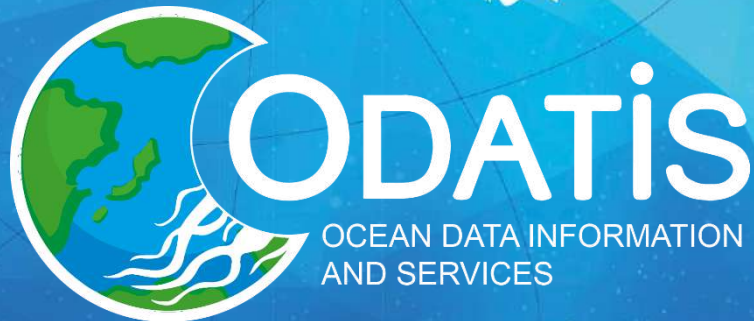
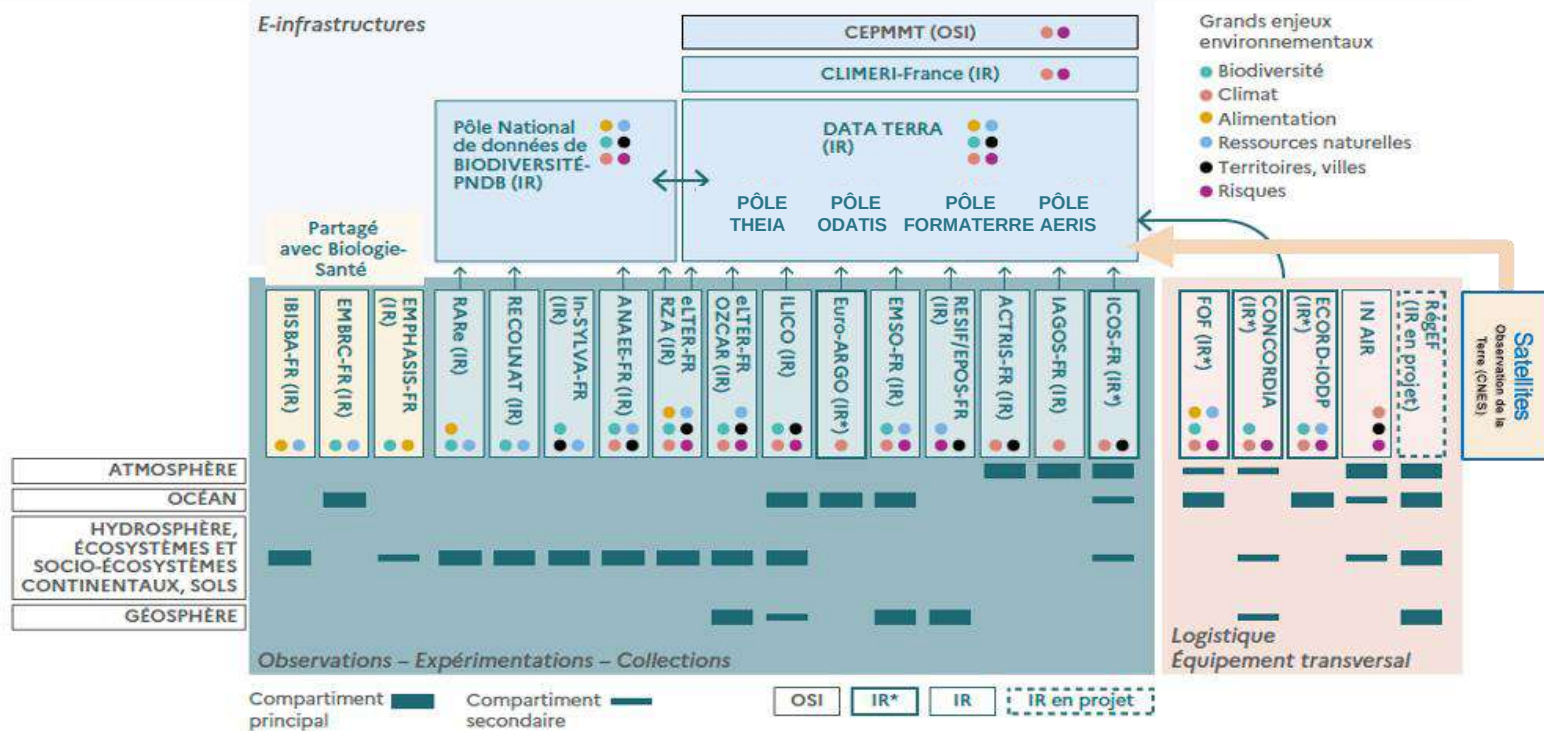




