



**DATA
TERRA**



ODATIS

Atelier COPiLOtE 2022

Mercredi 8 juin 2022

CDS-IS-SISMER



10/06/2022

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

Quels produits ont été évalués?

Produits évalués:

- Données des campagnes océanographiques françaises :
 - Données de physique-chimie
 - Données de géosciences
- Données d'échantillons issues des navires de la FOF
 - Echantillons biologiques
 - Echantillons géologiques (base BIGOOD)
- Données géographiques (Sextant)
- Données de monitoring côtier « Surval-Quadrige »
- Catalogue des campagnes françaises

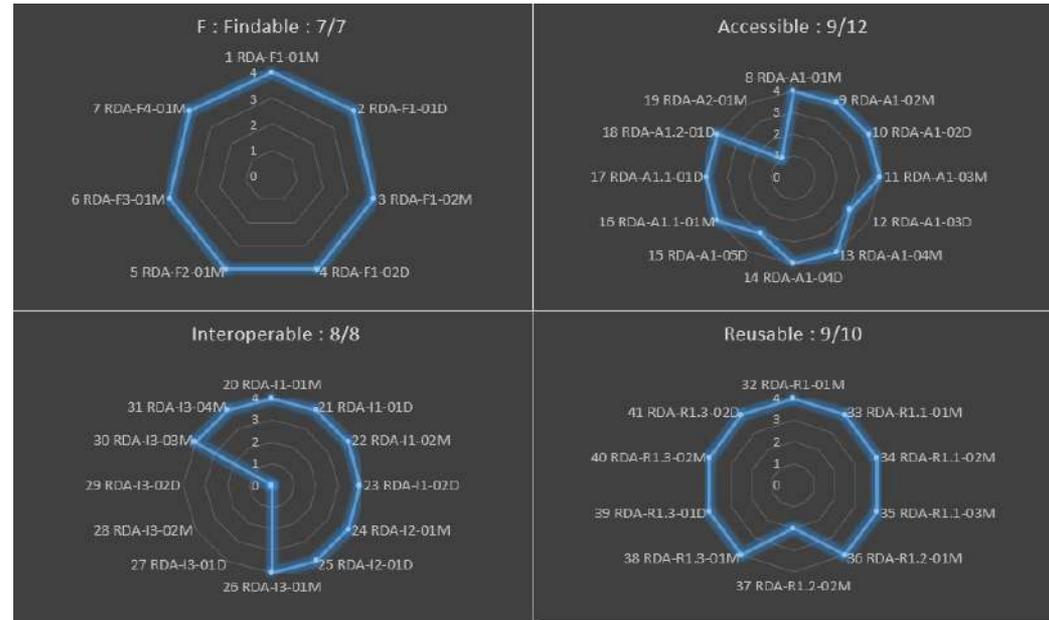
→ Ensemble du périmètre du CDS-IS-SISMER,
différent de celui soumis à CTS

Auto-évaluation n°1 – Données des campagnes

Diffusion via SDN:

F & I entièrement
implémentés lorsque
applicables :

→ DOI, PID, SparQL
endpoint, formats
standardisés, vocabulaire
FAIR (NERC-BODC), etc.



A & R bien implémentés : → Marine-ID, HTTPS, normes
communautaires pour les MD et formats de données, licence indiquée)

