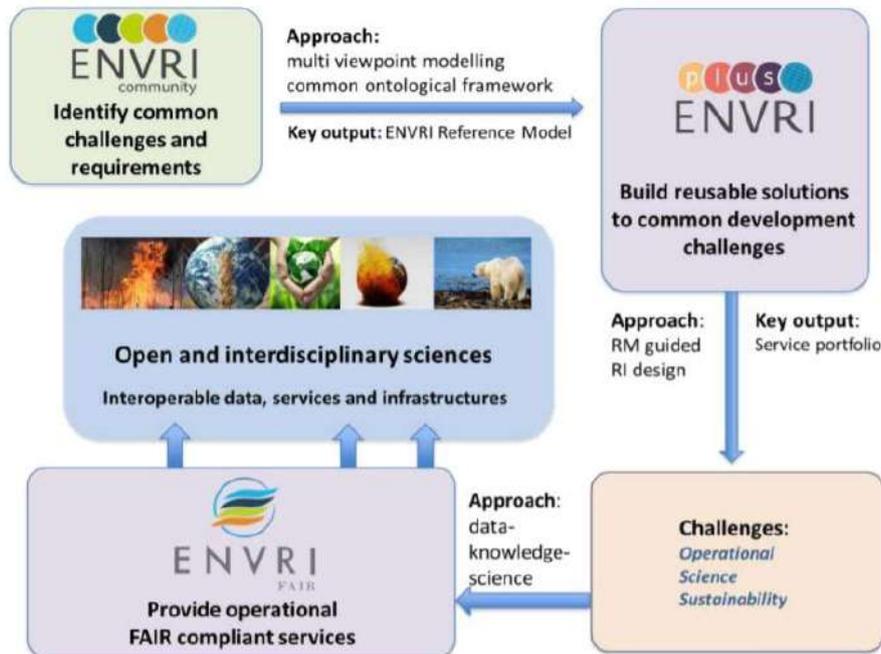


Figure 1. Les différents projets ENVRI

L'objectif principale d'ENVRI-FAIR est appliquer et d'implémenter les principes FAIR dans les IR et de lier ENVRI à EOSC (Figure 2). C'est un projet qui regroupe 37 partenaires et 13 IR qui sont sur la feuille de route ESFRI ou Landmark. Ce projet de l'appel d'offre INFRAEOSC-04-2018 a démarré en 2019 pour une durée de 4 ans (2019 – 2022) avec un budget de 19 M€. Ses coordinateurs principaux sont Andreas Petzold du ForschungsZentrum Jülich en Allemagne et Ari Asmi de l'University of Helsinki en Finlande. La mise en œuvre des principes FAIR est effectué par les IR elles-mêmes. Les standards et l'implémentation sont hiérarchisés au niveau des clusters, des sous-domaines et des IRs sous forme de blocs (Figure 2) et qui se déclinent sous forme de différents WPs (Figure 3).

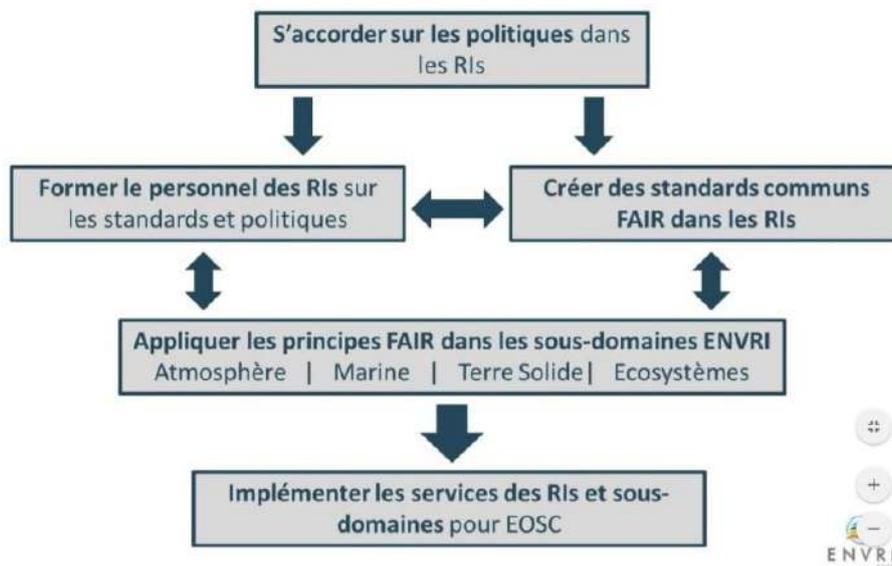


Figure 2. Principes d'implémentation du projet ENVRI-FAIR

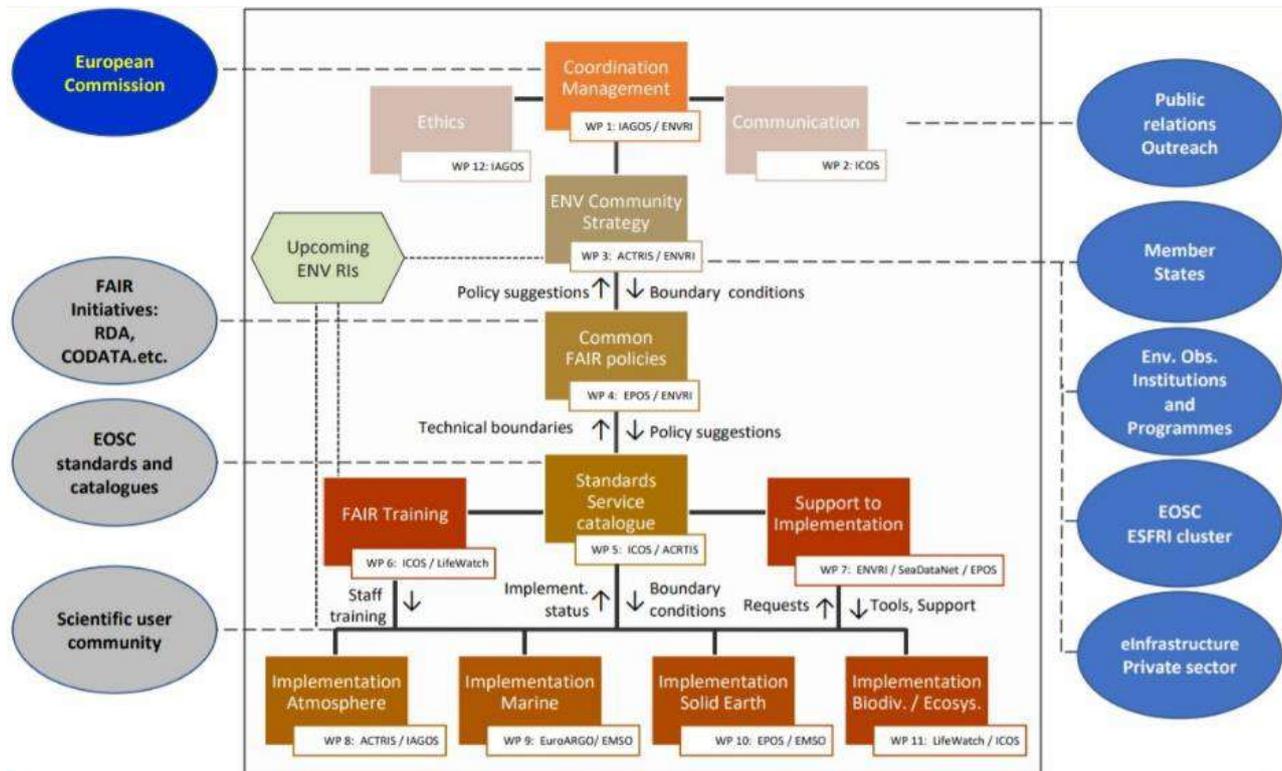


Figure 3. Les WPs du projet ENVRI-FAIR

Le WP9 s'intéresse plus particulièrement au domaine marin et ce WP est copiloté par TC.

