



DATA
TERRA



ODATIS

ODATIS

Certification des entrepôts,
FAIRisation des données et
Science ouverte

Assemblée Générale ODATIS

Erwann Quimbert & l'équipe du bureau exécutif ODATIS



ODATIS coordinateur de projet

- ANR Flash COPILOtE (2020-2022) : **CertificatiOn PôLe OcEan**
 - Un projet ANR vers la certification et la FAIRisation des Centres de Données et de Services du pôle Océan ODATIS
- ANR FNSO SO'ODATIS – projet non retenu 2023

Objectifs du projet ANR Flash COPiLOtE (2020-2022)

- COPiLOtE : **CertificatiOn PôLe OcEan**
- Un projet ANR vers la certification et la FAIRisation des Centres de Données et de Services du pôle Océan ODATIS
- Accompagner les Centre de données et de services (CDS) ODATIS à déposer un dossier de certification auprès de **Core Trust Seal (CTS)**
- Réaliser une évaluation du caractère **FAIR** des données gérées par les CDS ODATIS

Copilote – certification *Core Trust Seal*

- CTS est un organisme de certification des entrepôts de données, né de la fusion de 2 systèmes de certifications (DSA & WDS), sous l'égide de la RDA



<https://www.coretrustseal.org/>

Critères pour être certifié



L'entrepôt de données est évalué selon **16 critères** différents organisé en 3 thèmes principaux:

- Organisation de l'infrastructure
- Gestion des données numériques (données et métadonnées)
- Technologie et sécurité de l'infrastructure



CoreTrustSeal Trustworthy Digital Repositories
Requirements 2023-2025
Extended Guidance
V01.00

| | | | |
|----|---|-----|--|
| R1 | Mission/champ d'application | R9 | Procédures de stockage documentées |
| R2 | Licences | R10 | Plan de préservation |
| R3 | Continuité d'accès | R11 | Qualité de la données |
| R4 | Confidentialité/Ethique | R12 | Workflow de l'ingestion à la dissémination |
| R5 | Infrastructure organisationnelle | R13 | Découverte des données et identification |
| R6 | Avis d'experts | R14 | Réutilisation des données |
| R7 | Intégrité et authenticité des données | R15 | Infrastructure technique |
| R8 | Pertinence et compréhensibilité des données | R16 | Sécurité |

Intérêt de la certification



- Gage de qualité et de fiabilité
- Confiance pour les producteurs et utilisateurs de données
- Partage de bonnes pratiques
- Application des principes
- Gains en performance : interopérabilité
- Découverte des données et citation améliorées
- Pratiques et processus améliorés

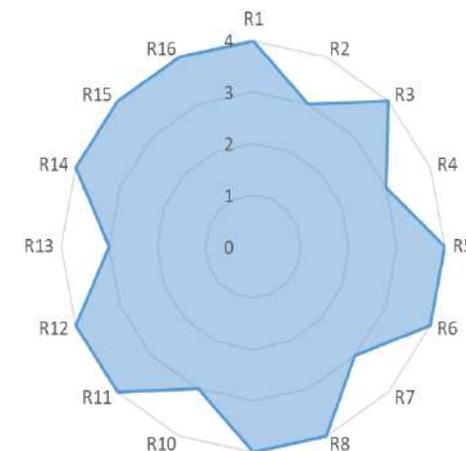
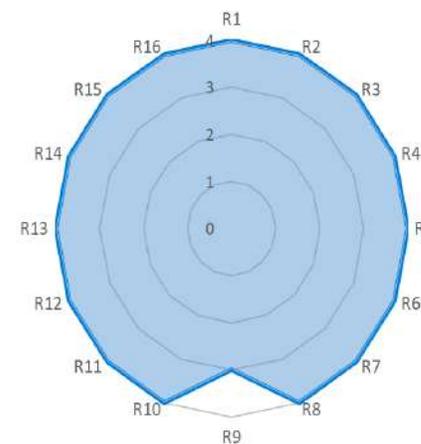
Méthode appliquée dans le cadre de COPiLOtE

