



DATA
TERRA



ODATIS

COPiLOtE

Utilisation du FAIR Data Maturity Model pour évaluer le caractère FAIR de vos données

Réunion annuelle RDA France

Erwann Quimbert & équipe projet COPiLOtE



25/11/2022

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

Le pôle Océan ODATIS

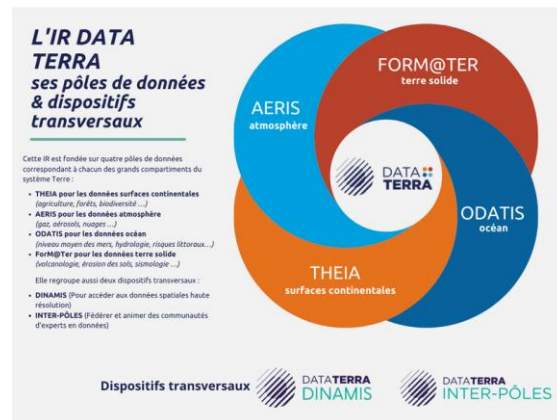
Missions

- Promouvoir et faciliter l'accès aux données d'observations réalisées dans l'océan ou à son interface avec les autres milieux, à partir de mesures in situ et de télédétection
- Fédérer au niveau national des activités de gestion de données et d'expertise scientifique en océanographie

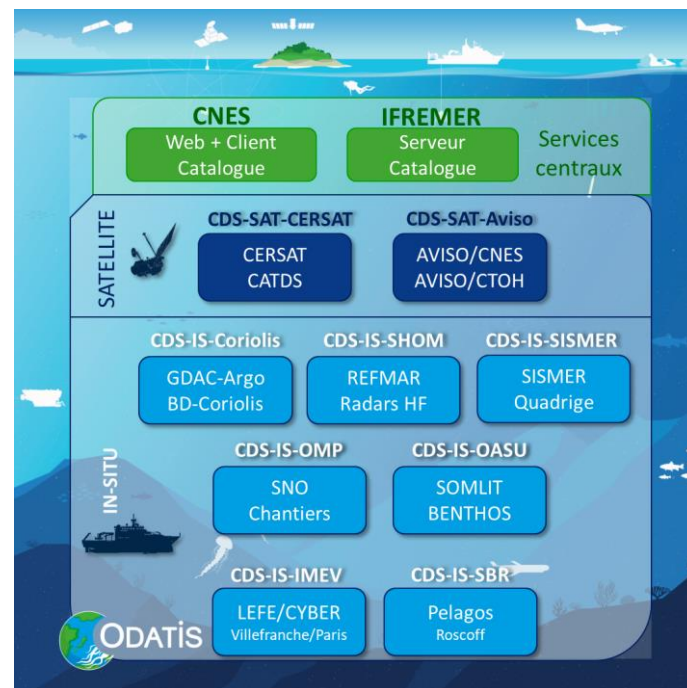
Partenaires



www.odatis-ocean.fr



Les centres de données et services d'ODATIS



Copilote - CertificatiOn PoLe OcEan



- Objectifs du projet ANR Flash COPILOtE (2020-2022)
 - Accompagner les 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
- 4 CDS ont déposé leur dossier de certification (en cours d'évaluation)

Observatoire Aquitain



des Sciences de l'Univers

CNRS UPMC

Station Biologique
Roscoff

SOMLIT (données d'océanographie côtière) et **KIDA** (astrophysique)

SNO PHYTOBS (phytoplancton) et **SNO BENTHOBS** (macrofaune)

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

1 filière AVISO+ (altimétrie de l'océan)

SH M

300 ans d'hydrographie



- 3 CDS ont fait leur demande de renouvellement : Coriolis, SISMER & Cersat



COPiLOTE – Auto-évaluation FAIR

- Un certain nombre d'initiatives différentes travaillent actuellement à la définition de cadres, de méthodes et de critères d'évaluation du caractère FAIR des données :

