



DATA
TERRA



ODATIS

FAIR Data Maturity Model

Retour d'expérience du projet COPiLOtE

pôle Océan ODATIS

Erwann Quimbert et l'équipe projet COPiLOtE



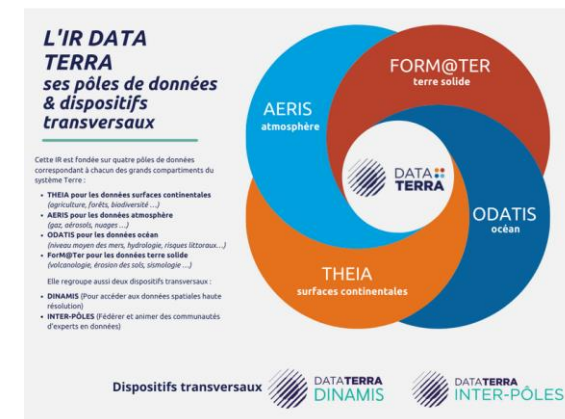
Le pôle Océan ODATIS

Missions

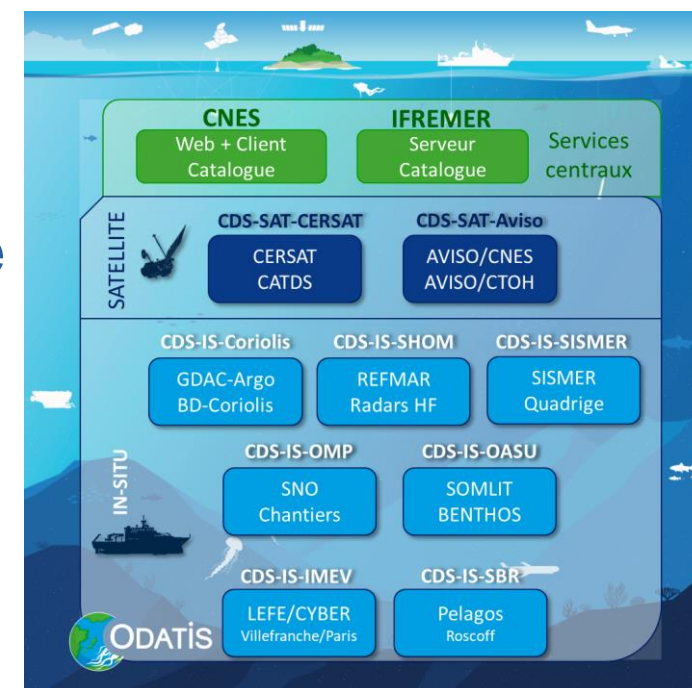
- Promouvoir et faciliter l'accès aux données d'observations réalisées dans l'océan ou à son interface avec les autres milieux, à partir de mesures in situ et de télédétection
- Fédérer au niveau national des activités de gestion de données et d'expertise scientifique en océanographie
- **Partenaires**



www.odatis-ocean.fr



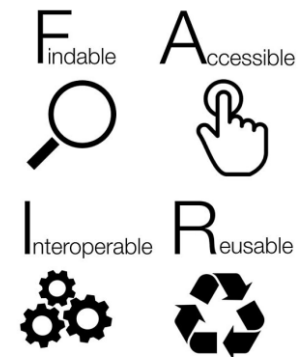
Les centres de données et services (CDS) d'ODATIS



Objectifs du projet ANR Flash COPiLOtE (2020-2022)

COPiLOtE : CertificatiOn PôLe OcEan

- Accompagner les Centres 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 : 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
- 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

	<i>Principe</i>				
<i>Priority</i>	Findable	Accessible	Interoperable	Reusable	Grand Total
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.

COPiLOtE – Auto-évaluation FAIR des CDS ODATIS

Méthodologie :

- 1) Choix de la méthode d'évaluation du caractère FAIR des données : FDMM
 - 2) Rédaction d'un guide :
 - 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>
 - Travail d'interprétation des critères/métriques FDMM effectué par ODATIS pour orienter l'évaluation de chaque critère.
 - 3) 2*2h de réunion par filière de données de l'équipe COPiLOtE avec chaque CDS pour :
 - Définir le périmètre des données à évaluer
 - Présenter la méthode & le guide et conduire l'évaluation
- permet d'établir un cadre commun pour homogénéiser les évaluations

COPiLOtE – Auto-évaluation FAIR des CDS ODATIS

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 pageLa 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:identifier%20sche					

