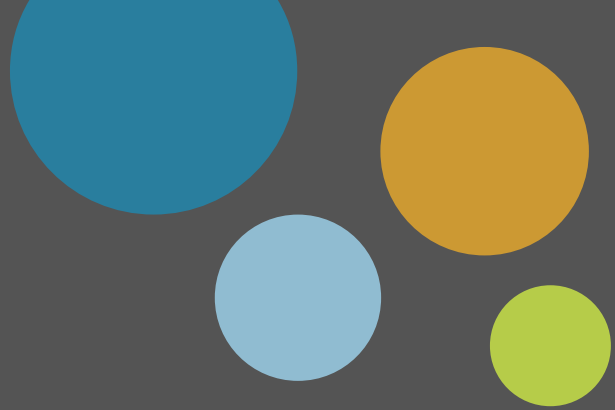


Gestion des données au SED00

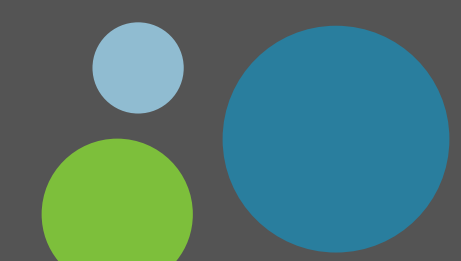
François ANDRE



06 mai 2021



Le SEDOO





Pôle numérique de l'OMP

Missions liées aux données

SEDOO

Gestion de données scientifiques

- Stockage des données
- Distribution
- Services à valeur ajoutée

DSI

Mise à disposition de l'infrastructure informatique

- Serveurs
- Stockage + **Archivage**
- Réseau
- Calcul

Type de données gérées

Données in situ d'observation de la Terre



- Données liées aux domaines: **Atmosphère, Océans, Surfaces continentales.**
- Métadonnées
- Données auxiliaires
- **Sites WEB**

+ASTRONOMIE

Périmètre d'action

LOCAL

Laboratoires de l'OMP

Services d'observation de l'OMP

NATIONAL

Infrastructures de Recherche nationales

Pôles nationaux de données

Campagnes

INTERNATIONAL

Infrastructures de recherche H2020

Programmes internationaux

PARTICIPATION À DES
GROUPES DE TRAVAIL

Mesures vs Données



Les chaînes d'acquisition des mesures sont souvent gérées par les laboratoires

Toutes les données de l'OMP ne sont pas gérées par le SEDOO



Ressources

Une équipe de développeurs

Ressources humaines:

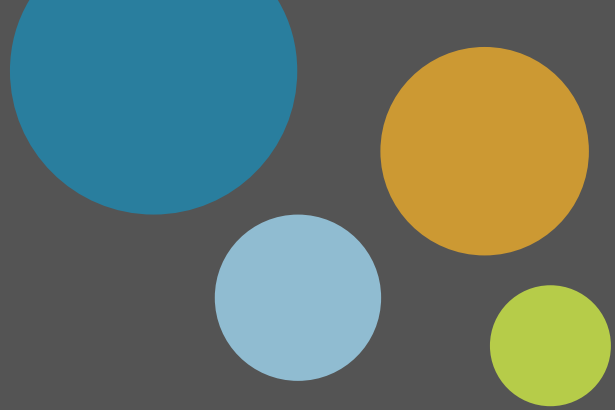
14 développeurs (50 % permanents – 50 % CDD)

Ressources matérielles:

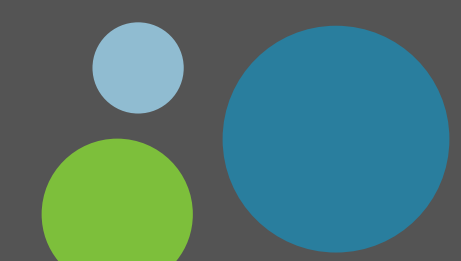
Une dizaine de serveurs

200 To de stockage





FAIR & TRUST



Mise en place d'une architecture idéale

On en est loin,
Mais on sait où aller

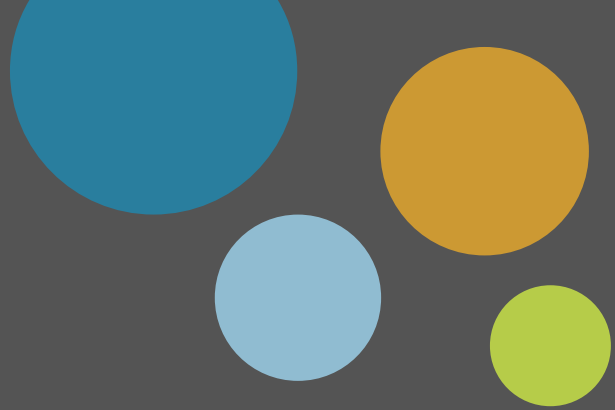


Les principes **FAIR** permettent d'évaluer la **qualité de la donnée distribuée**