- CDS Odatis
 - **formation RDA** sur la certification
 - **une lecture guidée** des critères CTS (cadre ODATIS)
- Accompagnement par le projet COPiLOtE pour la préparation du dossier CoreTrustSeal
 - Réunions de suivi régulières pour préparer le document (réponse critère par critère)
 - Auto-évaluation de tous les critères à la fin
 - Auto-évaluation croisée pour certains CDS

Auto évaluation des CDS



- Chaque CDS s'auto-évalue sur les 16 critères CTS et se donne **une note de 0 à 4**
 - 0 : non applicable
 - 1 : Non envisagé
 - 2 : En cours d'examen
 - 3 : En cours de mis en œuvre
 - 4 : Entièrement mis en œuvre
- Pour pouvoir prétendre à la certification il faut une note de **minimum 3** à chacun des 16 critères



COPiLOtE – Résultats certification

- 4 CDS ont déposé leur dossier de certification en octobre 2022



SOMLIT (données d'océanographie côtière) et **KIDA** (astrophysique)
 Retour revue avec commentaires



CNRS UPMC
Station Biologique
Roscoff
1872

SNO PHYTOBS (phytoplancton) et **SNO BENTHOBS** (macrofaune)
 Retour revue avec commentaires



300 ans d'hydrographie

1 filière : REFMAR (l'observation in-situ du niveau de la mer)
 Retour revue avec commentaires



1 filière AVISO+ (altimétrie de l'océan)
 Retour revue avec commentaires

- 3 CDS ont fait leur demande de renouvellement:



Coriolis, SISMER & Cersat Certification renouvelée – nov. 2023

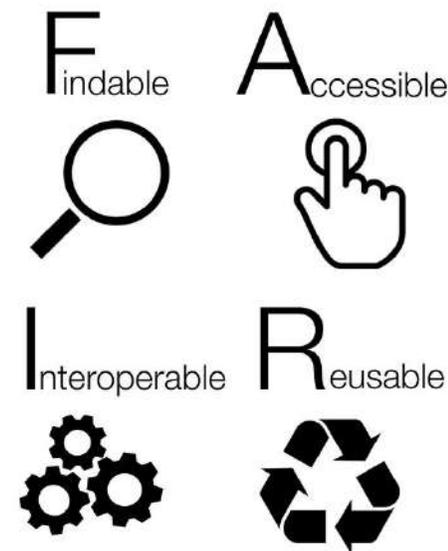
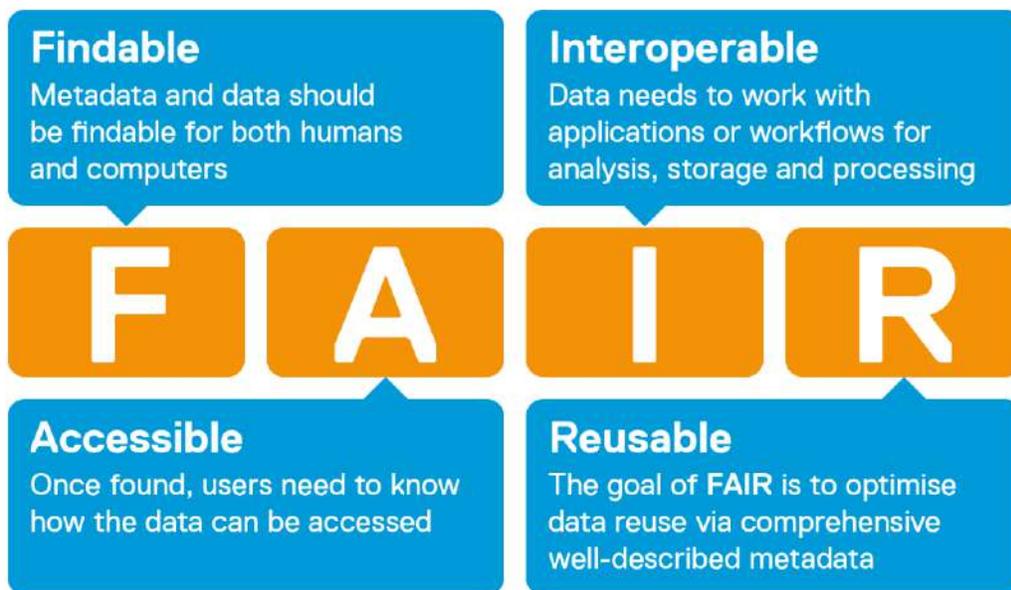
Bilan de ce travail de certification