- FAIR Implementation Profile



- FAIRsFAIR



- RDA - FAIR Data Maturity Model WG

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

Outils et méthodes pour évaluer la FAIRness des données



- 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 nous a paru la plus adaptée au contexte Odatis : réunion f2f, sensibilisation aux principes FAIR, évaluation de la maturité des CDS en matière de données FAIR
- Utiliser ces recommandations dans le cadre de l'auto-évaluation FAIR qui est à mener dans le projet COPiLOtE

Indicateurs FDMM

• 41 indicateurs ont été définies par le groupe de travail FDMM de la RDA

F : 7 indicateurs

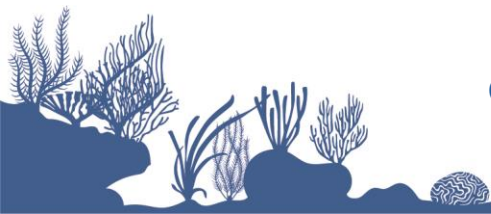
A : 12 indicateurs

I : 12 indicateurs

R : 10 indicateurs

ID	FAIR	Identifiant RDA	Priorité RDA	Indicateur
1	F1	RDA-F1-01M	Essentielle	Les métadonnées sont identifiées par un identifiant pérenne
2	F1	RDA-F1-01D	Essentielle	Les données sont identifiées par un identifiant pérenne
3	F1	RDA-F1-02M	Essentielle	Les métadonnées sont identifiées par un identifiant unique au niveau mondial
4	F1	RDA-F1-02D	Essentielle	Les données sont identifiées par un identifiant unique au niveau mondial
5	F2	RDA-F2-01M	Essentielle	Des métadonnées riches sont fournies pour permettre la découverte des données
6	F3	RDA-F3-01M	Essentielle	Les métadonnées comprennent l'identifiant des données
7	F4	RDA-F4-01M	Essentielle	Les métadonnées sont proposées de manière à pouvoir être moissonnées et indexées
8	A1	RDA-A1-01M	Essentielle	Les métadonnées contiennent des informations permettant à l'utilisateur d'accéder aux données
9	A1	RDA-A1-02M	Essentielle	Les métadonnées sont accessibles manuellement (c'est-à-dire avec une intervention humaine)
10	A1	RDA-A1-02D	Essentielle	Les données sont accessibles manuellement (c'est-à-dire avec une intervention humaine)
11	A1	RDA-A1-03M	Essentielle	L'identifiant de métadonnées renvoie à un enregistrement de métadonnées
12	A1	RDA-A1-03D	Essentielle	L'identifiant des données renvoie à un objet numérique
13	A1	RDA-A1-04M	Essentielle	Les métadonnées sont accessibles via un protocole standardisé
14	A1	RDA-A1-04D	Essentielle	Les données sont accessibles via un protocole standardisé
15	A1	RDA-A1-05D	Important	Les données sont accessibles automatiquement (c'est-à-dire par un programme informatique)
16	A1.1	RDA-A1.1-01M	Essentielle	Les métadonnées sont accessibles via un protocole d'accès libre
17	A1.1	RDA-A1.1-01D	Important	Les données sont accessibles via un protocole d'accès libre
18	A1.2	RDA-A1.2-01D	Utile	Les données sont accessibles via un protocole d'accès qui prend en charge l'authentification et l'autorisation
19	A2	RDA-A2-01M	Essentielle	Il est garanti que les métadonnées restent disponibles après que les données ne le soient plus
20	I1	RDA-I1-01M	Important	Les métadonnées utilisent une représentation des connaissances exprimée dans un format standardisé
21	I1	RDA-I1-01D	Important	Les données utilisent une représentation des connaissances exprimée dans un format standardisé
22	I1	RDA-I1-02M	Important	Les métadonnées utilisent une représentation des connaissances compréhensible par une machine
23	I1	RDA-I1-02D	Important	Les données utilisent une représentation des connaissances compréhensible par une machine
24	I2	RDA-I2-01M	Important	Les métadonnées utilisent des vocabulaires conformes aux principes FAIR
25	I2	RDA-I2-01D	Utile	Les données utilisent des vocabulaires conformes aux principes FAIR
26	I3	RDA-I3-01M	Important	Les métadonnées incluent des références à d'autres métadonnées
27	I3	RDA-I3-01D	Utile	Les données incluent des références à d'autres données
28	I3	RDA-I3-02M	Utile	Les métadonnées incluent des références à d'autres données
29	I3	RDA-I3-02D	Utile	Les données incluent des références qualifiées à d'autres données
30	I3	RDA-I3-03M	Important	Les métadonnées incluent des références qualifiées à d'autres métadonnées
31	I3	RDA-I3-04M	Utile	Les métadonnées comprennent des références qualifiées à d'autres données
32	R1	RDA-R1-01M	Essentielle	Une pluralité d'attributs précis et pertinents sont fournis pour permettre la réutilisation
33	R1.1	RDA-R1.1-01M	Essentielle	Les métadonnées comprennent des informations sur la licence sous laquelle les données peuvent être réutilisées
34	R1.1	RDA-R1.1-02M	Important	Les métadonnées font référence à une licence de réutilisation standard
35	R1.1	RDA-R1.1-03M	Important	Les métadonnées font référence à une licence de réutilisation compréhensible par une machine
36	R1.2	RDA-R1.2-01M	Important	Les métadonnées comprennent des informations sur la provenance selon des normes spécifiques à la communauté
37	R1.2	RDA-R1.2-02M	Utile	Les métadonnées incluent des informations de provenance selon un langage intercommunautaire
38	R1.3	RDA-R1.3-01M	Essentielle	Les métadonnées sont conformes à une norme communautaire
39	R1.3	RDA-R1.3-01D	Essentielle	Les données sont conformes à une norme communautaire
40	R1.3	RDA-R1.3-02M	Essentielle	Les métadonnées sont exprimées conformément à une norme communautaire compréhensible par les machines
41	R1.3	RDA-R1.3-02D	Important	Les données sont exprimées conformément à une norme communautaire compréhensible par les machines