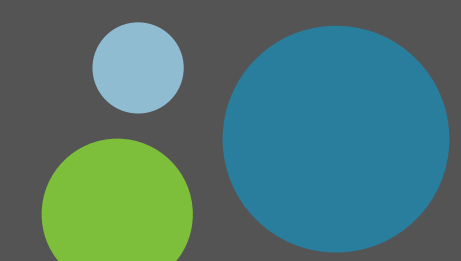
Les principes **TRUST** permettent d'évaluer notre **maturité en tant qu'entrepôt**



PRAGMATISME



FAIR





F comme Findable

La fiche de métadonnées au centre de tout

F1: Association avec un identifiant unique

F2: Description riche

F3: Affichage de l'identifiant unique dans la fiche de métadonnées

F4: Système de recherche des métadonnées



Catalogage

Exemple: <https://mtropics.obs-mip.fr/catalogue-m-tropics/>

The screenshot displays the MTROPICS catalog interface. At the top, there are 'RESET' and 'SEARCH' buttons. Below the search bar, a list of sites is shown with checkboxes: Dong Cao, Vietnam; Houay Pano, Laos; Huay Ma Nai, Thailand; Maddur, India; Mule Hole, India; and Nyong, Cameroon. A 'Parameters' section is also visible. The main area shows search results for '6 min rainfall data, Huay Ma Nai, Thailand'. A detailed view of this dataset is shown on the right, including an abstract, spatial extents, temporal extents, and sites. The abstract mentions 'Valentin et al., 2008' and provides a DOI link. A large blue banner with the text 'Catalogues adaptables' is overlaid on the bottom part of the screenshot.

Results found: 27

Full text search

Sites

- ☐ Dong Cao, Vietnam
- ☐ Houay Pano, Laos
- ☐ Huay Ma Nai, Thailand
- ☐ Maddur, India
- ☐ Mule Hole, India
- ☐ Nyong, Cameroon

Parameters

Temporal extents

Spatial extents

6 min rainfall data, Dong Cao, Vietnam

MTROPICS

6 min rainfall data, Houay Pano, Laos

MTROPICS

6 min rainfall data, Huay Ma Nai, Thailand

MTROPICS

Hydrochemical data, Maddur, India

MTROPICS

Hydrochemical data, Mule Hole, India

MTROPICS

6 min rainfall data, Huay Ma Nai, Thailand

MTROPICS

DATASET

DOWNLOAD

Abstract

Spatial extents

Temporal extents

Sites

Parameters

Contacts

Spatial extents

Dataset history

If using this dataset, please cite Valentin et al., 2008 <https://doi.org/10.1016/j.agee.2008.06.004>