Présentation de l'architecture Gaia Data « Exigences » vis-à-vis des Pôles

Joël Sudre (joel.sudre@data-terra.org)

Rappel de l'IR Data Terra dans le paysage national



Cheminement de la (meta)donnée: CDS (producteur) -> Pôle -> IR Data Terra

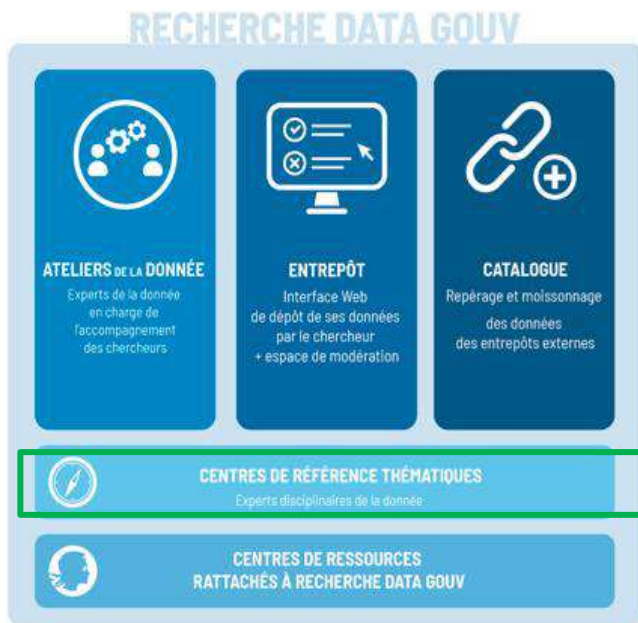
Producteur -> (Catalogue du CDS, de l'IR ILICO) -> Catalogue du pôle -> Catalogue de Data Terra



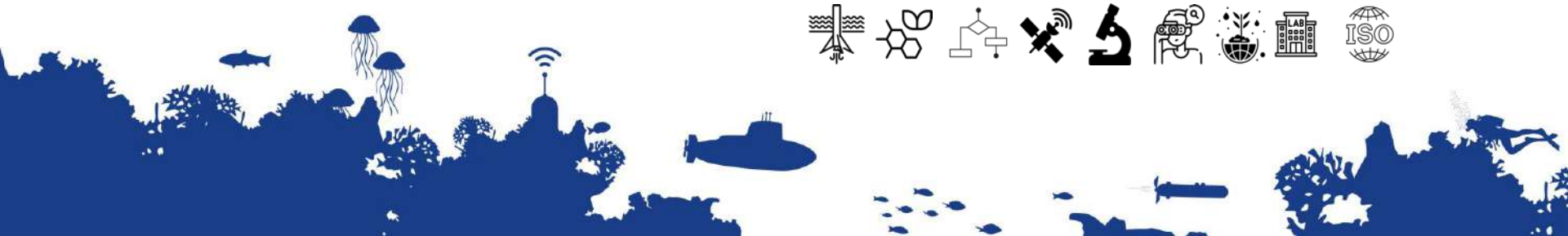
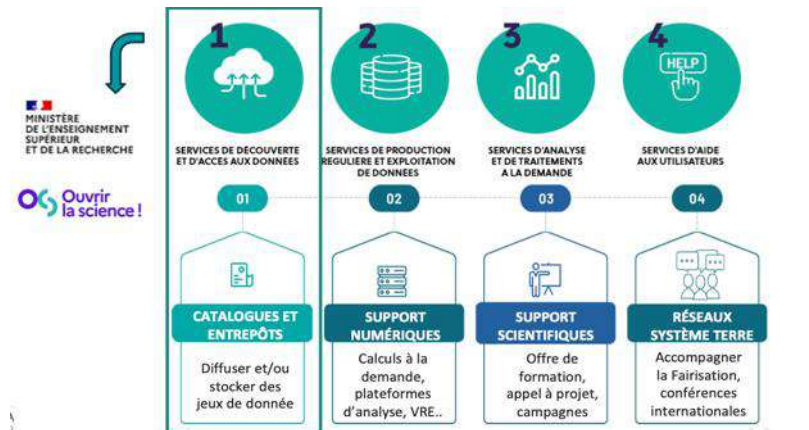
Recherche Data Gouv