COPiLOtE : Auto-évaluation FAIR des CDS ODATIS

- Questionnaire d'auto-évaluation avec chaque filière – Réunions pour répondre au questionnaire (2*2h de réunions par filière de données)
 - **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 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



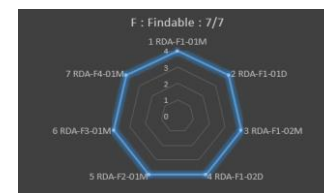
CAMPAGNES



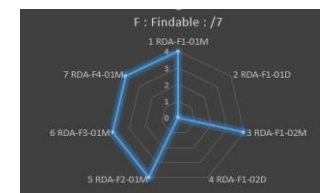
ECHANTILLONS



GEOSCIENCES



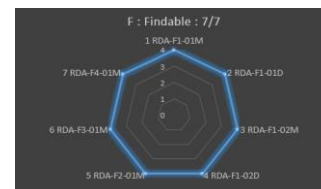
DONNEES GEO



SURVAL



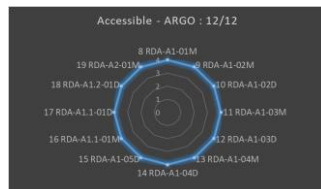
PHYSIQUE-CHIMIE



13/17

Auto-évaluations FAIR – PRINCIPE A

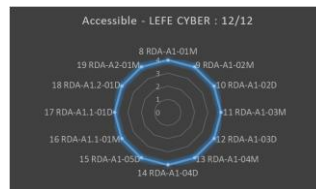
ARGO



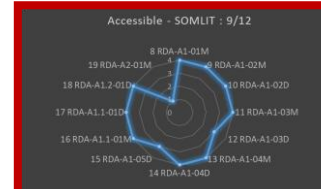
DBCP



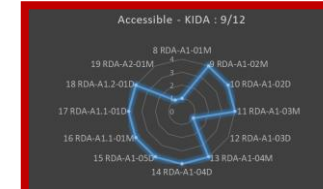
LEFE CYBER



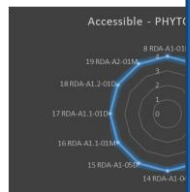
SOMLIT



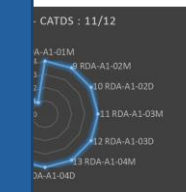
KIDA



PHYTO



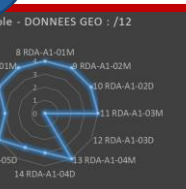
OS



CERSA

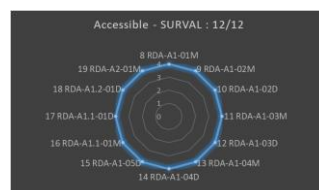


ES GEO

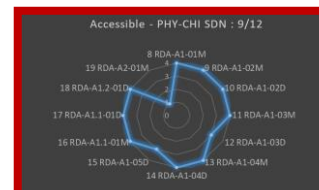


- Pas d'accès direct aux données mais passage par un panier d'extraction
- Les données ne sont pas accessibles automatiquement (c'est-à-dire par un programme informatique)
- Pas de page mémorial sur les métadonnées

SURVAL



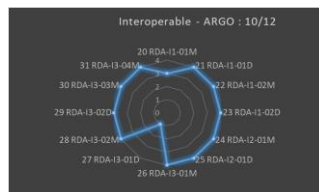
PHYSIQUE-CHIMIE



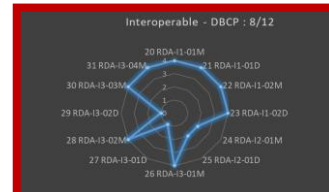
9/17

Auto-évaluations FAIR – PRINCIPE I

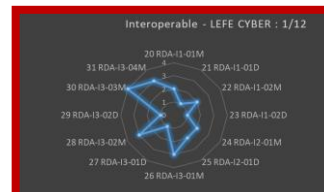
ARGO



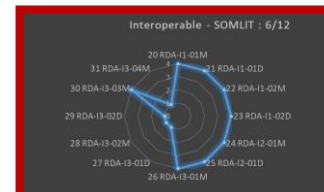
DBCP



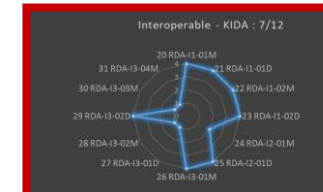
LEFE CYBER



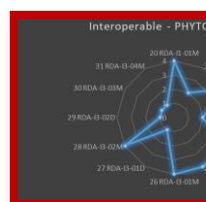
SOMLIT



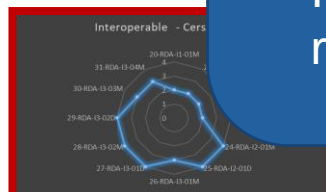
KIDA



PHYTOBIO



CERSA

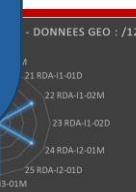


- Pas de “représentation de connaissances exprimée dans un format standardisé compréhensible par l’homme et la machine” (RDF/SPARQL/Websémantique)
- Utilisation de vocabulaire mais pas forcément FAIR
- Références croisées implémentées dans les métadonnées mais très peu dans les données