La première étape du projet a été d'évaluer la maturité FAIR de chaque IR au moyen de fichier Excel et dans l'avenir il est prévu d'utiliser les FIPs décrit dans la session 3.3 sur les principes FAIR. Des groupes de travail ont été mis en place sur 6 thématiques :

- Design of the ENVRI Catalogue of services,
- AAI implementation,
- Persistent Identifiers,
- Triple stores and data storage certification,
- Licences, citation and usage tracking,
- ENVRI-Hub design and architecture.

3.3.2. Introduction du WP9 Marine subdomain – (Valérie HARSCOAT)

VH introduit le WP9 d'ENVRI-FAIR qui s'intéresse au sous domaine marin (voir [202012_ODATIS_Atelier_Harscoat_EnvriFair.pdf](https://www.odatis-ocean.fr/fileadmin/documents/activites/ateliers/atelier_202012/202012_ODATIS_Atelier_Harscoat_EnvriFair.pdf)¹²).

Plusieurs infrastructures participent à ce WP9 : EURO-Argo, ICOS (Marine), EMSO, LifeWatch (Marine) et SeaDataNet. Ce WP est coordonné par IR EURO-ARGO et EMSO avec le support d'IFREMER. Les objectifs principaux de ce WP sont :

- D'améliorer le FAIRness dans les 5 IR concernées en prenant en compte les priorités des différents « utilisateurs » et en préparant l'interopérabilité des IRs par des vocabulaires communs pour les (méta)données, ceci afin d'aller vers une gestion plus coordonnée et plus homogène des données marines au niveau européen ;
- De piloter ce WP par les EOVS pour construire un système de données marines de qualité, interopérable et réutilisable, grâce à des objectifs pratiques communs via une analyse de la FAIRness ;
- D'implémenter dans chaque IR les améliorations FAIR pour les 3 niveaux (« utilisateurs », sous domaine marin, et ENVRI).

Pour chaque IR, la première phase a été de mener une analyse de la FAIRness (auto-analyse en termes de forces et de faiblesses et réponses au questionnaire sur les critères FAIR provenant du projet GO-FAIR). Cela a permis de mettre en place une feuille de route pour chaque IR pour améliorer la FAIRness (activités, priorités, planning). La deuxième phase a été consacrée à la spécification technique détaillée des améliorations propres à chaque IR en tenant compte des exigences pour le démonstrateur du sous domaine. La troisième phase est en cours et doit implémenter les améliorations dans chaque IR. Enfin la quatrième et dernière phase, sera dédiée à la mise en place d'un démonstrateur de la FAIRness du sous-domaine.

Le questionnaire a permis de comprendre les principes FAIR et leurs avantages pour les IR, d'évaluer le statu quo des données et services de chaque IR en termes de FAIR. Ce questionnaire a mis en lumière les lacunes et les points forts en matière d'information et de mise en œuvre, de comparer les différentes implémentations des IRs, d'évaluer et de prioriser les possibilités d'adoption de technologies pour les améliorer. Enfin cela a permis d'inclure dans la feuille de route les améliorations FAIR choisies.

Une stratégie commune aux IRs a été adoptée qui vise à améliorer le « backoffice » (de machine-to-machine) de chaque IR et des APIs dédiées aux systèmes de gestion des données. Le démonstrateur « EOVS » va être un cas d'usage de l'utilisation des services machine-to-machine et va permettre d'agréger des données de différents IRs pour un utilisateur « VIP » (Copernicus et EDMO-NET) du service.

¹²https://www.odatis-ocean.fr/fileadmin/documents/activites/ateliers/atelier_202012/202012_ODATIS_Atelier_Harscoat_EnvriFair.pdf
CR atelier technique décembre 2020



Un exemple de l'analyse FAIRness est donnée pour l'IR Euro-ARGO, voir la Figure 4 (slide 7 de la présentation de VH). Les faiblesses étant surtout concentrées au niveau de l'accès machine-to-machine, la Figure 5 représente le plan d'implémentation ENVRI-FAIR pour Euro-ARGO. Les cases en orange montrent les implémentations nécessaires pour obtenir des améliorations de la communication des informations au niveau du machine-to-machine.

- Les forces
 - Accès très ouvert aux données et pas d'enregistrement des utilisateurs
 - Prise en charge de nombreux protocoles d'accès aux données et aux métadonnées
 - Analyse de l'utilisation des données (via Spark, Cassandra, Elastic)
- Les faiblesses et les améliorations identifiées

	Faiblesses	Améliorations
FINDABILITY	le jeu de données global est facile à trouver et accessible pour un humain, mais pas de machine à machine	Mettre en place un service de moteur de recherche pour interroger le jeu de données Argo
ACCESSIBILITY	le jeu de données global est facile à trouver et accessible pour un humain, mais pas de machine à machine	Implémenter des API pour les accès machines à machines pour les données et les métadonnées Implémenter une API de description facilitant l'utilisation d'une API
INTEROPERABILITY and REUSABILITY	Des métadonnées détaillées mais certaines seulement décrites dans le manuel utilisateur, or l'utilisation de vocabulaire est indispensable pour faciliter les interfaces machine à machine et l'agrégation de données au niveau sous-domaine Un DOI pour le dataset complet Argo et des snapshots mensuels, mais seulement un humain peut identifier un cycle et un paramètre	Mise en place d'un vocabulaire pour toutes les tables du manuel utilisateur et les tables de configuration techniques Argo ; vocabulaire géré par le serveur NVS du BODC Attribuer un identifiant permanent PID à un paramètre acquis pendant un cycle d'un flotteur



Figure 4. Analyse FAIRness, l'exemple de l'IR Euro-ARGO.

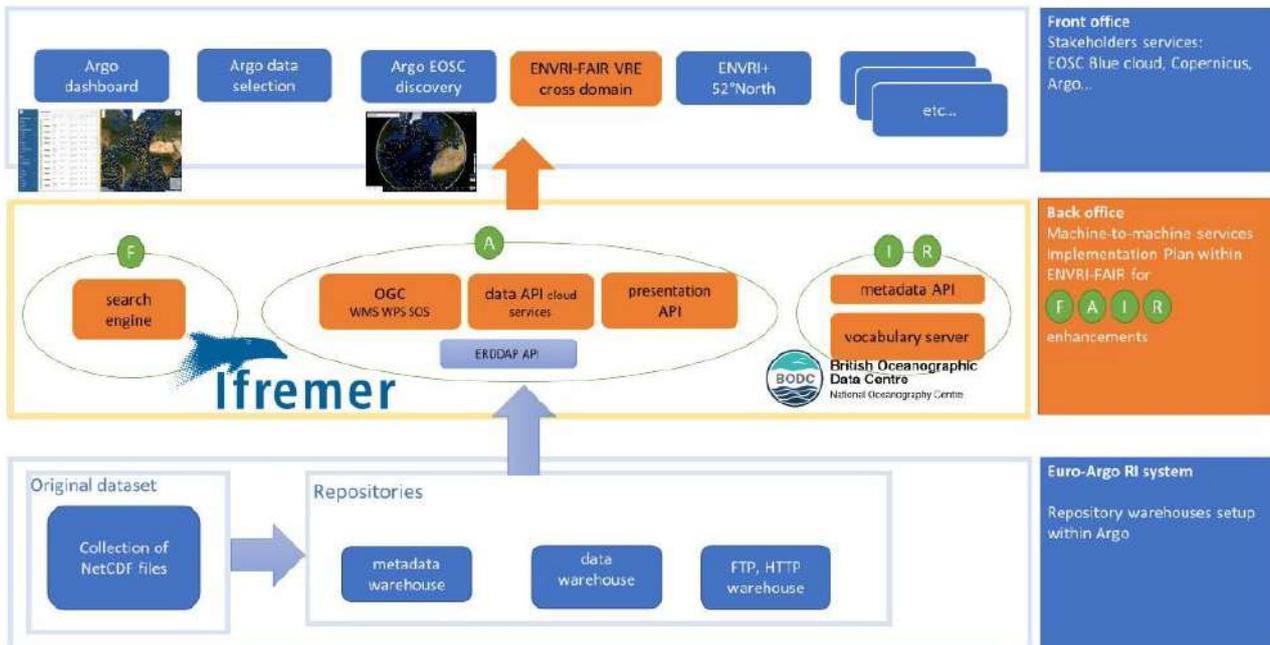


Figure 5. Plan d'implémentation ENVRI-FAIR pour Euro-Argo. Les cases en orange représentent les futures améliorations à implémenter.