- Les quatre CDS ont constaté son apport bénéfique principalement
 - Clarification écrite des flux de données
 - Ajout d'information
 - Ajout de documents annexes utiles
 - Prise en compte de l'ensemble des processus
 - Clarification des procédures internes concernant les filières évaluées
 - Mise en place de nouvelle procédure de réplication par exemple
 - Implication et collaboration de nombreux interlocuteurs par CDS
- Difficultés rencontrées
 - Définition du périmètre de la certification, définition de la granularité des informations à décrire
 - Un peu chronophage

COPiLOtE – Auto-évaluation FAIR des CDS ODATIS

- Les principes FAIR : un ensemble de recommandations pour gérer les données de la recherche et les rendre

Facile à trouver, **A**ccessible, **I**nteropérable et **R**éutilisable



COPiLOtE : Outil utilisé RDA - FAIR Data Maturity Model



- RDA - FAIR Data Maturity Model WG
 - RDA FAIR Data Maturity Model Specification and Guidelines Recommendation : <https://doi.org/10.15497/rda00050>
- Modèle d'auto-évaluation générique pour mesurer le niveau de maturité d'un jeu de donnée
- Approche proposée par la RDA/FDMM la plus adaptée au contexte ODATIS
- Permet :
 - D'informer et sensibiliser aux pratiques FAIR
 - D'adapter et nuancer l'évaluation en fonction des spécificités des données
 - De prendre en compte toutes les pratiques (différence avec outil centré sur l'interopérabilité)

Indicateurs du FAIR DATA MATURITY MODEL (FDMM)



- 41 indicateurs au total

- **F** : 7 indicateurs
- **A** : 12 indicateurs
- **I** : 12 indicateurs
- **R** : 10 indicateurs

Les indicateurs sont classifiés comme:

- **Essentiel**: de la plus haute importance
- **Important**: Accroît substantiellement le caractère FAIR
- **Utile**: bien à avoir, mais pas indispensable

| Priority | Principe | | | | Grand Total |
|--------------------|----------|------------|---------------|-----------|-------------|
| | Findable | Accessible | Interoperable | Reusable | |
| Essential | 7 | 8 | 0 | 5 | 20 |
| Important | 0 | 3 | 7 | 4 | 14 |
| Useful | 0 | 1 | 5 | 1 | 7 |
| Grand Total | 7 | 12 | 12 | 10 | 41 |

2 méthodes d'évaluation du FDMM



- Mesurer la réussite ou l'échec: Déterminer si une ressource évaluée répond aux exigences d'un indicateur exprimé sur une échelle binaire de réussite ou d'échec.
- Mesurer la maturité : Fournir une mesure dans laquelle une ressource évaluée répond aux exigences d'un indicateur exprimé suivant l'échelle
 - 0 = non applicable,
 - 1 = non envisagé,
 - 2 = en cours d'examen,
 - 3 = en cours de mise en œuvre,
 - 4 = entièrement mis en œuvre.



Même échelle de mesure
que la certification CTS

COPiLOtE : Auto-évaluation FAIR des CDS ODATIS

- Questionnaire d'auto-évaluation avec chaque filière – Réunions pour répondre au questionnaires (2 réunions par filière)
 - **CDS-IS-SBR** : pour PHYTOBS et BENTHOBS,
 - **CDS-IS-CORIOLIS** : pour ARGO, DBCP, OceanSITES-PIRATA, GOSUD-Ferry Box, Gliders
 - **CDS-IS-IMEV** : pour Lefe-Cyber
 - **CDS-IS-OASU** : pour SOMLIT et KIDA
 - **CDS-IS-SISMER** : pour les filières SISMER : Données géographiques (Sextant), Données des campagnes : de physique chimie et Géosciences, Catalogue des campagnes, Quadrige/SURVAL (monitoring côtier), Echantillons biologiques et géologiques
 - **CDS-IS-Shom** : pour RONIM/REFMAR
 - **CDS-SAT-CERSAT** : pour l'ensemble des produits + CATDS
 - **CDS-SAT-AVISO** : pour AVISO+

Quimbert Erwann, Fichaut Michèle, Maudire Gilbert (2022). **Guide principes FAIR**. Principes FAIR dans le contexte du pôle ODATIS.