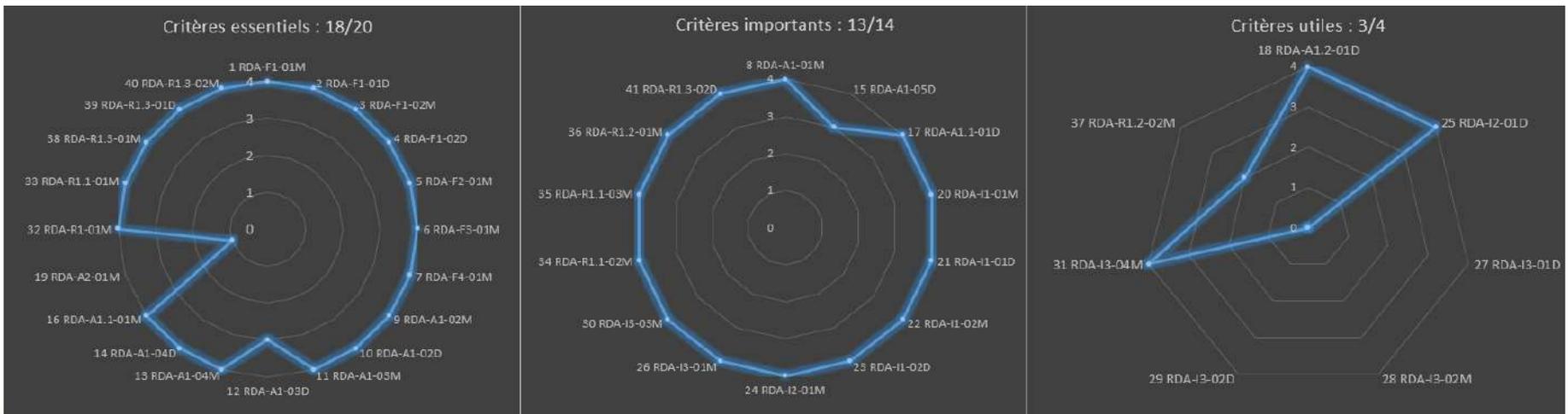
Mais possibilité d'amélioration: → API en surcouche du CDI en cours
d'implémentation, Page mémo des MD, Prov-O étudié dans ENVRI-FAIR

Auto-évaluation n°1 – Données des campagnes

Essentiels: Pas de page memo publique des métadonnées (mais conservés en deprecated), Identifiant renvoyant à un objet numérique (API au dessus du CDI en cours d'implémentation)

Importants: Données accessibles automatiquement (API en cours d'implémentation)

Utiles: Utilisation de PROV-O en cours d'analyse dans ENVRI-FAIR, données d'observation : références à d'autres données non applicable.



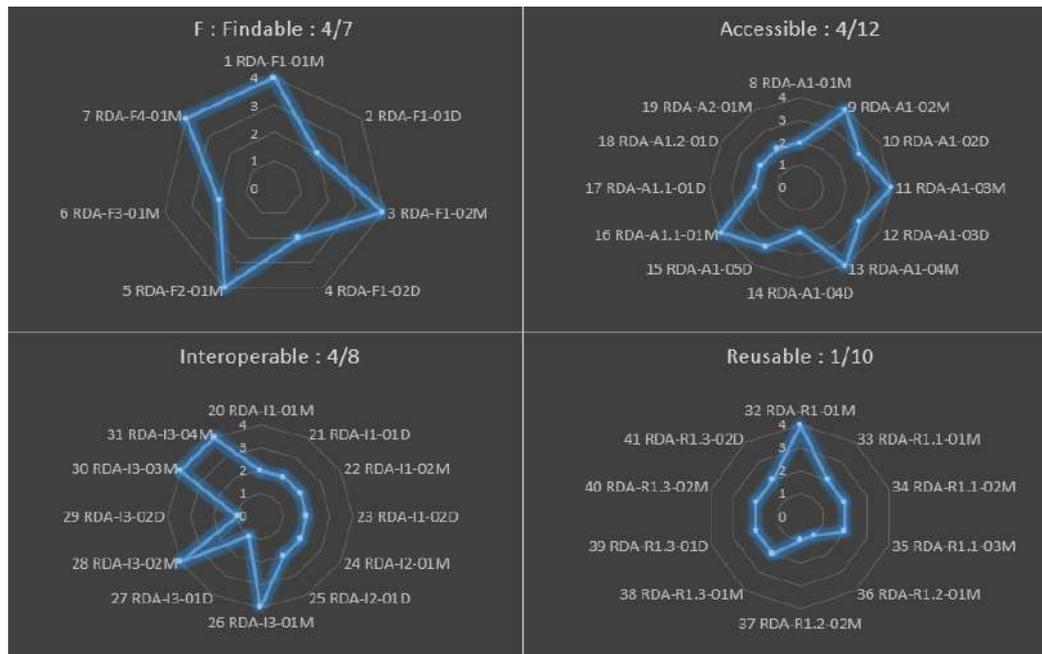
Auto-évaluation n°2 – Données d'échantillons

F : PID IGSN, riches MD, moissonnage via flux WFS. Mais Pas de PID pour données, non inclus dans MD

A : HTTPS, Landing page si IGSN mais information d'accès à ajouter, protocole standard, MA et AAI non implémenté

I : Références qualifiées à d'autres métadonnées et données via code IGSN mais pas pour les données, pas d'exposition RDF pour MD et données, vocabulaire FAIR utilisés pour données et MD mais à optimiser

R : Pas de norme communautaire utilisée pour les données & MD. MD riches mais pas d'information sur la licence ou la provenance.