3.3.3. Améliorations des principes FAIR dans Euro-ARGO – (Thierry CARVAL)

TC présente les implémentations qui ont été fait au niveau de l'IR Euro-ARGO afin d'améliorer la FAIRitude des données (voir [202012_ODATIS_Atelier_Carval_ArgoFair.pdf](https://fileadmin/documents/activites/ateliers/atelier_202012/202012_ODATIS_Atelier_Carval_ArgoFair.pdf)¹³). Une démarche parallèle est aussi effectuée dans les autres IR tel que SeaDataNet. L'idée était de proposer des services d'accès aux données et métadonnées qui n'existaient pas ou d'en améliorer l'efficacité dans les différentes IR européennes. Pour Argo, un moteur de recherche (OpenSearch) a été mis en place avec une API efficace et robuste sur les données et les métadonnées. Ce moteur est disponible via une image Docker et permet d'être utilisé sur des données in-situ et satellites en mode cloud. Les requêtes se présentent sous un format d'URL qui génère une page XML au format atom. Cette page contient toutes les URLs et les services d'accès à la donnée (FTP, descriptions des flotteurs sélectionnés et de leurs données). Le protocole Open API a permis de décrire des (méta)données que ce soit pour les humains mais aussi pour les machines. Un service ERDDAP a aussi été mis en place pour l'accès aux (méta)données. Des services OGC sont en cours d'implémentation (WMS, SensorThings, WPS (ENVRI-FAIR VRE) ainsi qu'entrer le vocabulaire d'ARGO dans le serveur de vocabulaire du BODC avec une ontologie associée permettant de donner du sens au vocabulaire.

Grace à ENVRI-FAIR, un tableau de bord des flotteurs a été développé avec un accès extrêmement rapide et robuste à l'ensemble des métadonnées du jeu de données des flotteurs ARGO (voir [dashboard](https://fleetmonitoring.euro-argo.eu/dashboard)¹⁴). Une autre API a aussi été développée qui permet, en fonction des positions des flotteurs ARGO, de voir la donnée des profils du flotteur (<https://dataselection.euro-argo.eu/>¹⁵). Cette API permet aussi de superposer d'autres données comme l'étendue des glaces, etc., et aussi d'exporter des données via un service d'extraction en batch.

Il est à noter que ces APIs travaillent sur plusieurs millions de métadonnées et sur un ensemble de 2 milliards d'observations disponibles. Sur ces APIs il est aussi possible de rajouter des services de l'OGC plus riches comme le WMS, SensorThings, WPS... Ces travaux sont encore en cours.

Les tables de vocabulaire ARGO sont servi par le serveur de vocabulaire du BODC. Il y a 25 tables de vocabulaire en ligne qui sont en « R » (voir [vocabulary_search/](https://www.bodc.ac.uk/resources/vocabularies/vocabulary_search/)¹⁷ ; [vocabulary_search/R03/](https://www.bodc.ac.uk/resources/vocabularies/vocabulary_search/R03/)¹⁸). Comme les vocabulaires sont évolutifs, ils sont gérés avec NVS-VOCABS GitHub ([ArgoVocabs](https://github.com/nvs-vocabs/ArgoVocabs)¹⁹).

Il est aussi nécessaire de développer l'ontologie ARGO qui est une sémantique sur les vocabulaires Argo. Cette ontologie en cours de création est une ontologie RDF conforme au modèle DCAT (EU) et à Schema.org (Google). Elle sera donc moissonnable par les catalogues DCAT et Google Data Search, interrogeable par un serveur Sparql endpoint, et s'intégrera dans la fédération des infrastructures de recherche du cluster ENVRI. Enfin elle va être le point d'entrée ARGO pour tout sur-ensemble sémantique RDF (Figure 6).

¹³https://www.odatis-ocean.fr/fileadmin/documents/activites/ateliers/atelier_202012/202012_ODATIS_Atelier_Carval_ArgoFair.pdf