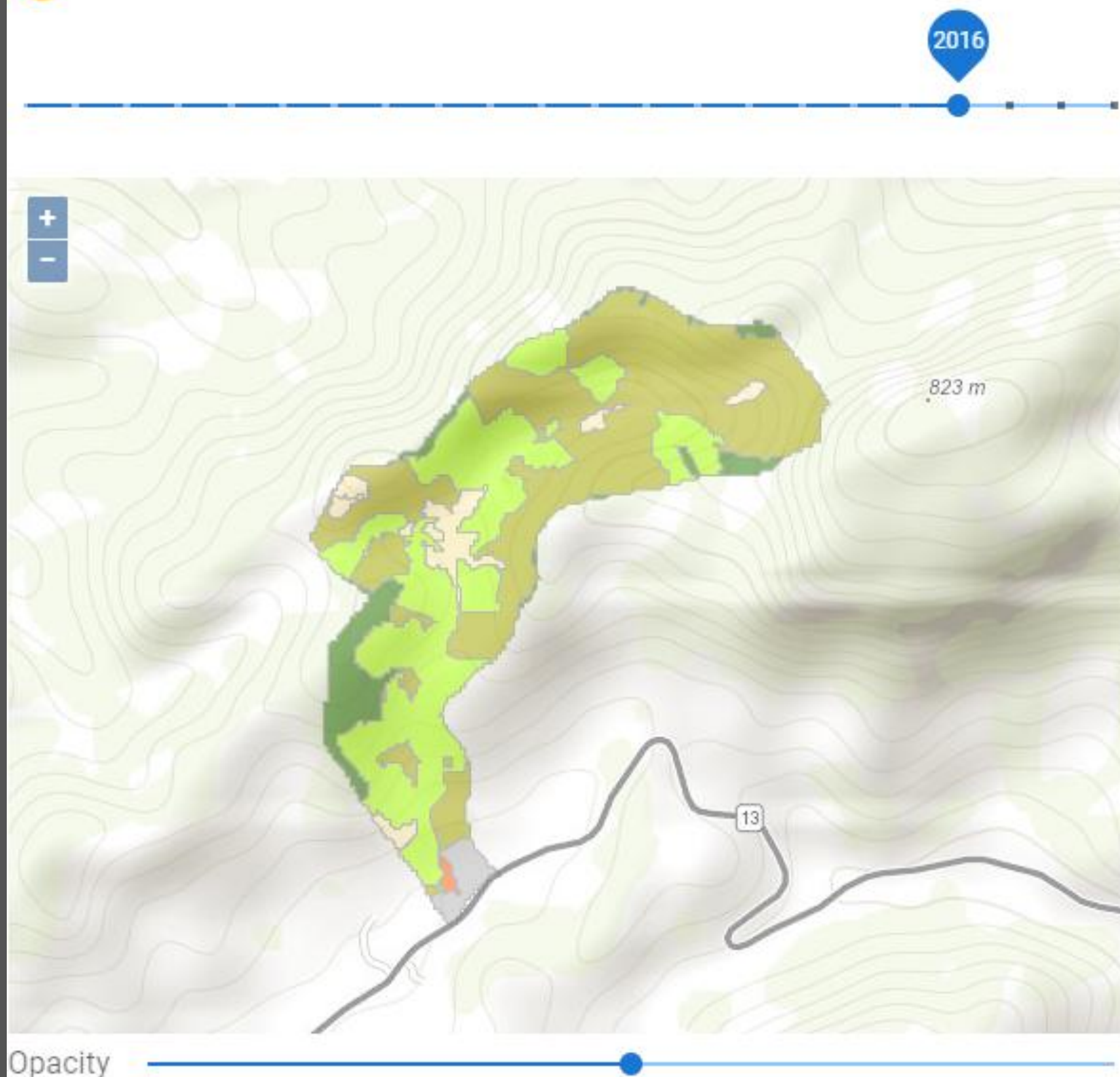
6-min time-step rainfall (mm) data from the Huay Ma Nai catchment weather station (Thailand).

1342

• **Fonctionnalités:** Stockage de métadonnées hétérogènes, Affichage personnalisé, Edition des métadonnées, Gestion par projet, Intégrable dans un site Wordpress

• **Point d'entrée vers le moissonnage :** **Datacite**, autres portails de métadonnées.

MISE EN ŒUVRE DES PRÉCONISATIONS INTERPÔLES



Exemple d'affichage personnalisé

Mise en place d'une visualisation permettant de comprendre les données avant de les télécharger



DOI

Hugonnet, R., McNabb, R., Berthier, E., Menounos, B., Nuth, C., Girod, L., Dussailant, I., Brun, F., & Kääb, A. (2021). Accelerated global glacier retreat in the twenty-first century - Dataset. [Data set]. Theia. <https://doi.org/10.7927/H4T9-6925>