COPiLoTe - Evaluer la FAIRness de vos données



Priorités FDMM

- Trois niveaux d'importance :

- **Essentielle**

- “un tel indicateur traite d'un aspect qui est de la plus haute importance pour atteindre l'équité dans la plupart des circonstances ou, à l'inverse, l'équité serait pratiquement impossible à atteindre si l'indicateur n'était pas satisfait.”

- **Important**

- “un tel indicateur aborde un aspect qui pourrait ne pas être de la plus haute importance dans des circonstances spécifiques, mais dont la satisfaction, si elle est possible, augmenterait substantiellement la FAIRness.”

- **Utile**

- “un tel indicateur aborde un aspect qui est agréable à avoir mais qui n'est pas nécessairement indispensable.”

Mesures FDMM

- Le groupe de travail FDMM a défini deux méthodes d'évaluation :
 - La première consiste à mesurer la conformité ou la non-conformité d'un indicateur par rapport aux exigences exprimées dans un indicateur. Une réponse binaire (réussite ou échec) est attribuée pour chacun des indicateurs.
 - La seconde consiste à attribuer un niveau de maturité à chaque indicateur de sorte que chaque mesure ait une des valeurs suivantes :
 - 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

Indicateurs FDMM – Principe F

ID	FAIR	Identifiant RDA	Priorité RDA	Indicateur
1	F1	RDA-F1-01M	Essentielle	Les métadonnées sont identifiées par un identifiant pérenne
2	F1	RDA-F1-01D	Essentielle	Les données sont identifiées par un identifiant pérenne
3	F1	RDA-F1-02M	Essentielle	Les métadonnées sont identifiées par un identifiant unique au niveau mondial
4	F1	RDA-F1-02D	Essentielle	Les données sont identifiées par un identifiant unique au niveau mondial
5	F2	RDA-F2-01M	Essentielle	Des métadonnées riches sont fournies pour permettre la découverte des données
6	F3	RDA-F3-01M	Essentielle	Les métadonnées comprennent l'identifiant des données
7	F4	RDA-F4-01M	Essentielle	Les métadonnées sont proposées de manière à pouvoir être moissonnées et indexées

Indicateurs FDMM – Principe F

D3.1 & D3.2 - Quimbert Erwann,
Fichaut Michèle, Maudire Gilbert
(2022). **Guide principes FAIR.**
Principes FAIR dans le contexte du
pôle ODATIS.
<https://doi.org/10.13155/87107>