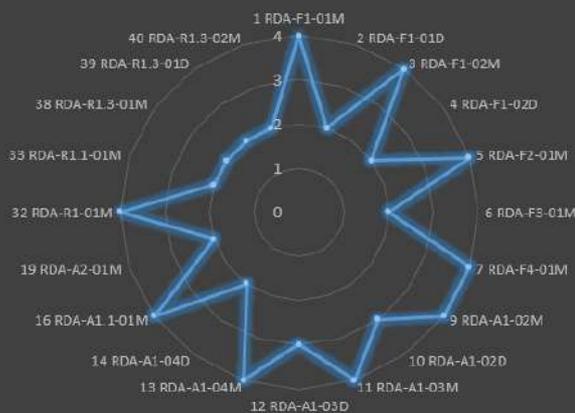
Auto-évaluation n°2 – Données d'échantillons

Essentiels: MD riches pour découverte et réutilisation, moissonnables, accessibles par HTTPS. PID pour MD mais pas données. Information pour accès aux données sur demande en cours d'implémentation. Norme communautaire pour données et MD à étudier.

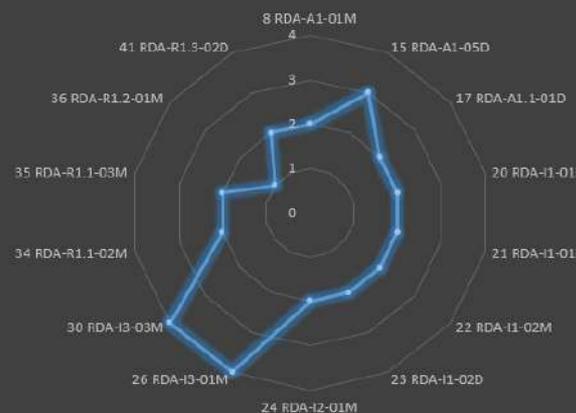
Importants: Références qualifiées (MD). Vocabulaire FAIR (NERC L22) dans MD à optimiser, idem données en RDF via SDN.

Utiles: MD riches en références à d'autres données (mais pas les données), voc. FAIR (P07) utilisé dans les données à optimiser, PROV-O non utilisé, pas d'authentification via protocole d'accès

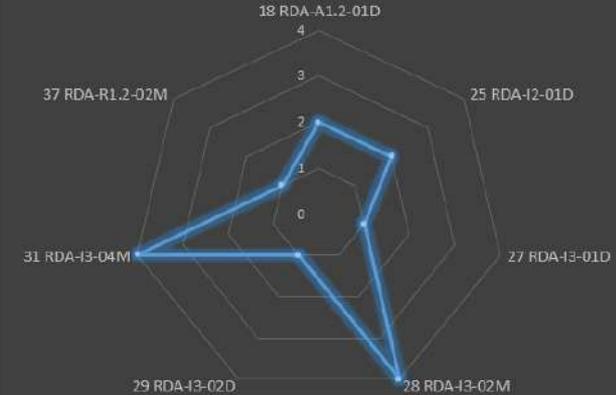
Critères essentiels : 9/20



Critères importants : 2/14



Critères utiles : 2/7



Auto-évaluation n°3 – Données géographiques (Sextant)