Data access

ELEVATION CHANGE MAPS

TIMES SERIES

Area

Alaska / Western Canada and USA

Period

Jan. 2000 - Dec. 2004

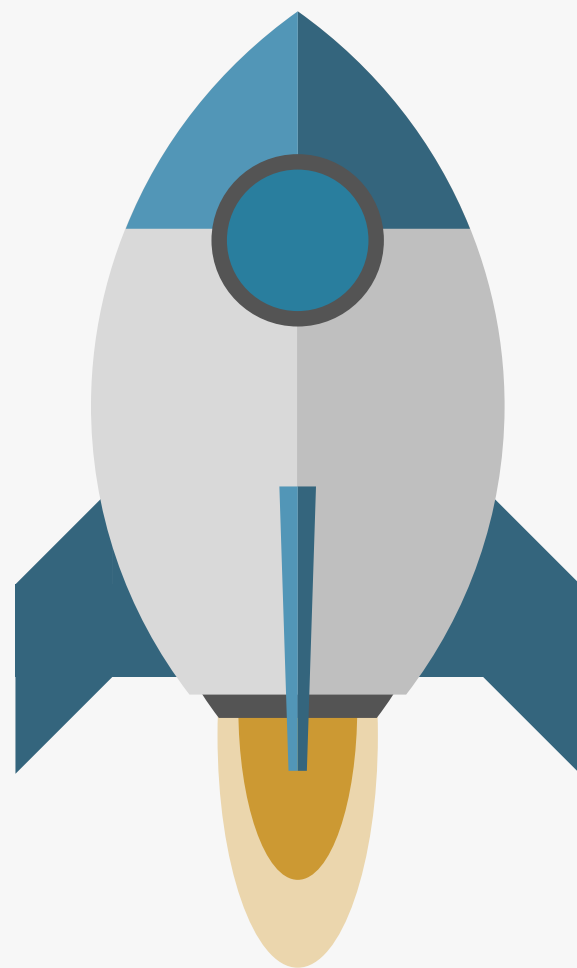
DOWNLOAD

Exemple d'affichage personnalisé

Mise en place d'un bloc de téléchargement permettant la mise à disposition d'un jeu de données de 50 go.



Points forts



Faible coût de mise en œuvre

Valorisation de la donnée

Offre intégrée pour les campagnes

Site Wordpress + Catalogue + DOI + Stockage



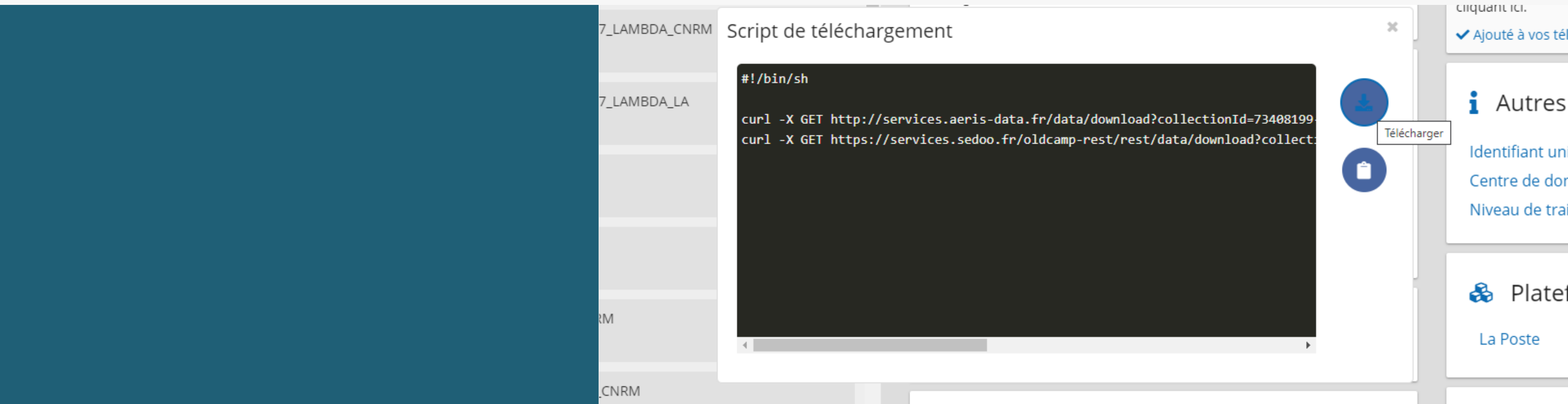
A comme Accessible

A1: Extraction des données / métadonnées via des protocoles ouverts

A2: Les métadonnées sont accessibles alors que les données ont disparues



Accessibilité des données via HTTP(s)



Cette méthode **n'est pas mise en œuvre pour tous les jeux de données**

Cette méthode **n'est pas adaptée à toutes les volumétries**

Une **documentation** spécifique à chaque jeu de données **devraient être ajoutée**

Ce n'est pas l'accès privilégié demandé par nos utilisateurs

L'ACCÈS VIA D'AUTRES
PROTOCOLES SONT ÉTUDIÉS



Accessibilité des métadonnées

DOI
Tombstones

Mise en place manuelle

Permet aussi d'indiquer une version non disponible d'un jeu de données.