¹⁴ <https://fleetmonitoring.euro-argo.eu/dashboard>

¹⁵ <https://dataselection.euro-argo.eu/>

¹⁶ <https://dataselection.euro-argo.eu/>

¹⁷ https://www.bodc.ac.uk/resources/vocabularies/vocabulary_search/

¹⁸ https://www.bodc.ac.uk/resources/vocabularies/vocabulary_search/R03/

¹⁹ <https://github.com/nvs-vocabs/ArgoVocabs>

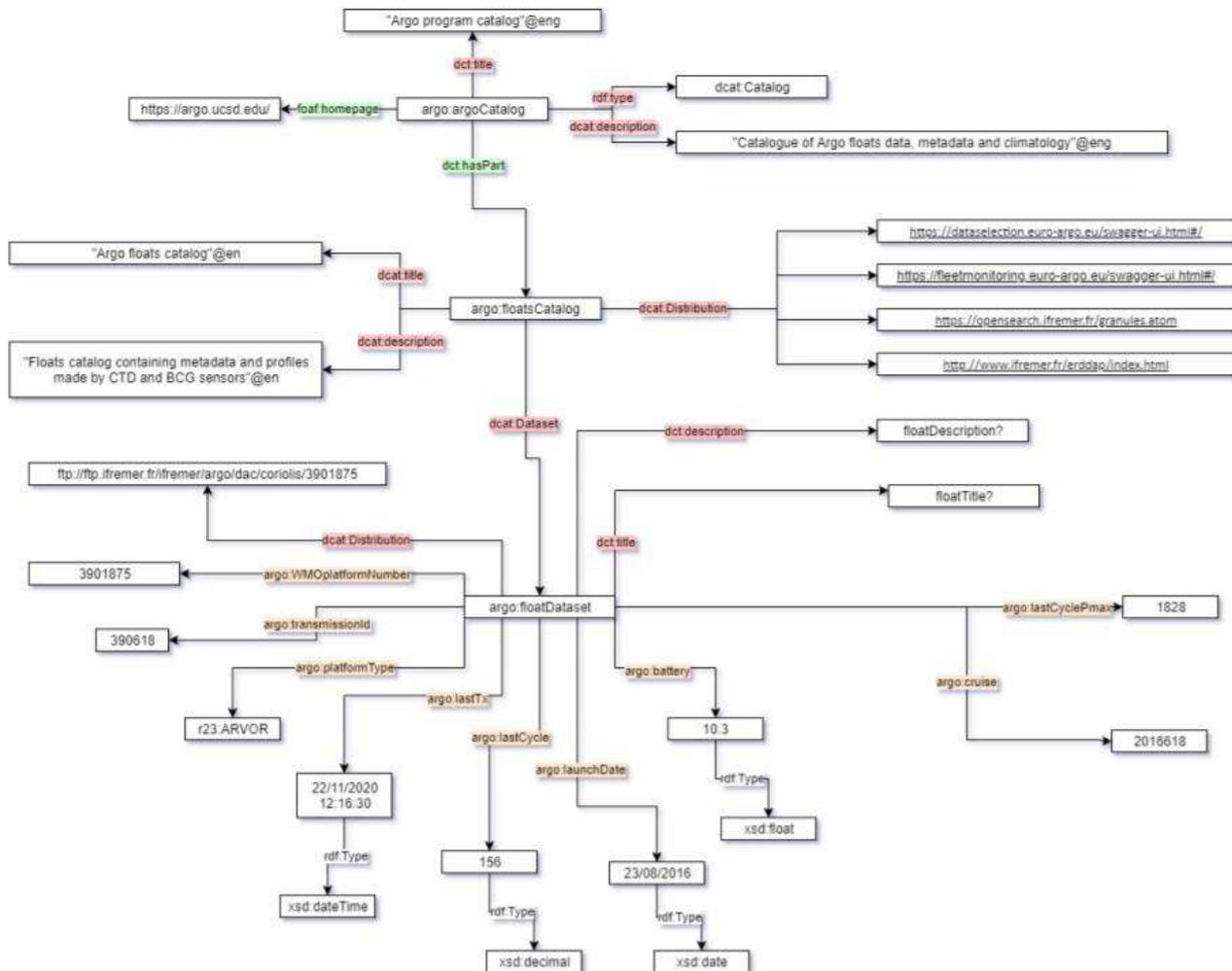


Figure 6. Schéma de l'ontologie ARGO

3.4. Atmosphère - AERIS – (Damien BOULANGER)

DB présente l'implémentation des principes FAIR dans le pôle AERIS (voir 202012_ODATIS_Atelier_Boulanger_FairAeris.pdf²⁰) en reprenant la liste des principes FAIR :

- F1 :
 - Assignation de DOI pour les jeux de données finaux que le pôle veut rendre citables ;
 - Il est prévu d'utiliser aussi les PID ePIC (basés sur les handles) pour les jeux de données internes (gestion de workflow en interne sur les jeux de données brutes ou intermédiaires par exemple). Ces PID vont être utilisés dans ENVRI-FAIR pour IAGOS et ACTRIS.

²⁰https://www.odatis-ocean.fr/fileadmin/documents/activites/ateliers/atelier_202012/202012_ODATIS_Atelier_Boulanger_FairAeris.pdf
CR atelier technique décembre 2020





- F2 :
 - AERIS s'est doté d'un format de métadonnées pivot ;
 - Mapping de ce format avec ISO 19115/INSPIRE et DataCite ;
 - Dans ENVRI-FAIR il est prévu de faire aussi le mapping vers les profils de métadonnées WMO notamment le WMO core metadata profile pour WIS et pour WIGOS.
- F3 :
 - Catalogue de (méta)données AERIS ;
 - Ce catalogue est aussi dans DataCite via les DOI ;
 - L'implémentation des fiches de métadonnées est en cours sur schema.org afin d'indexer ces fiches sur Google Data Search ;
 - Il y a aussi des travaux sur la recherche sémantique dans ENVRI-FAIR.
- A1.1 :
 - Utilisation de Geonetwork pour le moissonnage des métadonnées : CSW ou OAI-PMH ;
 - Implémentation de serveurs THREDDs en préparation pour fournir les données via OPeNDAP, WCS, etc.
- A1.2 :
 - SSO AERIS basé sur la solution Keycloak pour l'authentification ;
 - Authentification via ORCID ou EduGAIN ;
 - Gestion fine et centralisée des autorisations pour toutes les applications AERIS (solution en cours de déploiement).
- A2 :
 - DMP en lien avec le GT Data Terra.
- I1 :
 - Profils de métadonnées ISO 19115 disponibles dans plusieurs projets.
- I2 :
 - Thésaurus AERIS basés sur GCMD (plateformes, instruments et variables) et les noms standard de la convention CF ;
 - Publication de ces vocabulaires en cours avec SKOSMOS.
- R1 :
 - Certaines métadonnées sont obligatoires pour les producteurs de données ;
 - Il est prévu pour les PI de fournir des fiches mieux renseignées pour qu'elles soient plus lisibles.



- R1.1 :
 - Plusieurs types de licences selon les projets ;
 - Besoin de rendre ces fiches machine-actionnable en se basant sur ce qui va être proposé dans ENVRI-FAIR.
- R1.2 :
 - Travail sur la provenance très peu avancé dans la communauté AERIS mais il est prévu, dans le cadre d'ENVRI-FAIR (IAGOS, ACTRIS), en utilisant l'ontologie PROV-O (standard W3C), et aussi d'utiliser les PIDINST de RDA pour fournir les métadonnées des instruments.
- R1.3 :
 - ISO 19115/INSPIRE ;
 - WMO metadata profiles.

Le catalogue AERIS est consultable en suivant ce lien : <https://www.aeris-data.fr/catalogue/22>. Ce catalogue est une nouvelle version en cours de développement qui va permettre aux PI d'éditer en ligne leurs métadonnées. La fiche est remplie en ligne et une fois cette fiche validée, il est possible d'attribuer un DOI qui est ensuite introduit dans cette même fiche.

La certification des entrepôts pour IAGOS et ECCAD est en cours et permettra d'avoir un modèle pour les autres projets.

Certains travaux sont effectués sur la citation dynamique par la mise en place d'un Querystore où les requêtes des utilisateurs vont être enregistrées dans une base centrale pour tous les projets AERIS. Des PID seront mis sur chacune des requêtes en utilisant des fragments pour générer les landing page spécifique à la requête. Les landing pages seront filtrés par les fragments.

3.5. Discussion à la suite de la Session FAIR

GM lance la discussion et précise que les travaux qui ont été entrepris sur les données ARGO sont l'objectif type à avoir sur l'ensemble des jeux de données et ces travaux sont en pointe dans ce qui se fait actuellement sur les données et métadonnées. Cet objectif est non seulement celui d'ODATIS mais aussi de l'ensemble des pôles et de IR Data Terra en général. L'objectif est de pouvoir visualiser, interroger, moissonner, superposer, télécharger des données provenant de différentes sources. Pour cela nous avons besoin de nous améliorer sur les métadonnées riches (comme DB l'a présenté), mais aussi sur les services d'accès comme l'a montré TC et sur la description des services pour qu'ils puissent être actionnables par un humain ou via une machine (norme OpenAPI, ou OPENDAP qui aide à cela). Les services OPENDAP et ERDDAP ont été évoqués dans les ateliers précédents. Les métadonnées qui sont dans le catalogue ODATIS sont élaborées et moissonnées dans divers centres de données et on voit bien une différence de complétude et de richesse entre les différentes fiches de métadonnées qui sont dans le catalogue ODATIS. Dans le cadre du projet PHIDIAS, une étude a

²¹ <https://www.aeris-data.fr/catalogue/>

²² <https://www.aeris-data.fr/catalogue/>

été menée pour regarder le degré de FAIRness (ou FAIRitude) de ces fiches en utilisant le modèle pivot SOSA qui a été évoqué par DB et qui serait le modèle commun pour l'IR Data Terra. Par exemple, pour décrire les métadonnées et les données observées dans les fiches, il est utilisé la norme ISO 19115 avec des mots clés. Cependant derrière ces mots clés, la référence n'est souvent pas mise pour décrire le mot clé, et une machine est incapable de trouver la donnée qui est présentée avec ce mot clé si la référence n'est pas mise (ex. La température : si la référence n'est pas mise sur ce mot clé une machine est incapable de trouver la donnée de température associée au mot clé). Il est donc nécessaire d'aller plus loin que le mot clé (qu'un humain peut facilement interpréter) pour qu'une machine soit capable de l'interpréter en tant que variable observée. Il y a donc un niveau de complétude dans données à améliorer si l'on veut rendre machine actionnable les jeux de données. Dans 80% des fiches d'ODATIS nous avons ce type de problème.

La deuxième remarque de GM est sur la certification CoreTrustSeal et les principes FAIR. On voit bien que beaucoup des principes FAIR se retrouvent dans CTS et inversement. Cet aspect est extrêmement important pour chaque CDS car dans le cahier des charges ODATIS pour les CDS, chaque CDS doit respecter les principes FAIR et les engagements de la certification CTS. Bien que dans le cahier des charges ces deux notions sont distinctes, il existe un mapping de ces deux notions qui doit être pris en compte dans chaque CDS. Ces points doivent être travailler en commun et vont certainement faire l'objet d'autres ateliers ODATIS.