1	RDA-F1-01M	Les métadonnées sont identifiées par un identifiant pérenne			
Principe FAIR associé	F1	Critère CoreTrustSeal	R13	Priorité RDA	Essentielle
Description	<p>RDA "Cet indicateur évalue si les métadonnées sont identifiées ou non par un identifiant permanent. Un identifiant persistant garantit que les métadonnées resteront trouvables dans le temps et réduit le risque de liens brisés."</p> <p>FsF "Nous faisons une distinction entre l'unicité et la persistance d'un identifiant. Une URL HTTP est globalement unique, mais peut ne pas être persistante car l'URL des données peut ne pas être accessible (problème de liens brisés) ou les données disponibles sous l'URL d'origine peuvent être modifiées (problème de dérive du contenu). Les identificateurs basés, par exemple, sur le Handle System, DOI, ARK sont à la fois uniques et persistants. Ils sont maintenus et régis de manière à rester stables et résolubles sur le long terme. L'identifiant persistant (PID) d'une ressource peut pointer vers une page d'atterrissage contenant des métadonnées et des informations supplémentaires sur la manière d'accéder au contenu des données. »</p>				
Evaluation					
Obligation Odatis	<ul style="list-style-type: none"> Avoir un identifiant pérenne (PID) de la métadonnée qui renvoie à une landing page La génération d'identifiants persistants pour les métadonnées du CDS peut être réalisée par le pôle Odatis 				
Méthode d'évaluation	<p>RDA "La persistance d'un identifiant est déterminée par l'engagement de l'organisation qui l'attribue et le gère. L'évaluation de cet indicateur doit donc tenir compte de la politique de persistance de cette organisation. Un tel engagement pourrait être exprimé par une université ou un institut de recherche, par une infrastructure de recherche ou par une organisation qui émet des identifiants formels, comme l'International DOI Foundation. Une façon possible d'évaluer cet indicateur est de vérifier que l'identifiant utilisé pour les métadonnées est répertorié dans un service de registre comme FAIRsharing, approuvé par la RDA."</p>				
Liens utiles					
<ul style="list-style-type: none"> Liste de schémas d'identifiants uniques au monde compilés par FAIRsharing : https://fairsharing.org/standards/?q=&selected_facets=type_exact:identifieur%20sche 					

Indicateurs FDMM – Principe A

ID	FAIR	Identifiant RDA	Priorité RDA	Indicateur
8	A1	RDA-A1-01M	Essentielle	Les métadonnées contiennent des informations permettant à l'utilisateur d'accéder aux données
9	A1	RDA-A1-02M	Essentielle	Les métadonnées sont accessibles manuellement (c'est-à-dire avec une intervention humaine)
10	A1	RDA-A1-02D	Essentielle	Les données sont accessibles manuellement (c'est-à-dire avec une intervention humaine).
11	A1	RDA-A1-03M	Essentielle	L'identifiant de métadonnées renvoie à un enregistrement de métadonnées
12	A1	RDA-A1-03D	Essentielle	L'identifiant des données renvoie à un objet numérique
13	A1	RDA-A1-04M	Essentielle	Les métadonnées sont accessibles via un protocole standardisé
14	A1	RDA-A1-04D	Essentielle	Les données sont accessibles via un protocole standardisé
15	A1	RDA-A1-05D	Important	Les données sont accessibles automatiquement (c'est-à-dire par un programme informatique)
16	A1.1	RDA-A1.1-01M	Essentielle	Les métadonnées sont accessibles via un protocole d'accès libre
17	A1.1	RDA-A1.1-01D	Important	Les données sont accessibles via un protocole d'accès libre
18	A1.2	RDA-A1.2-01D	Utile	Les données sont accessibles via un protocole d'accès qui prend en charge l'authentification et l'autorisation
19	A2	RDA-A2-01M	Essentielle	Il est garanti que les métadonnées restent disponibles après que les données ne le soient plus