MULE HOLE AND MADDUR DATA

The datasets for the measurements at the Indian sites of Mule Hole and Maddur are now split into two separate parts. You can access it by clicking on the following links.

Mule Hole

- [Hydrochemical data](#)
- [Hydrological data](#)
- [Meteorological data](#)

Maddur



I comme Interoperable

-
- I1:** Données et métadonnées peuvent être comprises par des automates
 - I2:** Les vocabulaires utilisés sont FAIR
 - I3:** Les métadonnées & données lient sémantiquement d'autres données et métadonnées



DATA CITE

ODATIS

Limites de l'interopérabilité

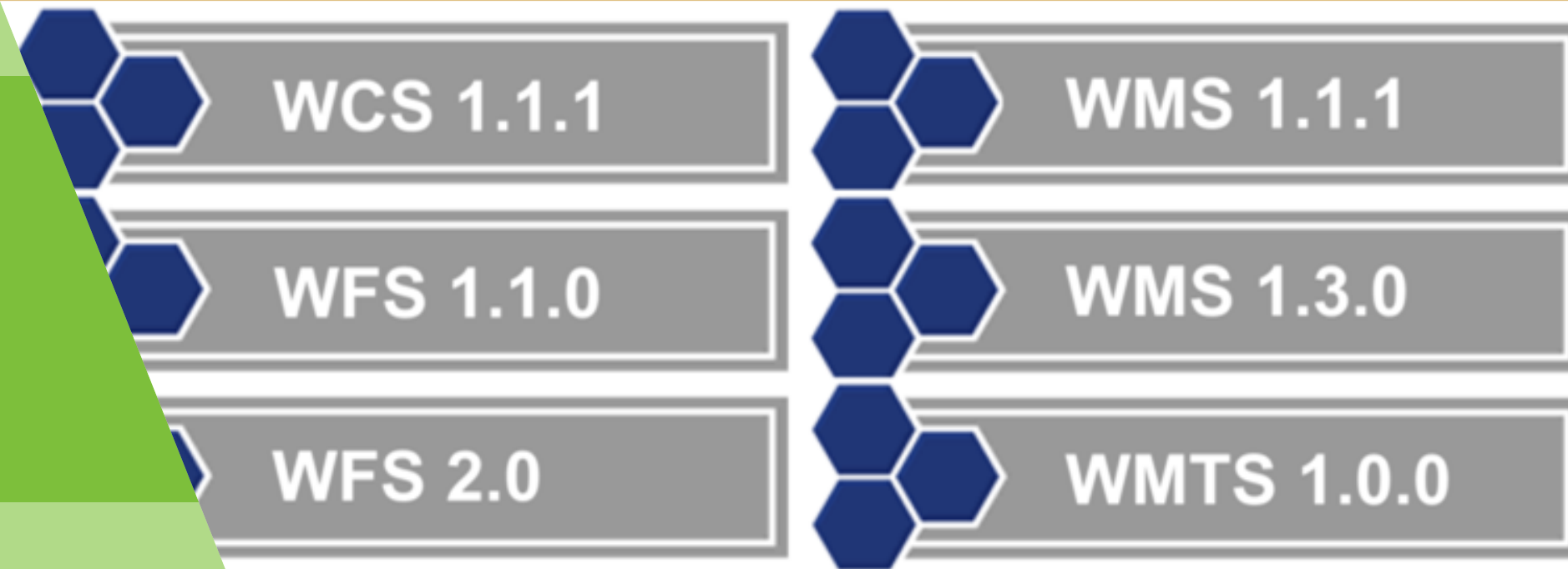
Vocabulaires, protocoles & granularités différentes