Un Google doc a été mis à disposition des participants pour lister les questions et afin que les réponses soient mises à disposition pour une lecture ultérieure (voir ANNEXES 6.1).

4. Sessions identifiants permanents

4.1. PID dans et pour ODATIS et COPiLoTe – (Gilbert MAUDIRE)

GM présente une introduction aux identifiants permanents (Persistent Identifier - PID) pour le pôle ODATIS. Les PID sont un des éléments indispensables aux principes FAIR et ils sont très importants dans le contexte du pôle ODATIS. L'idée est de faire adopter dans tous les centres du pôle ODATIS et même au-delà (car cela concerne l'IR Data Terra dans son ensemble) des principes homogènes. Cette adoption permettra de faciliter l'accès aux données pour les utilisateurs. L'homogénéité concerne par exemple la publication des données qui doit se faire avec la mise en place d'un DOI sur le jeu de données afin d'en améliorer la traçabilité. Les landing page associées au DOI doivent être le plus homogènes possibles où l'on doit trouver à minima les mêmes informations (métadonnées riches, niveau de granularité du jeu de données, etc.). Plusieurs sujets sont encore en évolution au sujet des PID/DOI, comme par exemple est-ce que l'on met des DOI sur des requêtes de données afin de pouvoir facilement rejouer le même jeu de données ?

L'objectif est donc d'harmoniser nos pratiques au sujet des PID et de la mise en place de DOI sur les jeux de données.

4.2. Principes, Guidelines pour différents jeux de données & bibliométrie – (Frédéric MERCEUR)

FM présente des informations générales sur les PID, sur ORCID, les DOI de données et SEANOE (voir [202012_ODATIS_Atelier_Merceur_PID.pdf](#)²³).

Un PID est un identifiant pérenne d'une ressource. Cette chaîne de caractère est unique au niveau mondial et est associée à un résolveur. Cette chaîne de caractère et ce résolveur forment une URL qui est redirigée de façon transparente pour un usagé soit vers la ressource elle-même, soit vers une landing page (page d'atterrissage) qui va décrire la ressource et qui va fournir un lien vers cette ressource. Ce lien est pérenne car il est possible de mettre à jour le lien de redirection auprès de la base de données du PID. Un PID peut-être attribué (à quasiment tout !) à des documents, des données, des personnes, des échantillons, des expériences, des logiciels, des modèles, des définitions de variables, ... Les PID sont, bien entendu, indispensables si on veut mettre en place des principes FAIR.

Il existe plusieurs systèmes de PID comme :

- DOI,
- ORCID,
- ROR,
- Handle,
- IGSN,
- ...

Comment choisir un PID ?

Il est conseillé de lire la ressource suivante : Madden, Frances, van Horik, René, van de Sandt, Stephanie, Lavasa, Artemis, & Cousijn, Helena. (2020, May 28). *Guides to Choosing Persistent Identifiers Version 3* (Version 3). Zenodo. <http://doi.org/10.5281/zenodo.4192174>²⁴.

Les critères de choix doivent être pris en fonction de la pérennité de l'organisme qui porte le PID, choix de la communauté, la reconnaissance du PID par les éditeurs de revues, la conformité du schéma de métadonnées, ... Dans certains cas les choix sont assez simples comme pour la citation publique de jeux de données ou d'un document : le DOI s'impose, ou pour l'identification publique de contributeurs l'ORCID devient incontournable. Cependant le choix peut être plus compliqué comme pour l'identification d'organismes biologiques (DOI, FAO, ...).

ORCID est un système de PID qui permet d'identifier et de distinguer les contributeurs de la recherche (scientifiques, ingénieurs, techniciens, etc.). L'identifiant ORCID est une liste de 4 blocs

²³https://www.odatis-ocean.fr/fileadmin/documents/activites/ateliers/atelier_202012/202012_ODATIS_Atelier_Merceur_PID.pdf

²⁴ <http://doi.org/10.5281/zenodo.4192174>

de 4 numéros (et lettres), ex : [0000-0001-8876-5179](https://orcid.org/0000-0001-8876-5179)²⁵. Ce système est porté par un organisme à but non lucratif basé aux Etats-Unis d'Amérique. Un consortium « Communauté française ORCID » est lancé en octobre 2019 par le gouvernement français. Ce consortium est porté par contrat et inclue 36 établissements français (dont les établissements de recherche français) afin de faciliter l'implémentation de l'identifiant dans les établissements. La landing page d'un ORCID comprend le nom de la personne ainsi que son employeur et la plupart du temps la liste de ces publications. Il est à noter que l'obtention d'un ORCID est une démarche personnelle et qu'un établissement ne peut pas contraindre un employé à en avoir un. Un problème est qu'il existe de nombreuses landing pages vides dans ORCID. Pour éviter ce problème, l'IFREMER a développé à partir des API d'ORCID un formulaire où l'utilisateur autorise ARCHIMER (la base de données documentaire de l'IFREMER pour son personnel) à venir alimenter et mettre à jour automatiquement sa page ORCID. Les informations sont mises à jour quotidiennement pour chaque utilisateur.

Pour SEANOE, l'ensemble des dépôts bénéficie d'un DOI auprès de DataCite. De ce fait, au moment de l'envoi par SEANOE du remerciement du dépôt, le contributeur est informé que DataCite peut aller automatiquement alimenter son compte ORCID avec les dépôts effectués dans DataCite.

Un DOI peut être mis sur une publication mais aussi sur un jeu de données. Un DOI est une chaîne de caractère avec un préfixe qui est attribué à l'organisme qui va attribuer ce DOI et un suffixe. La combinaison des deux (préfixe, suffixe) est une chaîne unique au niveau mondial. Ce DOI comme tout PID est une redirection vers une URL pointant vers la page d'atterrissage. Pour les centres désirant attribuer des DOI, il est nécessaire de choisir une agence de DOI (CrossRef, DataCite, ...). Il existe une dizaine d'agences mais chaque agence est spécialisée. CrossRef est spécialisé dans l'attribution de DOI aux documents (ex : les publications). Pour attribuer des DOIs aux données, la plupart des centres passent par DataCite qui est spécialisé dans l'attribution de DOI aux jeux de données. Le représentant de DataCite en France est INIST qui permet de souscrire un contrat (sous forme d'un abonnement de 180€/an) pour attribuer un nombre illimité de DOI. Une fois le contrat souscrit, il est possible de demander un ou plusieurs préfixes. L'Ifremer a par exemple demandé 7 préfixes (pour SEANOE, le SISMER, ARCHIMER, etc.).

DataCite recommande de mettre un suffixe le plus court et plus opaque possible afin que le suffixe ne soit pas porteur de sens.

Pour activer un DOI, il y a deux possibilités :

- Soit le demandeur passe par des formulaires web proposés par DataCite,
- Soit le demandeur passe par l'API de DataCite.

Dans les deux cas, pour activer un DOI il est nécessaire de poster l'URL de la page d'atterrissage et de fournir 5 métadonnées obligatoires : DOI, Creator, Title, Publisher, Publication Year et un ensemble de métadonnées optionnel. L'URL et les métadonnées obligatoires doivent être maintenues à jour. La page d'atterrissage doit expliquer comment accéder à la donnée (URL,

²⁵ <https://orcid.org/0000-0001-8876-5179>

formulaire d'enregistrement, page de login, email, services de sélection des données), ou si la donnée n'est plus accessible. Il est important de noter qu'un DOI ne peut pas être supprimé.

Pour choisir un entrepôt pour publier un jeu de donnée, on peut :

- Vérifier la disponibilité d'un entrepôt institutionnel,
- Vérifier les recommandations des éditeurs d'articles (certains éditeurs imposent une liste d'entrepôts),
- Explorer les annuaires d'entrepôts (ex : <https://fairsharing.org/databases/>²⁶).