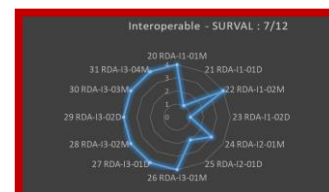
CATOS



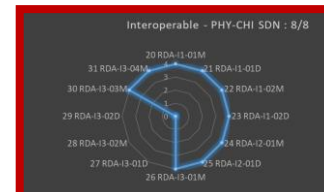
GEO



SURVAL



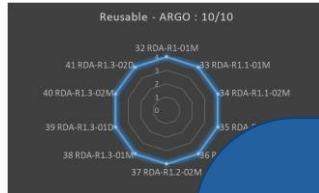
PHYSIQUE-CHIMIE



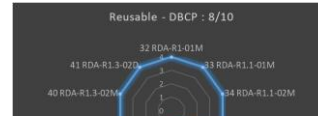
1/17

Auto-évaluations FAIR – PRINCIPE R

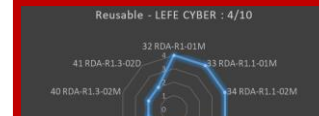
ARGO



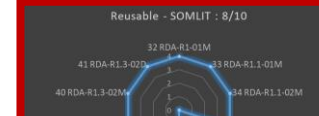
DBCP



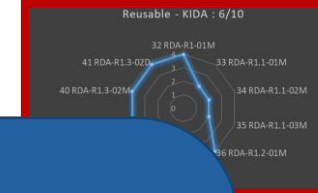
LEFE CYBER



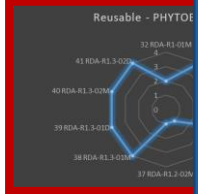
SOMLIT



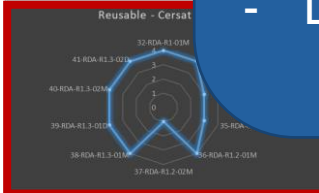
KIDA



PHYTO

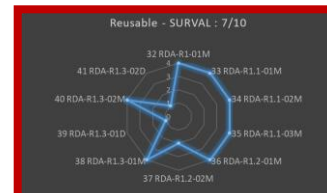


CERSA

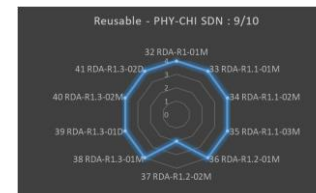


- Pas de licence de spécifiée seulement les conditions d'accès
- Dans certains cas, licence à standardiser
- Pas d'information de provenance selon un langage intercommunautaire (PROV-O)
- Pas de norme utilisée dans certains cas pour les métadonnées (page HTML, développement 'maison'...)
- Données fournis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

SURVAL



PHYSIQUE-CHIMIE



7/17

Auto-évaluation FAIR des données Odatis

- 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 des références qualifiées à d'autres métadonnées et données quand c'est possible (ORCID, Archimer)
 - Utiliser des licences standards (AVISO+)

Intérêt de cette auto-évaluation

- Pour les CDS

- Exercice très intéressant : recul sur les pratiques
- Sensibiliser les thématiciens et parties-prenantes à l'intérêt de la FAIRisation
- 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)

Atelier FIP pour la communauté ODATIS

FIP : FAIR Implementation Profile

- La communauté GO FAIR a lancé le développement de profils de mise en œuvre FAIR (FIP) exploitables par des machines
- Le FIP est une collection de choix de mise en œuvre FAIR faits par une communauté de pratique pour chacun des principes FAIR.
- Effectuer un état des lieux des outils, des technologies et des standards relatifs aux données et métadonnées de votre CDS
- Décrire et exposer les pratiques FAIR actuelles de son CDS/sa filière de données de façon FAIR, sous forme de nanopublication exposée en RDF via Triple Store et réutilisable par d'autres communautés
- Contribution à cette initiative GO FAIR et utilisée par plusieurs communautés (IR d'ENVRI-FAIR, projet FAIR-IMPACT, PNDB, etc.)