→ Mise en place au cas par cas

→ Choix du standard

FORMATS

PROTOCOLES



Limites de l'interopérabilité

Protocoles d'interopérabilité parfois inadaptés

Formats peu installés dans les communautés

→ Complexité de mise en œuvre / Dette technique



R comme Reusable

R1: Licence sur les données

R2: Gestion de la provenance

R3: Respect des standards du domaine



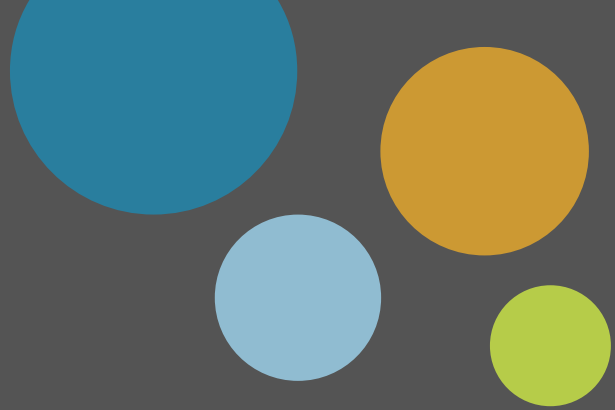


Cas d'école: les licences

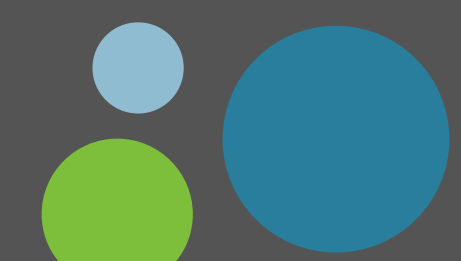


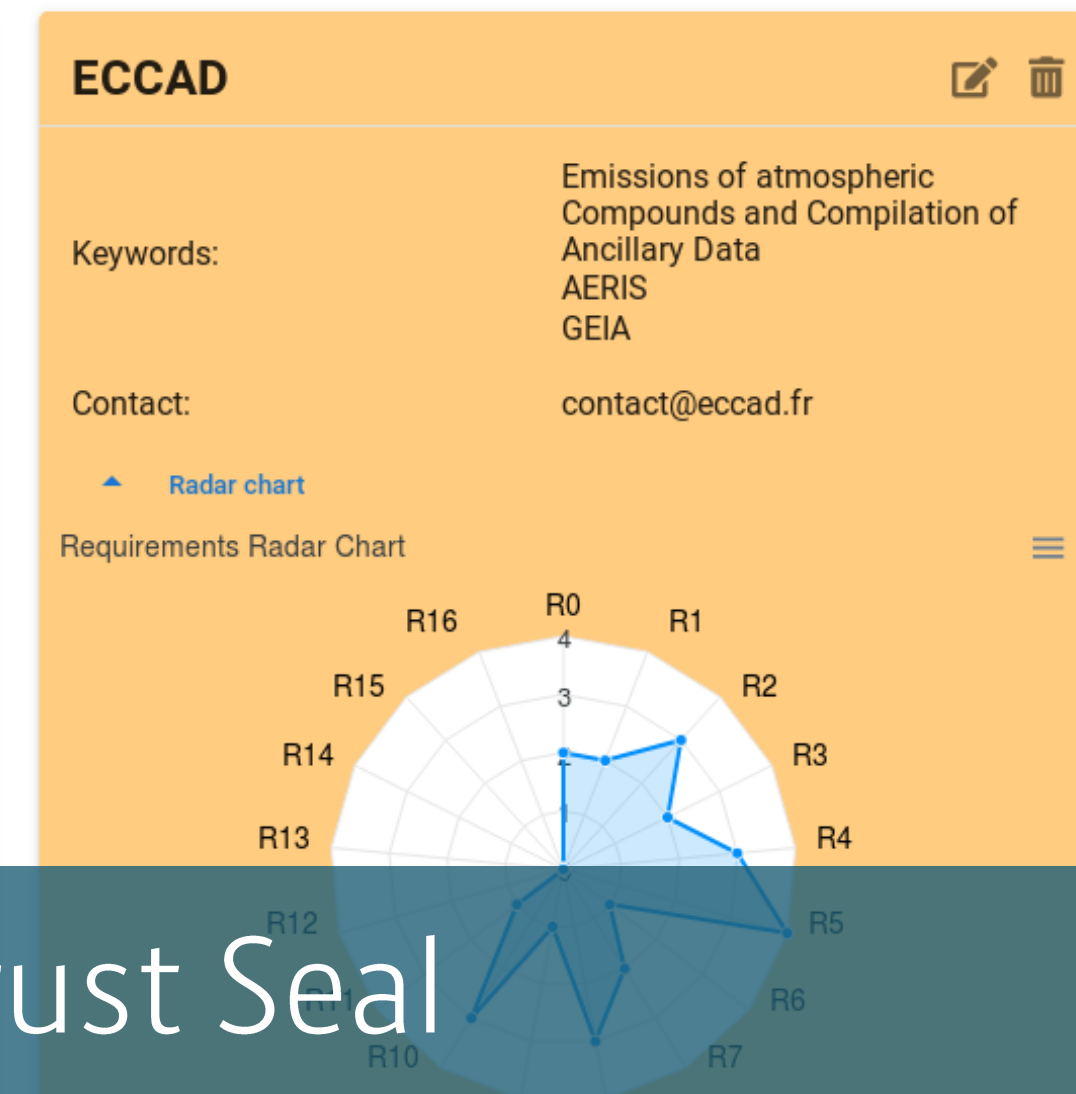
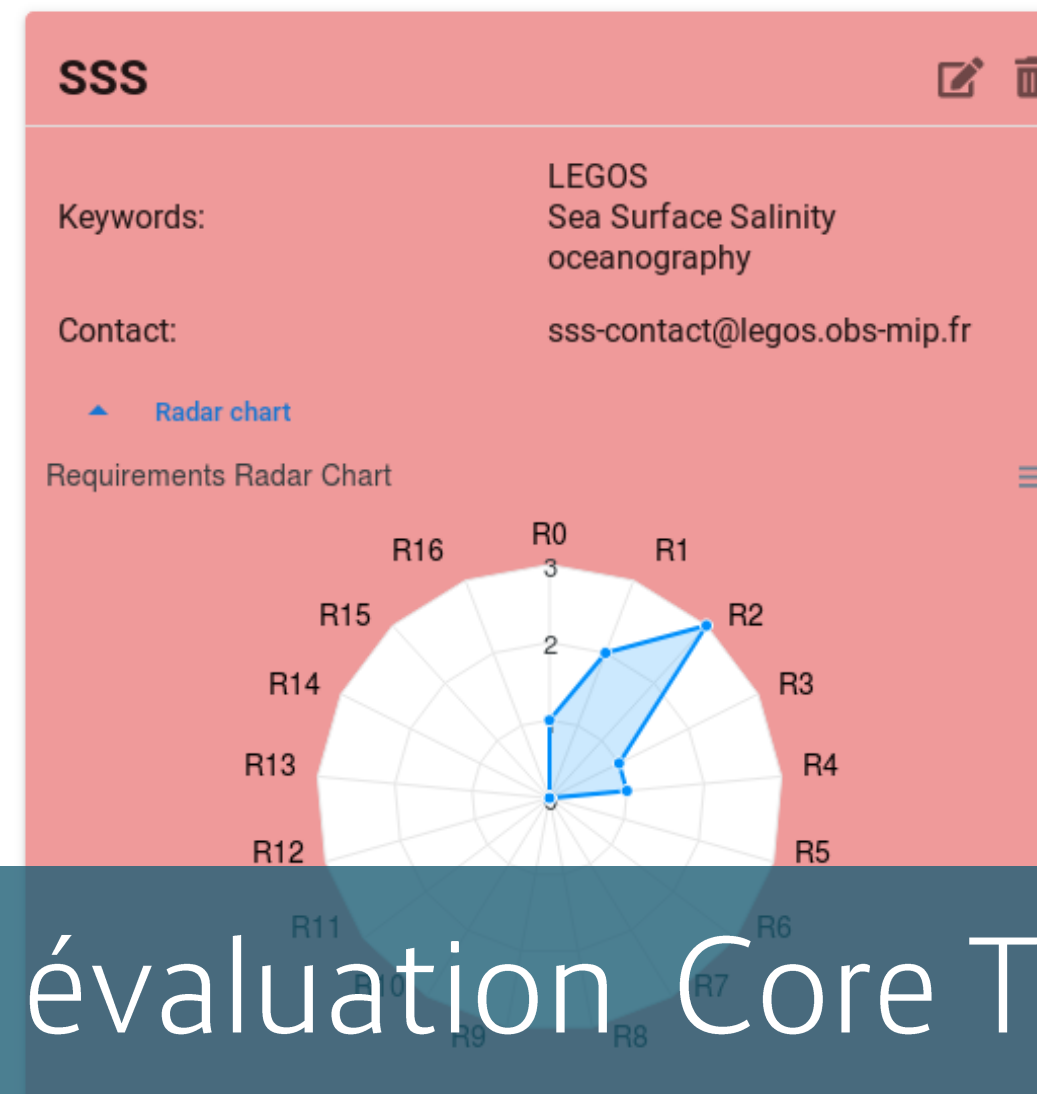
Qui décide ?