Comment choisir la granularité d'un DOI ?

Pour cela, il est conseillé de se souvenir de la finalité du DOI qui est la citation du jeu de données dans un article scientifique. Il est donc conseillé d'imaginer comment le jeu de données va être utilisé dans un article scientifique. Il est d'ailleurs possible d'attribuer un DOI à plusieurs niveaux d'un même jeu de donnée. Il faut donc réfléchir à la granularité du DOI à l'intérieur d'un jeu de données en fonction de la taille des fichiers, de leur utilisation, du nombre de fichiers, des paramètres, des régions d'étude, des versions, etc.

Pour la gestion des versions, le choix doit être effectué en fonction :

- Du besoin de citation spécifique des différentes versions,
- De la reproductibilité,
- Des contraintes techniques liées au volume des données,
- De l'entrepôt retenu,
- ...

Dans SEANOE, l'utilisateur n'a pas la possibilité de modifier lui-même le jeu de données. SEANOE a donc une discussion avec l'utilisateur pour voir ce qu'il veut modifier. Il y a donc plusieurs possibilités comme d'obtenir un nouveau DOI, ou de conserver le même DOI pour l'ensemble des versions. La deuxième solution est préconisée car elle permet de simplifier les consignes aux auteurs et d'avoir une seule page d'atterrissage pour les différentes versions ce qui donne une meilleure visibilité dans Google. De plus cela permet aux usagers d'accéder à la version la plus récente du jeu de données et de laisser les versions plus anciennes en accès à la demande. Il est aussi possible d'attribuer des fragments pour citer une version spécifique dans un seul DOI.

Pour le choix des auteurs, il est possible soit d'avoir une liste de personnes, soit une liste de collectivité d'auteurs (associés à une liste de contributeurs). Ce choix a une incidence sur le choix des versions car un auteur arrivant en cours de projet peut avoir la paternité de jeux de données précédent son arrivée (ce qui peut être problématique).

²⁶ <https://fairsharing.org/databases/>
CR atelier technique décembre 2020



SEANOE a été mis en place pour offrir une alternative à PANGAEA. SEANOE (<https://www.seanoe.org/28>) est :

- Un entrepôt de jeux de données marines recommandés par ODATIS, SeaDataNet, Elsevier, PLOS, ...,
- Un service accessible gratuitement à l'ensemble de la communauté internationale,
- Un service où les données sont obligatoirement en libre accès après un embargo éventuel de 2 ans avec une licence CC,
- Un service où tous les jeux de données disposent d'un DOI,
- Un service où les dépôts sont validés en 24h (ouvrables) maximum, avec un maximum de 100Go par dépôt,
- Un outil optimisé pour le référencement dans Google.

SEANOE permet d'avoir une veille sur les citations et sur les articles qui ont cités la donnée. Ce qui permet aussi d'avoir une analyse (pour les grands jeux de données) des citations sous forme graphique (voir planche 38). Si le jeu de donnée déposé dans SEANOE est intéressant pour être déposé dans EMODnet, celui-ci est automatiquement poussé via EMODnet ingestion qui permet au centre de données européen d'organiser son ingestion.

Un Google doc a été mis à disposition des participants pour lister les questions et afin que les réponses soient mises à disposition pour une lecture ultérieure (voir ANNEXES 6.2).

5. Synthèse des 2 jours des sessions de support par CDS

A la suite de cette demi-journée de présentations, des sessions individuelles d'une heure trente pour chaque CDS se sont succédées. Elles ont permis de prendre en compte chaque particularité de chaque CDS devant faire leur autoévaluation (CTS) dans l'ANR COPiLOtE. Une synthèse orale d'une demi-heure a été proposée à l'ensemble des CDS. Cette synthèse n'est pas mis à disposition dans ce compte rendu car elle revêt un caractère confidentiel interne aux participants du projet ANR COPiLOtE.

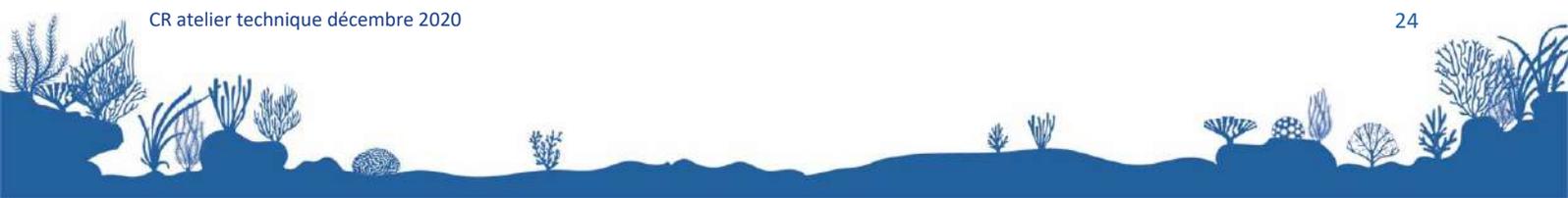
²⁷ <https://www.seanoe.org/>

²⁸ <https://www.seanoe.org/>

6. ANNEXES

6.1. Questions/Réponses sessions FAIR

<p>Dimitri : Ce projet ENVRI, est-ce qu'il y a des conséquences en terme d'obligations pour les centres de recherche</p>	<p>Damien : Les centres devront respecter certaines consignes, mais ce sera fait en collaboration, mais oui ça peut impliquer les scientifiques</p>
<p>Nicolas : Quelles sont les ressources humaines consacrées pour l'exemple Argo FAIR services enhancements qui vient d'être présenté par Thierry ?</p>	<p>Thierry: les API et le vocabulaire dans NVS sont financés par le projet ENVRI-FAIR, les IHM (tableau de bord, sélection de données) sont financés par les projets européens MOCCA et EA-RISE. Une dizaine de personnes travaillent à cette activité.</p> <p>ERDDAP évolue pour gérer les vocabulaires SeaDataNet, c'est un effort financé par la NOAA</p>
<p>Nicolas : Quelles sont les spécialités de la dizaine de personnes travaillant sur ce projet ?</p>	<p>Thierry: Les gestionnaires de vocabulaire, d'ontologie, sparql (BODC, SeaDataNet)</p> <p>Des data scientists (Euro-Argo)</p> <p>Des informaticiens (Ifremer, Altran, Capgemni)</p>
<p>Pascal : Quelle est la technologie utilisée pour le moteur à facette de https://fleetmonitoring.euro-argo.eu/dashboard?</p> <p>Pascal : Prévu que ce code soit partagé</p>	<p>FRED : Développement utilisant le mécanisme à facettes d'elastic search.</p> <p>Thierry : Tout ce qui est metadata est indexé par ElasticSearch. Et tout ce qui est data est indexé par Cassandra.</p> <p>Fred : Le code de l'appli global non, par contre développé sous formes d'applications Angular qui elles devraient l'être. Le 2nd composant c'est la gestion de la carte qui a été déposé par-dessus l'appli Leaflet.</p> <p>Thierry : Sur principe d'accord pour partager, le code pourrait être déposé dans un GitHub ODATIS ?</p> <p>Contact : Fred Merceur</p>