<https://doi.org/10.13155/8710>

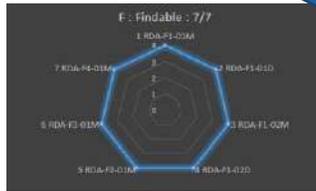
Auto-évaluations FAIR – PRI

DOI non disponible pour le moment,
pas de moissonnage possible
Réflexion en cours pour une reprise du
catalogue sur les outils du pôle Odatis

En cours de labélisation

DOI ont été pré-générés mais ne
pointent pas vers des LP pour le
moment, prévu avec les crédits de fin
d'année

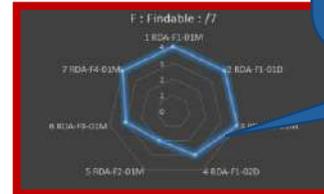
PHYTOBS



BENTHOBS



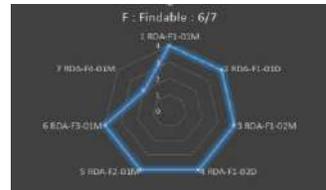
REFMAR



CERSAT



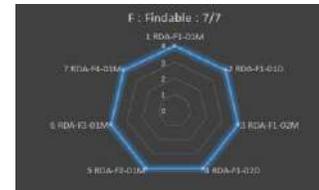
CAMPAGNES



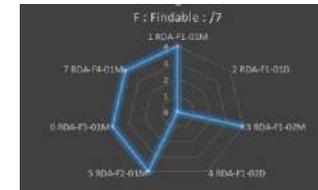
ECHANTILLONS



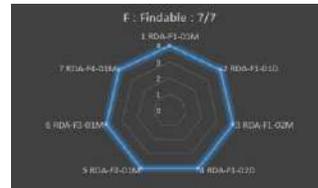
GEOSCIENCES



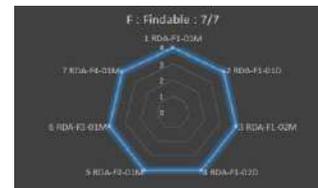
DONNEES GEO



SURVAL



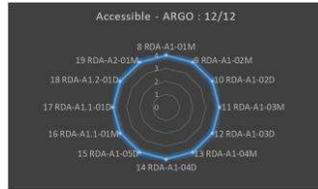
PHYSIQUE-CHEMIE



13/17

Auto-évaluations FAIR – PRINCIPE A

ARGO



DBCP



LEFE CYBER



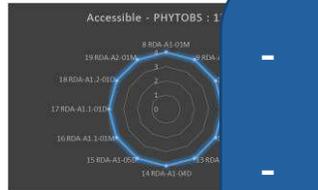
SOMLIT



KIDA



PHYTOBS



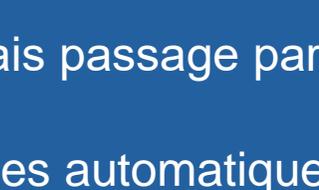
BENTHOS



BONUM



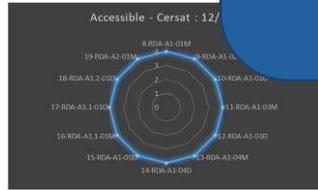
AVISO



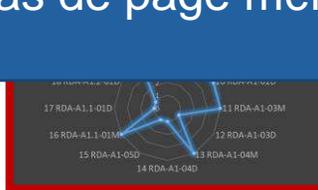