Indicateurs FDMM – Principe I

ID	FAIR	Identifiant RDA	Priorité RDA	Indicateur
20	I1	RDA-I1-01M	Important	Les métadonnées utilisent une représentation des connaissances exprimée dans un format standardisé
21	I1	RDA-I1-01D	Important	Les données utilisent une représentation des connaissances exprimée dans un format standardisé
22	I1	RDA-I1-02M	Important	Les métadonnées utilisent une représentation des connaissances compréhensible par une machine
23	I1	RDA-I1-02D	Important	Les données utilisent une représentation des connaissances compréhensible par une machine
24	I2	RDA-I2-01M	Important	Les métadonnées utilisent des vocabulaires conformes aux principes FAIR
25	I2	RDA-I2-01D	Utile	Les données utilisent des vocabulaires conformes aux principes FAIR
26	I3	RDA-I3-01M	Important	Les métadonnées incluent des références à d'autres métadonnées
27	I3	RDA-I3-01D	Utile	Les données incluent des références à d'autres données
28	I3	RDA-I3-02M	Utile	Les métadonnées incluent des références à d'autres données
29	I3	RDA-I3-02D	Utile	Les données incluent des références qualifiées à d'autres données
30	I3	RDA-I3-03M	Important	Les métadonnées incluent des références qualifiées à d'autres métadonnées
31	I3	RDA-I3-04M	Utile	Les métadonnées comprennent des références qualifiées à d'autres données

Indicateurs FDMM – Principe R

ID	FAIR	Identifiant RDA	Priorité RDA	Indicateur
32	R1	RDA-R1-01M	Essentielle	Une pluralité d'attributs précis et pertinents sont fournis pour permettre la réutilisation
33	R1.1	RDA-R1.1-01M	Essentielle	Les métadonnées comprennent des informations sur la licence sous laquelle les données peuvent être réutilisées
34	R1.1	RDA-R1.1-02M	Important	Les métadonnées font référence à une licence de réutilisation standard
35	R1.1	RDA-R1.1-03M	Important	Les métadonnées font référence à une licence de réutilisation compréhensible par une machine
36	R1.2	RDA-R1.2-01M	Important	Les métadonnées comprennent des informations sur la provenance selon des normes spécifiques à la communauté
37	R1.2	RDA-R1.2-02M	Utile	Les métadonnées incluent des informations de provenance selon un langage intercommunautaire
38	R1.3	RDA-R1.3-01M	Essentielle	Les métadonnées sont conformes à une norme communautaire
39	R1.3	RDA-R1.3-01D	Essentielle	Les données sont conformes à une norme communautaire
40	R1.3	RDA-R1.3-02M	Essentielle	Les métadonnées sont exprimées conformément à une norme communautaire compréhensible par les machines
41	R1.3	RDA-R1.3-02D	Important	Les données sont exprimées conformément à une norme communautaire compréhensible par les machines

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, Quadrigé/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+