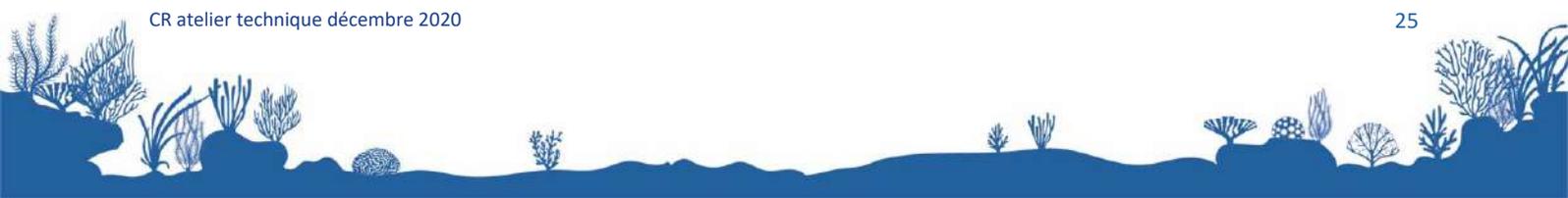




<p>Dominique : ARGO : L'ontologie Argo a-t-elle utilisée une ontologie "cadre"? Interpole semble vouloir utiliser SOSA. L'exercice a-t-il été fait d'essayer de l'utiliser?</p>	<p>On utilise les éléments qui nous paraissent pertinents : DCAT au niveau catalogue, DCAT-AP pour les services, FOAF et ORCID pour identifier des personnes, etc...</p> <p>Gilbert : SOSA est un méta modèle qui va plus loin dans la description des observations; utilisé dans Data Terra - DCAT est la norme au niveau européen EOSC mais a minima il faut utiliser Geo DCAT AP</p> <p>Damien : DCat AP est prévu d'être utilisé pour le catalogue ENVRI dans le projet Envri-FAIR</p>
<p>Dominique : DOI : pour les jeux de données "évolutifs" des fragments de DOI peuvent être utilisés, faut-il garder les données de chaque "version" pour permettre la reproductibilité des études déjà publiées?</p>	<p>Fred : j'en parle dans ma présentation à suivre</p> <p>Dominique : si je comprends bien c'est donc au cas par cas et un jeu de données doifié peut disparaître si mis à jour et l'ancienne version considérée comme non "importantes" ou trop "lourde"</p>

6.2. Questions/Réponses sessions identifiants permanents

<p>Est-ce que l'API ArchiMer va aussi dans la direction ORCID > Archimer. J'entends par là, si une information est mise à jour sur ORCID sera-t-elle transmise à ArchiMer ?</p>	<p>Fred : Non cela ne va que dans un sens. Message, vous déposez tout dans ArchiMer et nous on s'occupe de dupliquer là où nécessaire.</p> <p>Gilbert : Incitez à déposer au plus près (ODATIS/CDS/SeaNoe). Pour fédérer l'information à un nouveau de proximité et les données qui vont derrière.</p>
<p>Maurice: quelle est la position/politique de ODATIS sur la possibilité d'attribution de DOI par l'Inist ? pas mal d'instituts ont un contrat avec l'Inist qui attribue un préfixe de DOI... et donc on peut/pourrait ne pas passer par Seanoe par ce biais, c'est dommage, mais ça offre des possibilités pour identifier des données longue traîne).</p> <p>Maurice : Est-ce qu'on suit la recommandation ODATIS qui est de DOIser dans SEANOE. Est-ce</p>	<p>Fred : Tout le monde passe par l'INIST pour avoir un contrat DataCite. Pour attribuer des DOI DataCite, on a dû souscrire un contrat avec Inist.</p> <p>Laurent : Apparemment la politique de DataCite a changé, il faut passer directement par DataCite sans passer par l'Inist.</p> <p>Yvan : Bizarre car pour PNDB est en train de passer par l'Inist.</p>





qu'il ne faut pas donner des préconisations pour éclaircir la situation ?

Laurent : Dépend sans doute du type de structure. Inist sans doute passer par l'Inist que pour les institutions publiques.

Gilbert : Plusieurs éléments de réponse.

1) Data Terra va mettre en place un système de dépôt et d'assignation de DOI commun à l'ensemble des pôles pour ceux qui n'en n'ont pas encore. Pour l'Océan le besoin est déjà couvert par Seanoe, mais avec des échanges de metadata avec le système commun. C'est le BRGM qui est chargé de mettre en place ce système commun (neutre, inter organisme, pris en charge par DT).

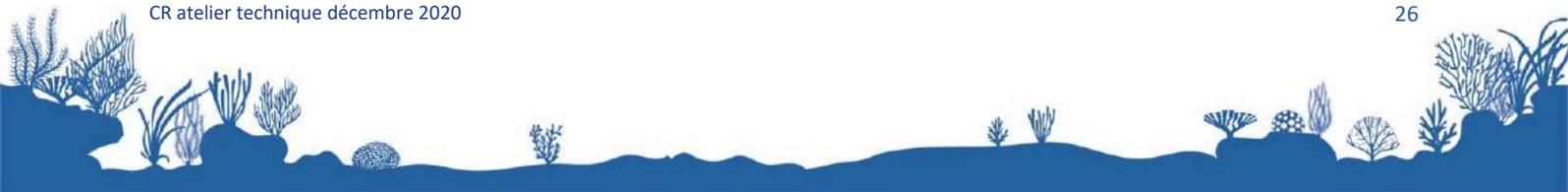
2). Il faut éviter d'avoir trop de systèmes et ne pas fractionner les choses. L'une des recommandations ODATIS/DT c'est de ne pas avoir une multiplicité des systèmes de dépôt. L'idée ce ne serait pas d'avoir plus d'un dépôt par organisme. Éviter d'avoir un système de dépôt par OSU et plutôt à l'Insu. Une autre recommandation, c'est que c'est que les DOI soient agnostiques, comme présenté par Fred, sans nom de structure (qui peuvent évoluer) afin d'être en concordance avec l'aspect de pérennité des DOI. DataCite recommande des DOI courts et opaques.

Yvan : Pour l'histoire de DOI, pouvoir référencer des identifiants alternatifs. Ex. jeux de données avec 2 DOI mais des metadata un peu différents. Importance de pouvoir les renseigner.

Gilbert : Lien entre DOI et catalogue des pôles. Le catalogue, à terme, devrait référencer des DOI, mais ça voudrait dire un alignement de la granularité entre catalogue et DOI. Quel est votre avis sur le sujet ? Source de complexité ?

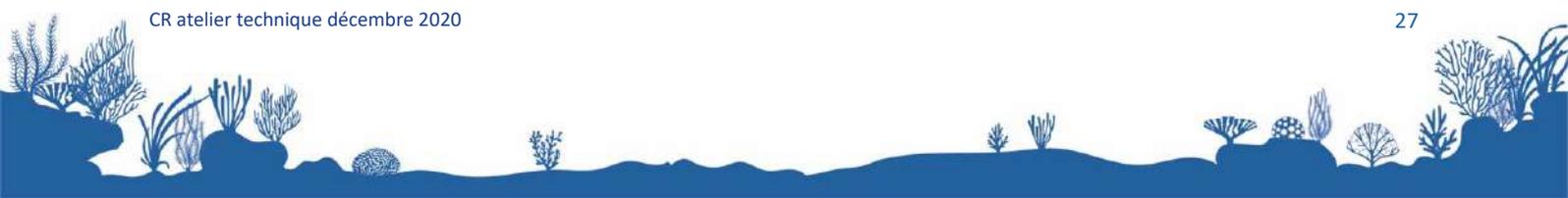
Yvan : Réflexion à pousser sur les fragments des jeux de données.