Outil évalué plutôt que les données

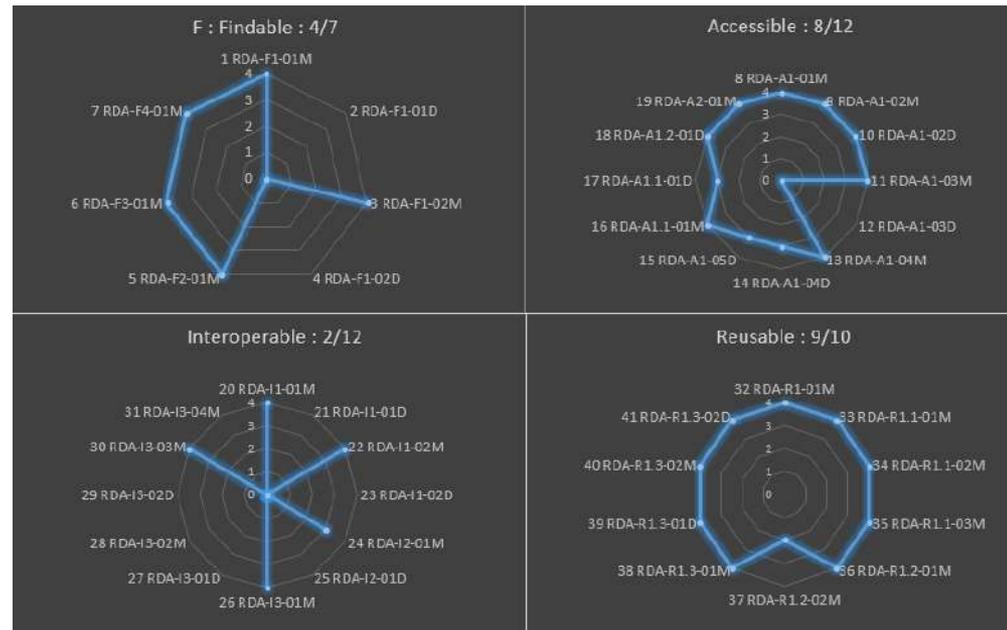
F entièrement implémenté lorsque applicable : UUID, ISO 19115, CSW, etc.

A : Ldap, protocoles normalisés, landing page, etc.

→ Mais protocoles « maisons »
FILE et COPYFILE
propriétaires, non standards et machine-actionable

I : OGC-API-Records, références ressources liées, vocabulaire FAIR en cours d'implémentation.

R : Licence, normes communautaires, attributs riches, mais description de la provenance à étudier (PROV-O).



Auto-évaluation n°3 – Données géographiques (Sextant)

Essentiels: Protocoles standardisés : valables pour WFS, WMS, HTTPS mais pas pour protocoles FILE, COPYFILE

Importants: Vocabulaires FAIR : en cours d'implémentation (Serveur Ifremer), Protocoles automatiques et libres : valables pour WFS, WMS, HTTPS mais pas pour protocoles FILE, COPYFILE

Utiles: Utilisation de PROV-O ou autre langage intercommunautaire à étudier

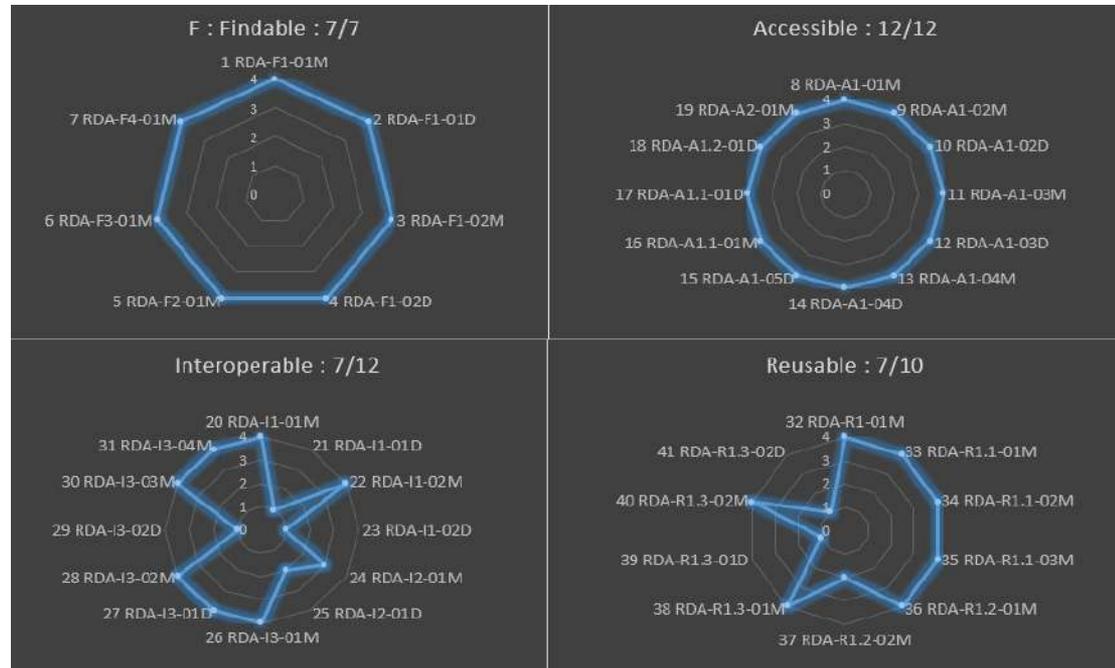


Auto-évaluation n°5 – Données de monitoring côtier « Surval-Quadriges »