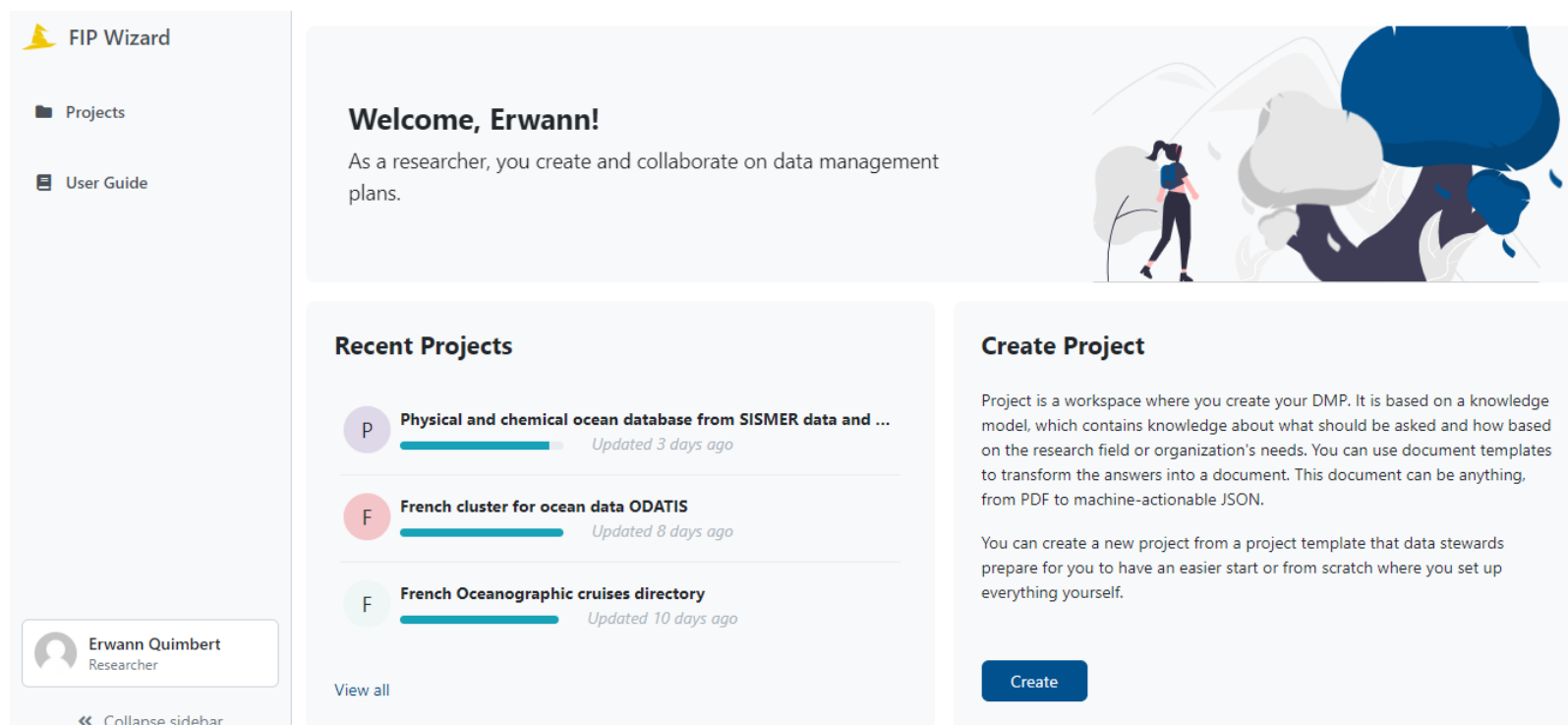
Journée Interopérabilité et Innovation – 23/01/2024