Joël : Remarque. Le CNRS (message JP Vilotte) voudrait que les OSU soient des acteurs au





	<p>niveau des DOI. Il faudrait clarifier pour éviter que ça aille dans tous les sens.</p> <p>Maurice : Objet de ma remarque. D'où l'importance de besoin de positionnement et recommandation ODATIS/DT.</p> <p>Gilbert : Le niveau OSU est contre-productif. 2 alternatives : soit fait par les organismes (ou par institut), soit fait par les infrastructures (ex. PNDB, DT). Sinon au bout de 10 ans la moitié des DOI ne seront plus entretenus. Ce sera évoqué en réunion Data Terra</p> <p>Yvan : Du côté Terre Vivante, OSU pas aussi structuré que DT sauf Milieu marin un peu plus en avance. Et accueil plutôt agréablement que le sujet soit discuté.</p> <p>Fred : Quand on attribue un DOI, il faut s'engager sur un temps très long. On s'engage contractuellement à maintenir des LP qui seront lues dans très longtemps (ex d'article encore cités 100ans après).</p> <p>Gilbert : ça implique de pas trop distribuer les choses. SEANOE, c'est environ 1 à 2 équivalamment temps-plein. Si on multiplie ce temps-là par tous les OSU, c'est autant de temps en moins pour le cœur de métier (essayer que toute les data soit bien décrite, mise à disposition). Il faut factoriser au mieux.</p>
<p>Maurice : est-ce que on peut modifier le lien vers une landing page? exemple un DOI ->http:X ... si X disparaît est ce qu'on peut faire pointer le DOI vers un autre hébergeur http://Y ? (</p>	<p>Oui, la landing page du DOI peut être mise à jour</p>
<p>Maurice: est-ce qu'on peut mettre un DOI sur des données "brutes" (non traitées) ... puis un autre DOI sur le "même" dataset mais sur des données traitées et publiées?</p> <p>exemple EMSO : données de capteur brutes qui remontent du fond... puis les "mêmes" données filtrées traitées, etc.</p>	<p>Fred : Oui, contractuellement et techniquement tout est possible. Ex EMSO, a fait le choix d'attribuer des DOI pour données brutes, ce qui nous laisse plusieurs années pour les traiter par la suite. Et quand on a le temps, on va retraiter les données et les convertir dans format pivot (ex netcdf). Amender un DOI de</p>



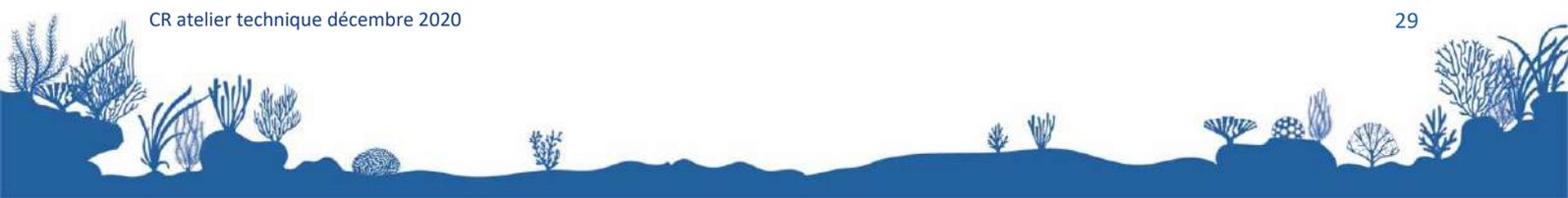


	<p>données brutes avec un DOI de data “retraitées”.</p> <p>Laurent : Les CDS peuvent prendre l’initiative de prendre des DOI pour des datasets qu’ils stockent ce qui nous amène à parfois avoir un même jeu de data à différents endroits mais avec différents DOI.</p> <p>Gilbert: Problématique pour l'utilisateur (peut porter à un confusion). Ne me choque pas qu’il y ait un DOI par niveau de traitement. C’est normal, car chaque traitement ajoute un caractère subjectif. Une des recommandations les DOI qui référencent des données dérivées, devraient référencer les DOI des data dont ils sont dérivés. Garder un peu l’héritage, l’historique du dataset (source d’origine, ...). Par contre avoir un même dataset quasi identique référencé dans différents endroits, ce n’est pas si top que ça pour les utilisateurs. Mais ça arrivera, mais l’idée est d’éviter cela au travers d’ODATIS/DT (éviter une telle redondance).</p>
<p>Maurice : ce n’est pas ce que font EMSO Açores ou ils mettent un DOI chaque année</p> <p>oui Thierry on en avait parlé , mais ça me semblait beaucoup par rapport à ce qu' a dit Fred dans son exposé, ou certains choisissent de DOIser un seul jeu de donnée qui évolue (mais ça reste un choix j’en conviens)</p>	<p>Thierry : Pour EMSO-Açores, les DOI sont gérés par déploiements (qui durent environ un an)</p>
<p>Dimitry: est-ce que les informations que l’on fournit sur SEANOE pour obtenir un DOI sont suffisantes pour fournir des métadonnées “riches” dans le sens FAIR?</p>	<p>Fred : en terme de respect des préconisations FAIR, pour les métadonnées, SEANOE s’en sort plutôt bien. Quelques points d’améliorations ont été repérés et sont en cours de réflexion (ex : ajout de métadonnées sous forme de liste de références, ex : liste de paramètres SeaDataNet).</p>
<p>Pascal: est-il légal de modifier une fiche d’ORCID d’autrui à partir des infos d’un annuaire interne?</p>	<p>Thierry : Il n'y a pas de modification de fiche "d'autrui" dans l'annuaire Ifremer. C'est la personne qui demande explicitement à l'annuaire de tenir compte de son ORCID</p>





	<p>Fred : c'est le scientifique qui donne l'autorisation à un service (ex : DataCite, Archimer) de venir enrichir sa page ORCID. Le scientifique peut révoquer cette autorisation à tout moment. Cette autorisation qui doit être donné sur le site ORCID, est conservé sur la forme d'une clé dans chaque système.</p>
<p>Pascal: Est-ce que l'IFREMER propose un service d'attribution d'IGSN? Est-ce prévu au niveau de TERRA DATA?</p>	<p>Thierry: oui, les échantillons géologiques des campagnes océanographiques françaises ont un IGSN géré par Ifremer-Sismer.</p>
<p>Pascal: ORCID pas adapté aux ingénieurs et techniciens. Quelle est la meilleure approche pour identifier les personnes en dehors des chercheurs?</p>	<p>Joël: Je ne vois pas pourquoi ORCID n'est pas adapté aux ingénieurs? Peux-tu préciser pourquoi stp?</p> <p>Pascal: on attache surtout des publications dans ORCID. Les ingénieurs ont très peu de publications. On trouve référence à leur travail sous forme de remerciements dans les publications.</p> <p>Thierry: les jeux de données publiés avec DOI ont intérêt à lister précisément leurs auteurs et contributeurs (scientifique, ingénieur, technicien) avec ORCID. ORCID n'est pas limité aux publications classiques. A ma connaissance, il n'y a pas de problème ORCID pour ingénieurs et techniciens.</p> <p>Fred : Le "C" de ORCID c'est pour Contributor. ORCID n'est pas réservé aux scientifiques, il est dédié à l'ensemble des contributeurs de la recherche : Scientifiques, Ingénieurs, techniciens (c'est une réponse de ORCID à la même question posée lors d'une présentation)</p>
<p>Clémence : Quelle est la différence entre PID et URI? Les PID sont-ils des sous-groupes d'URI?</p>	<p>Fred : "Un URI peut être de type « locator » ou « name » ou les deux. Ex :</p> <p>http://www.wikipedia.org/ urn:isbn:0-395-36341-1"</p> <p>Source: https://fr.wikipedia.org/wiki/Uniform_Resource_Identifier</p>





	<p>Il n'y a donc pas cette notion de "redirection" comme dans les PID. C'est ce mécanisme de redirection qui rend les PID pérennes car il est possible de mettre à jour l'URL de la redirection auprès de l'organisme qui gère le PID.</p>
<p>Pascal: Est-ce qu'un DOI peut être mis sur des conditions expérimentales plutôt que sur les données?</p>	<p>Thierry: c'est faisable, c'est un DOI sur des métadonnées.</p> <p>Pascal: Les métadonnées peuvent-elles exister sans les données?</p> <p>Thierry: un DOI sur un échantillon (géologique, biologique) est un exemple de DOI sans données (d'où vient-il, où est-il, ...).</p> <p>Chaque campagne océanographique française a un DOI, c'est aussi un exemple de DOI sans données</p> <p>Pascal: Pour les échantillons (carottes), nous utilisons des IGSN. Quelle différence entre IGSN et DOI?</p> <p>Thierry: c'est un choix pertinent, voir la présentation de Fred "Comment choisir son PID"</p>