Auto-évaluations – PR

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

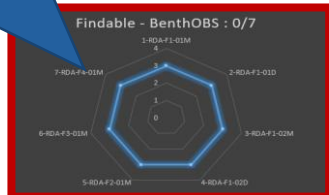
En cours de labélisation

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



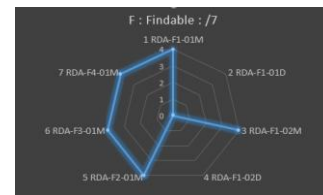
ECHANTILLONS



GEOSCIENCES



DONNEES GEO



SURVAL



PHYSIQUE-CHIMIE

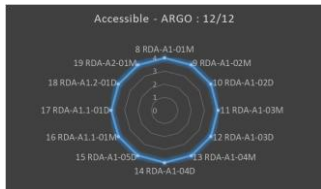


13/17

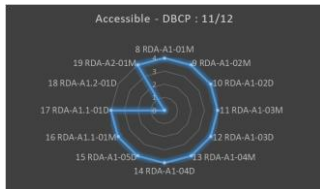
Evaluer la FAIRness de vos données

Auto-évaluations – PRINCIPE A

ARGO



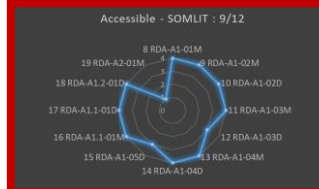
DBCP



LEFE CYBER



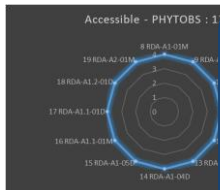
SOMLIT



KIDA



PHYTOBS



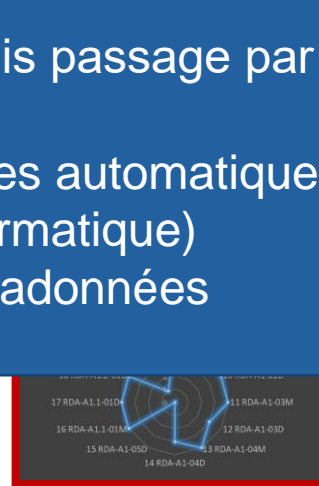
BENTHOS

- 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 mémoriale sur les métadonnées

BONUM



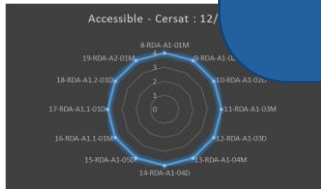
AVISO



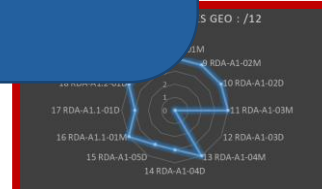
CATDS



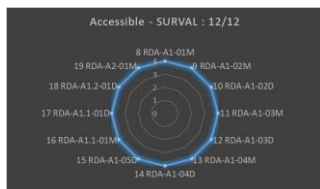
CERSAT



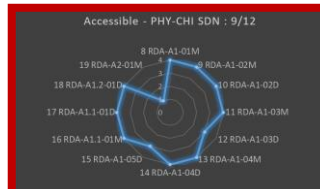
GEO



SURVAL



PHYSIQUE-CHIMIE

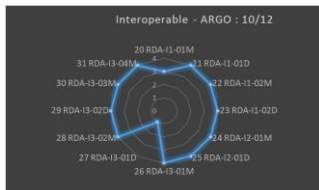


9/17

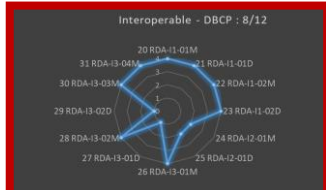
Evaluer la FAIRness de vos données

Auto-évaluations – PRINCIPE I

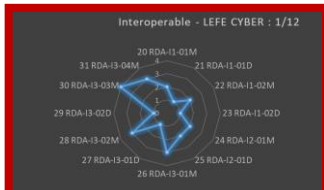
ARGO



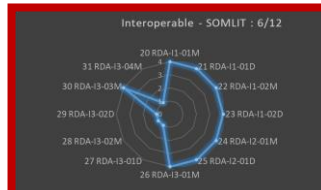
DBCP



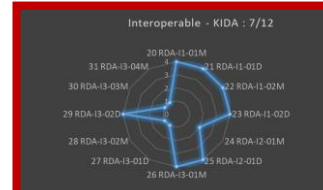
LEFE CYBER



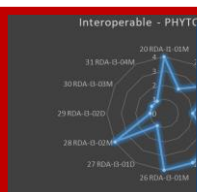
SOMLIT



KIDA

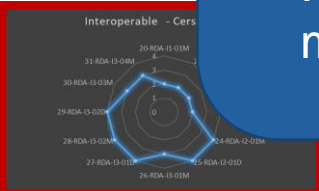


PHYTOBIO

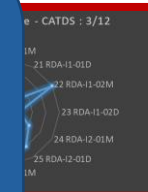


- Pas de “représentation de connaissances exprimée dans un format standardisé compréhensible par l’homme et la machine” (RDF/SPARQL/Websémantique)
- Utilisation de vocabulaire mais pas forcément FAIR
- Références croisées implémentées dans les métadonnées mais très peu dans les données