FIP : les questions

FAIR principle	Question	FAIR enabling resource types	Your answers
F1	What globally unique, persistent, resolvable identifiers do you use for metadata records?	Identifier type	e.g. PURL, DOI
F1	What globally unique, persistent, resolvable identifiers do you use for datasets?	Identifier type	
F2	Which metadata schemas do you use for findability?	Metadata schema	
F3	What is the technology that links the persistent identifiers of your data to the metadata description?	Metadata-Data linking mechanism	
F4	In which search engines are your metadata records indexed?	Search engines	
F4	In which search engines are your datasets indexed?	Search engines	
A1.1	Which standardized communication protocol do you use for metadata records?	Communication protocol	
A1.1	Which standardized communication protocol do you use for datasets?	Communication protocol	
A1.2	Which authentication & authorisation technique do you use for metadata records?	Authentication & authorisation technique	
A1.2	Which authentication & authorisation technique do you use for datasets?	Authentication & authorisation technique	
A2	Which metadata longevity plan do you use?	Metadata longevity	
I1	Which knowledge representation languages (allowing machine interoperation) do you use for metadata records?	Knowledge representation language	
I1	Which knowledge representation languages (allowing machine interoperation) do you use for datasets?	Knowledge representation language	
I2	Which structured vocabularies do you use to annotate your metadata records?	Structured vocabularies	
I2	Which structured vocabularies do you use to encode your datasets?	Structured vocabularies	
I3	Which models, schema(s) do you use for your metadata records?	Metadata schema	
I3	Which models, schema(s) do you use for your datasets?	Data schema	
R1.1	Which usage license do you use for your metadata records?	Data usage license	
R1.1	Which usage license do you use for your datasets?	Data usage license	
R1.2	Which metadata schemas do you use for describing the provenance of your metadata records?	Provenance model	
R1.2	Which metadata schemas do you use for describing the provenance of your datasets?	Provenance model	

FIP : l'outil de saisie des FIP

- Outil collaboratif en ligne [FIP Wizard](https://fip-wizard.ds-wizard.org/) permet de générer automatiquement une première version du document FIP



<https://fip-wizard.ds-wizard.org/>

Journée Interopérabilité et Innovation – 23/01/2024

FIP – matrice de convergence

- Les FIP finalisés sont rassemblés dans une matrice de convergence FAIR où chaque colonne représente une communauté et chaque ligne une ressource FAIR déployée pour des principes FAIR spécifiques. De cette façon, les alignements potentiels des FIP de différentes communautés peuvent être identifiés afin d'optimiser la convergence sur la réutilisation des ressources existantes et l'interopérabilité entre les données et services FAIR de chaque communauté.

		Communities									
		A	B	C	D	E	F	G	H	I	J
F	Resource 1	0	1	0	0	0	0	0	0	0	0
	Resource 2	1	1	1	0	1	1	1	1	1	0
A	Resource 3	1	1	0	0	0	0	0	0	1	0
	Resource 4	0	0	1	1	1	1	1	1	1	1
I	Resource 5	1	0	0	0	0	0	0	0	1	0
	Resource 6	0	1	1	0	1	1	1	1	1	0
R	Resource 7	1	0	0	0	0	0	0	0	0	0
	Resource 8	1	0	1	0	1	1	1	1	1	0

FIP Fingerprint FIP reuse

En savoir plus

- Guidelines : <https://osf.io/4bfcy/>

- Informations générales sur les FIP

Reusable FAIR Implementation Profiles as Accelerators of FAIR Convergence, in Grossmann G., Ram S. (eds) Advances in Conceptual Modeling. ER 2020. Lecture Notes in Computer Science, vol 12584. Springer, Cham.

https://doi.org/10.1007/978-3-030-65847-2_13

- FIP supports de formation sur le portail ENVRI FAIR :
<https://training.envri.eu/course/view.php?id=56>



DATA
TERRA



ODATIS

Merci pour votre attention !



23/01/2024

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

Indicateurs du FAIR DATA MATURITY MODEL (FDMM)

Facile à trouver : 7 indicateurs, **tous essentiels**

Identifiants



1. **F1-01M:** Les métadonnées sont identifiées par un identifiant pérenne
2. **F1-01D:** Les données sont identifiées par un identifiant pérenne
3. **F1-02M:** Les métadonnées sont identifiées par un identifiant mondialement unique pérenne
4. **F1-02D:** Les données sont identifiées par un identifiant mondialement unique pérenne
5. **F2-01M:** Des métadonnées riches permettent la découverte des données
6. **F3-01M:** Les métadonnées comprennent l'identifiant des données
7. **F4-01M:** Les métadonnées sont proposées de manière à pouvoir être moissonnées et indexées



Indicateurs du FAIR DATA MATURITY MODEL (FDMM)

12 indicators for **ACCESSIBLE** (8 essentiels, 3 importants, 1 utile)



Protocole d'accès, Authentification, Autorisation, identifiants

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

Indicateurs du FAIR DATA MATURITY MODEL (FDMM)

12 indicateurs for Interopérable (7 importants, 5 utiles)

Standard, FAIR, Lisible par machine, linked data



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

Indicateurs du FAIR DATA MATURITY MODEL (FDMM)

10 indicateurs for Réutilisable (5 essentiels, 4 importants, 1 utile)



Licence, norme communautaire

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