-
- Acteurs variés (internationaux , publics/privés, ...)
 - Utilisateurs variés



Trust





Outil libre d'auto évaluation Core Trust Seal

L'outil CRUSOE va être repris par le **Comité National de la Science Ouverte**



Organisation

R1 à R6

Pérennité

Expertise

Maturité (licences)

FAIR

SNO &
SEDOO



Gestion des données

R7 à R14

DMP

Procédures documentées

Qualification de la données

Découverte et Réutilisabilité

FAIR

SEDOO



Infrastructure

R15 et R16

Stockage / Archivage

Sécurité

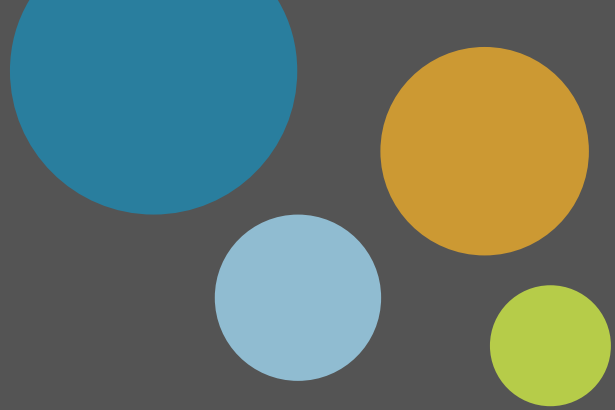
DSI &
SEDOO



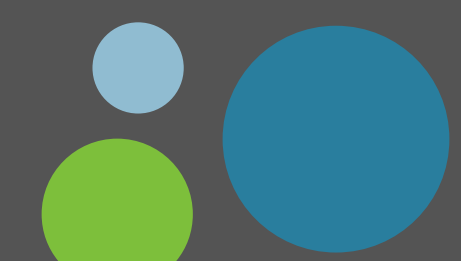
Certification de IAGOS en cours

Doubles challenges

- **Services manquants** (procédures documentés, archivage pérenne, ...)
- **Certification par SNO** → Collaboration avec la partie «Mesure»



Conclusion





Apports des démarches FAIR / TRUST

Explicitation des processus existants

Identification des lacunes

Valorisation des acteurs

Intégration avec les autres défis :

**Volumes des données / Logistique des données /
Conteneurisation / Réduction de la dette technique / Lien
avec les futurs centres de données locaux / Services riches
autour des données / Mise en place d'une gouvernance des
données / ...**





KEEP
CALM
AND
IGNORE
STANDARDS

Pragmatisme

La **maîtrise des données** que nous stockons et des outils que nous utilisons est primordiale.

La mise en place de standards peut s'avérer coûteuse et porteuse de dettes techniques



Thank You / Merci

You can visit us on www.sedoo.fr



Jun Akizaki - <http://thepopp.com>

Used Font: [Route 159 Family](#) & [Open Sans Family](#)

Icon: [Font Awesome](#), the author is [Dave Gandy](#) (Changed the color by Photoshop)
[Typicons](#), the author is [Stephen Hutchings](#) (Changed the color by Photoshop)

Offre d'hébergement WEB

<https://www.sedoo.fr/hebergement-wordpress/>

An International Study of the Marine
Biogeochemical Cycles of Trace Elements and
Isotopes



[HIGHLIGHTS](#)



[DOoR PORTAL](#)



[CRUISES & MAPS](#)



[COOKBOOK](#)



[PUBLICATIONS](#)



[MAILING LIST](#)

News

2021 CALL FOR SCOR

Science Highlights

~150 sites institutionnels

Follow us

Tweets by @geotracés



SEDoo | Observatoire Midi-Pyrénées