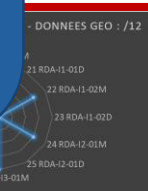
CERSA



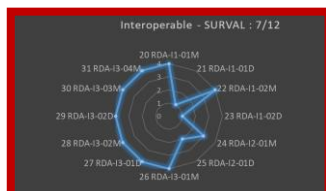
CATDS



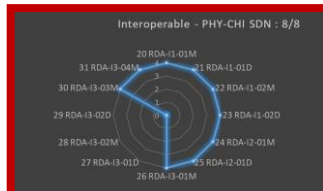
GEO



SURVAL



PHYSIQUE-CHEMIE

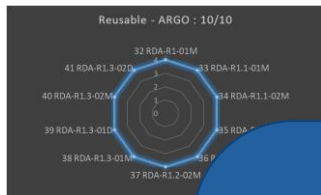


1/17

Evaluer la FAIRness de vos données

Auto-évaluations – PRINCIPE R

ARGO



DBCP



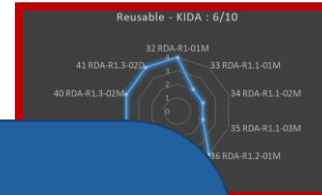
LEFE CYBER



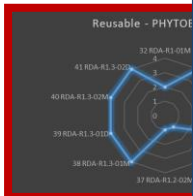
SOMLIT



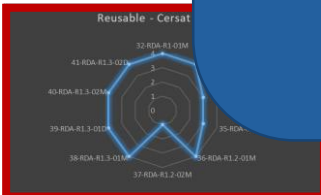
KIDA



PHYTO

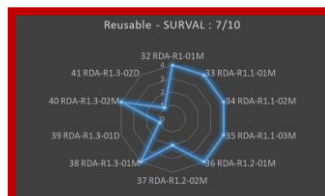


CERSA



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

SURVAL



PHYSIQUE-CHEMIE



7/17

Evaluer la FAIRness de vos données

Intérêt de ce travail d'auto-évaluation des données des CDS

- Pour les CDS

- Exercice jugé très intéressant par les CDS, permet de prendre du recul sur ses pratiques
- Cette évaluation FAIR est un moyen de tester objectivement si une ressource a rempli avec succès les exigences de FAIRitude établie par la communauté ODATIS. L'idée est aussi d'identifier les points souhaitables pour des améliorations et les prioriser.
- Les CDS utilisent cette évaluation pour déterminer où leurs pratiques peuvent être améliorées pour atteindre un niveau plus élevé de FAIRitude.

- Pour le pôle Odatis

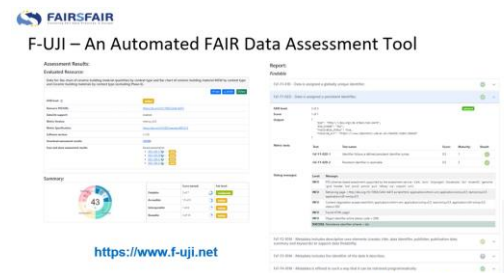
- Permet d'avoir une vision globale et partagée du degré de FAIRitude de chaque CDS
- Permet d'identifier des outils que le pôle Odatis pourrait mettre en œuvre pour améliorer la FAIRitude des données des CDS

Conclusion

- FAIRitude des données dépend beaucoup de l'outil de diffusion
- Participation à des projets européens (ENVRI FAIR, SeaDataNet) améliore souvent les pratiques (eg. Argo, banque de physique Chimie...)
- Pistes d'amélioration:
 - Amélioration de l'utilisation de vocabulaire FAIR (ajout URI, ou utilisation de vocabulaire FAIR existant: NERC-BODC, GCMD etc.)
 - Données liées, websémantique, SPARQL endpoint, RDF
 - Etude de PROV-O (ENVRI, Sextant)
 - Enrichir les métadonnées avec liens vers d'autres outils de diffusion (SDN, EMODnet, Sextant, etc.)
 - Enrichir les métadonnées avec références qualifiées à d'autres métadonnées et données quand c'est possible (ORCID, Archimer ...)

Travaux en cours (d'ici fin déc.)

- Analyse comparative avec l'outil F-UJI



- Atelier FIP pour la communauté ODATIS
- Document de synthèse à rédiger pour la fin du projet (Décembre 2022) - Piste d'amélioration pour chaque CDS à inclure dans ce document



**DATA
TERRA**



ODATIS

Merci de votre attention



25/11/2022

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