CATDS



CERSAT



ODATIS



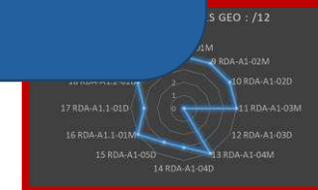
ODATIS



ODATIS

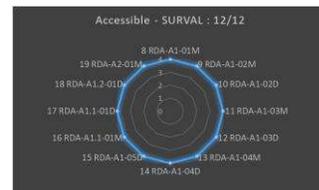


ODATIS GEO

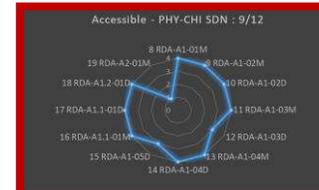


- Pas d'accès direct aux données mais passage par un panier d'extraction
- Les données ne sont pas accessibles automatiquement (c'est-à-dire par un programme informatique)
- Pas de page mémorial sur les métadonnées

SURVAL



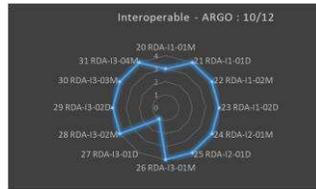
PHYSIQUE-CHEMIE



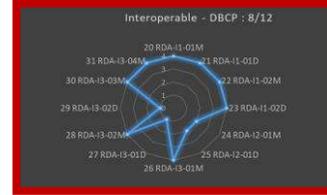
9/17

Auto-évaluations FAIR – PRINCIPE I

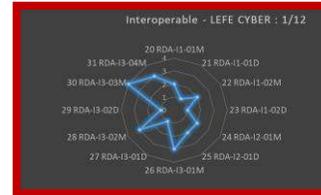
ARGO



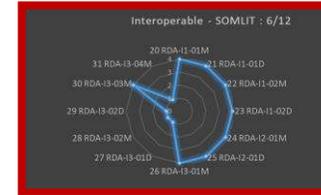
DBC



LEFE CYBER



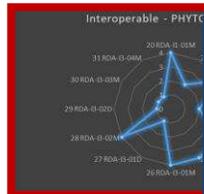
SOMLIT



KIDA



PHYTOBIO

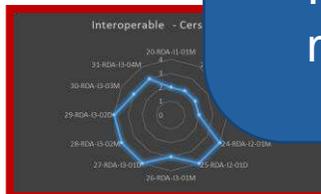


- Pas de “représentation de connaissances exprimée dans un format standardisé compréhensible par l’homme et la machine” (RDF/SPARQL/Websémantique)
- Utilisation de vocabulaire mais pas forcément FAIR
- Références croisées implémentées dans les métadonnées mais très peu dans les données