Diffusion via
Surval et Sextant:

F & A entièrement
implémentés :

→ DOI, URI (code
Sandre +
datagouv), CSW,
ISO19115 & HTTPS



I : MD en RDF, riches en références à d'autres MD et données. Vocabulaire FAIR en cours d'implémentation (serveur de voc. Odatis, codes Sandre et SAR), Références dans les données via citation de taxon à étudier

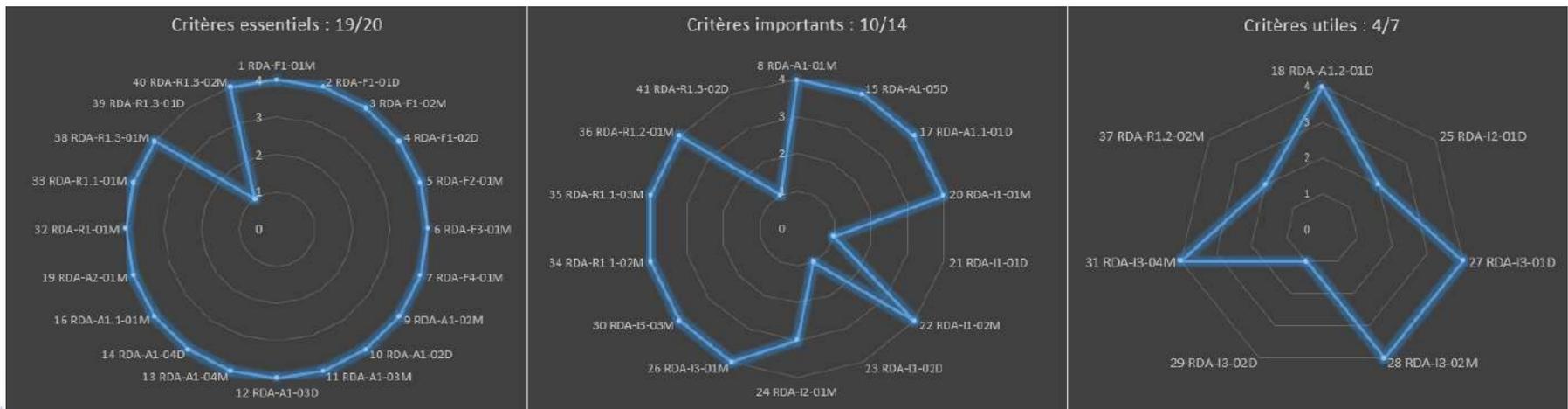
R : MD riches pour réutilisation, norme communautaire & MA + provenance, licence Etalab standard et MA. PROV-O à étudier. Données diffusées en CSV: non normé, non MA, possibilité de diffusion via ODV pourrait être étudiée

Auto-évaluation n°5 – Données de monitoring côtier « Surval-Quadrige »

Essentiels: norme communautaire pour les données non implémentée (CSV)

Importants: pas de rep. des connaissances dans format standardisé et MA pour les données (SPARQL/RDF), Vocabulaire FAIR en cours d'implémentation

Utiles: Vocabulaire FAIR pour les données et utilisation de PROV-O en cours d'examen, références qualifiées à d'autres données (citation taxon) pourrait être envisagé