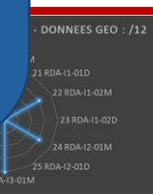
CATDS



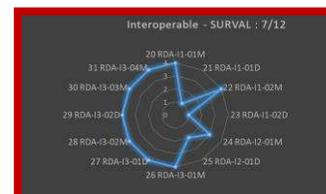
CERSA



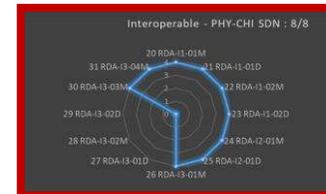
DONNEES GEO



SURVAL



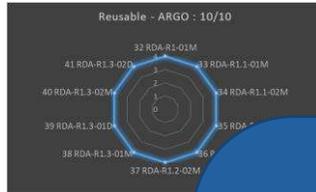
PHYSIQUE-CHEMIE



1/17

Auto-évaluations FAIR – PRINCIPE R

ARGO



DBCP



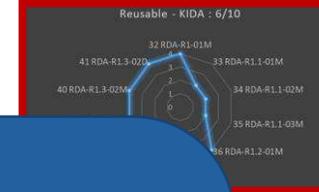
LEFE CYBER



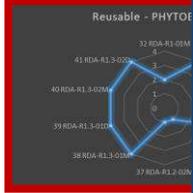
SOMLIT



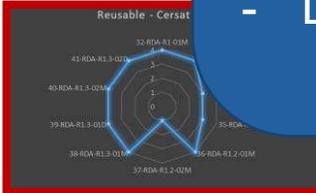
KIDA



PHYTO

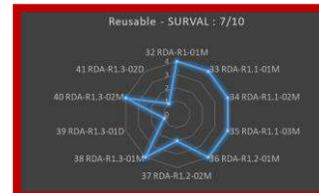


CERSA

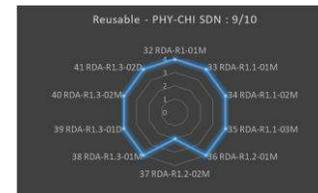


- Pas de licence de spécifiée seulement les conditions d'accès
- Dans certains cas, licence à standardiser
- Pas d'information de provenance selon un langage intercommunautaire (PROV-O)
- Pas de norme utilisée dans certains cas pour les métadonnées (page HTML, développement 'maison'...)
- Données fournis au format excel

SURVAL



PHYSIQUE-CHIMIE



7/17

| 19

Intérêt de cette auto-Evaluation

- Pour les CDS

- Exercice très intéressant : recul sur ses pratiques
- Permet d'identifier les améliorations à apporter aux ensembles de données, en particulier concernant les critères essentiels

- Pour le pôle ODATIS

- Tester si une ressource a rempli avec succès les exigences de FAIRitude établie par la communauté ODATIS
- Avoir une vision globale & partagée sur le caractère FAIR des données de l'ensemble des CDS
- Identifier des outils à mettre en œuvre pour améliorer le caractère FAIR des données des CDS (ex: SPARQL endpoint)

Projet SO'ODATIS



| | |
|--|----|
| WP1 – Gestion et coordination du projet SO'Odatis | 2 |
| Tâche 1.1 – Gestion administrative & financière | 2 |
| Tâche 1.2 - Coordination opérationnelle et technique | 2 |
| Tâche 1.3 – Communication via site web et médias sociaux | 2 |
| WP2 – Offre de services sur les publications | 3 |
| Tâche 2.1 : Publication scientifique | 3 |
| Tâche 2.2 : Connexion de l'archive institutionnelle Archiver avec HAL | 5 |
| Tâche 2.3 : POC archive documentaire pour la littérature grise | 5 |
| WP3 – Données, codes & publications pour Odatis | 6 |
| Tâche 3.1 - Mise à niveau des outils du pôle Odatis : SEANOE & Sextant | 6 |
| Tâche 3.2 - Outils de monitoring & tableau de bord | 7 |
| Tâche 3.3 - Web sémantique & vocabulaires | 8 |
| Tâche 3.4 - Data paper | 8 |
| Tâche 3.5 - Interconnexion avec les initiatives nationales & internationales | 9 |
| WP4 – Bonnes pratiques et accompagnement des communautés | 9 |
| Tâche 4.1 - Formation et production de guides sur les bonnes pratiques | 10 |
| Tâche 4.2 - Accompagner les porteurs de projets de recherche sur les activités de gestion de données | 11 |
| Tâche 4.3 - Actions d'information sur les outils proposés | 11 |

- Les objectifs poursuivis par l'appel à projets sont les suivants :
 - Favoriser le continuum entre publications, données et codes
 - Favoriser la reproductibilité ou la traçabilité de la démarche scientifique

<https://www.ouvrirlascience.fr/appel-a-projets-publication-et-edition-scientifiques-ouvertes-3eme-edition/#calendrier>

A venir - Projet SO'ODATIS

Offre de services sur les publications

- Proposer une épi-revue de type diamant
- Interconnexion de l'archive institutionnel Archimer avec l'archive nationale HAL
- POC archive pour la littérature grise

Données, codes, logiciels

- Mise à niveau entrepôt et catalogue du pôle
- Proposer des outils de monitoring et tableaux de bord
- Initier une démarche vers le web des données
- Booster la génération de Data Paper

Bonnes pratiques et accompagnement des communautés

- Guides de bonnes pratiques
- Accompagner les porteurs de projet sur les activités de gestion de données
- Communications sur la Science Ouverte



**DATA
TERRA**



ODATIS

Merci de votre attention



14/12/2023

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