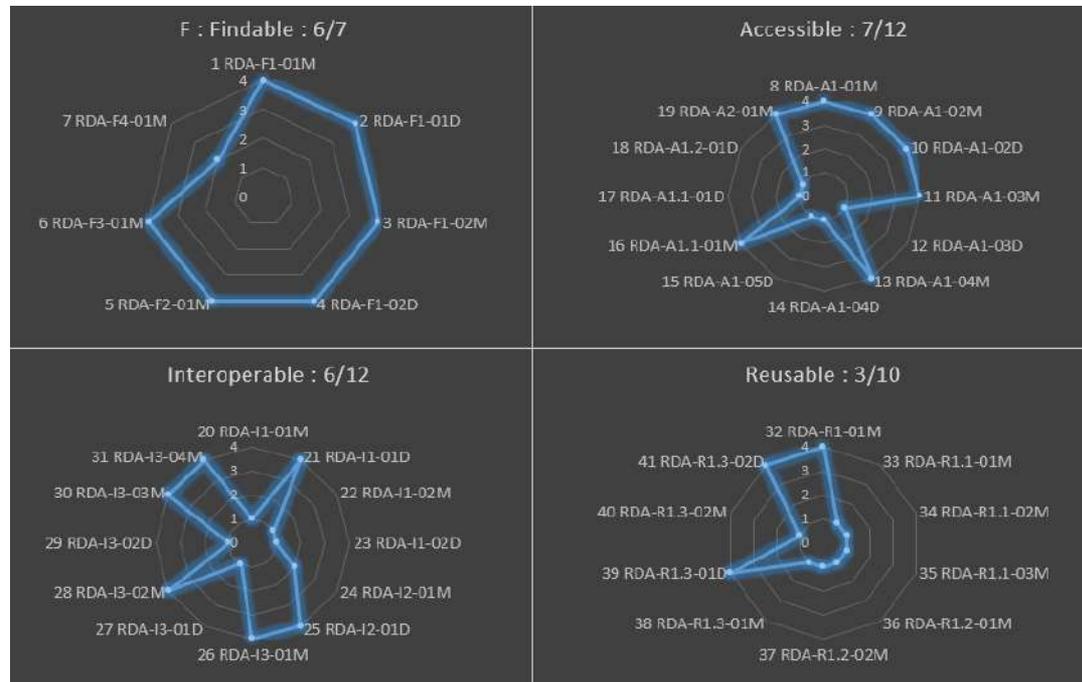
Auto-évaluation n°4 – Catalogue des campagnes

F plutôt bien implémenté :
DOI, description plutôt riche,
etc. Mais pas de moissonnage
possible

A implémenté pour les MD
(DOI, HTTPS, conditions
d'accès), mais pas entièrement
pour les données : accès par
panier (non standard, non MA)

I : Données plus standardisées, voc FAIR, mais pas de références faites. MD riches en références qualifiées (vidéo, publi, campagnes, etc.), mais voc. non FAIR, et pas d'ontologie standardisée utilisée pour rep. des connaissances. Ni données ni MD ne sont exposées en RDF.

R : norme communautaire pour les données (NetCDF, Medatlas) mais pas pour MD. Pas d'information sur la licence ou la provenance, bien que MD riches.



Auto-évaluation n°4 – Catalogue des campagnes

Essentiels: MD riches pour découverte et réutilisation, ID unique et pérenne pour MD et données. Données accessibles mais pas de façon MA, données standardisées mais pas MD, pas d'info de licence, seulement conditions d'accès.

Importants: MD riches en références à d'autres MD, et info d'accès. Données via norme communautaire et format standard. Pas d'info de provenance, pas d'accès aux données automatique possible, MD ont des vocabulaires issus de thesaurus mais pas FAIRisés.

Utiles: MD riches en références à d'autres données (mais pas les données), voc. FAIR utilisé dans les données. PROV-O non utilisé, pas d'authentification via protocole d'accès



Analyses et pistes d'améliorations

→ Critères essentiels plutôt bien implémentés.
FAIRisation dépend beaucoup de l'outil de diffusion.

→ Pistes d'amélioration:

- Etude de PROV-O (ENVRI, Sextant)
- Enrichir les MD avec liens vers d'autres outils de diffusion (SDN, EMODnet, Sextant, etc.)
- Enrichir les MD avec références qualifiées à d'autres MD et données quand c'est possible
- Amélioration de l'utilisation de vocabulaire FAIR (ajout URI, ou utilisation de vocabulaire FAIR existant: NERC-BODC, etc.)

Conclusion

Une méthode très utile, mais optimisable:

Déclinaison du critère en fonction de la donnée ou de la métadonnée: ne fonctionne pas bien pour tous **les produits** (identifiants: que considérer comme « identifiant pérenne de la donnée », références à d'autres données : que considérer comme « référence à la donnée », etc.)

Critères liés aux MD : évaluation de la FAIRness de l'outil de diffusion plus que des données, or multiplicité possible.

Critères 27 à 29 (référence à d'autres données): pour données d'observation « brutes », pas forcément applicable.

Critère 19 (MD disponibles après effacement des données): est-ce souhaitable pour une BDD évolutive ? Snapshots ?

Questions / Points de discussion?



**DATA
TERRA**



ODATIS

Merci de votre attention



10/06/2022

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