Data Terra => Centre national de Référence Thématique système Terre et Environnement

Un écosystème au service du partage et de l'ouverture des données de recherche FÉDÉRER, ACCOMPAGNER, PARTAGER, OUVRIR, RÉUTILISER



- **Les centres de référence thématiques de Recherche Data Gouv :**
- Définissent les normes internationales de description des données
- Définissent les normes internationales de diffusion des données (ouverture, période d'embargo, accès restreint)
- Définissent et diffusent les bonnes pratiques de collecte, documentation, traitement, et diffusion des données
- Définissent la liste des entrepôts de données de référence de leur domaine thématique (nationaux et/ou internationaux) vers lesquels orienter les chercheurs pour le dépôt des données et que Recherche Data Gouv moissonnera
- Contribuent à la définition de l'arborescence thématiques des données de l'entrepôt Recherche Data Gouv
- Soutiennent l'articulation entre les dispositifs thématiques spécialisés et Recherche Data Gouv



Préconisations générales!



Sur les données du système Terre et de l'environnement les jeux de données (+ de 15000!) sont encore très hétérogènes:

- Nécessité de « FAIRISER » les données ET les métadonnées
- Utilisation du web sémantique et des serveurs de vocabulaire
- Avoir des ontologies qui se correspondent entre les différents domaines
- Mettre en place des formations sur les bonnes pratiques et sur le FAIR:
 - ➡ Rendre l'utilisation des entrepôts nationaux courants
 - Changer les pratiques sur les données
 - Prendre l'habitude d'écrire un PGD



- Dans les entrepôts
- Dans les différents catalogues:
 - des CDS
 - des Pôles (ODATIS mais aussi les autres)
 - des IRs d'observation
- Dans les VREs (Jupyter – Galaxy)
- Pour la visualisation



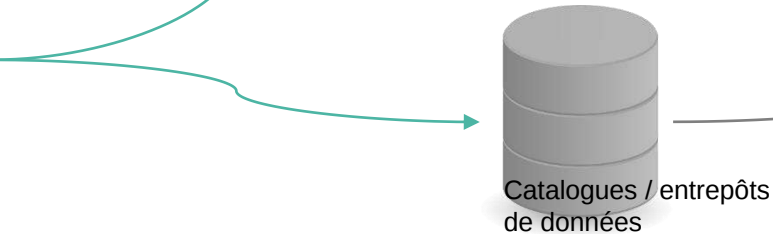
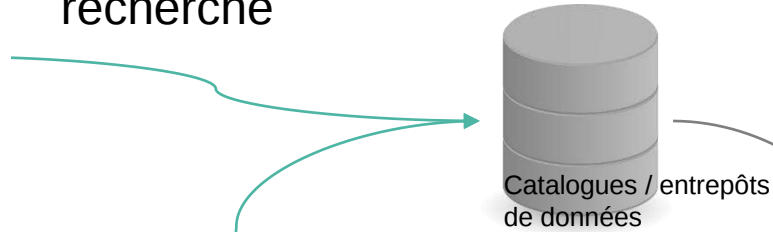
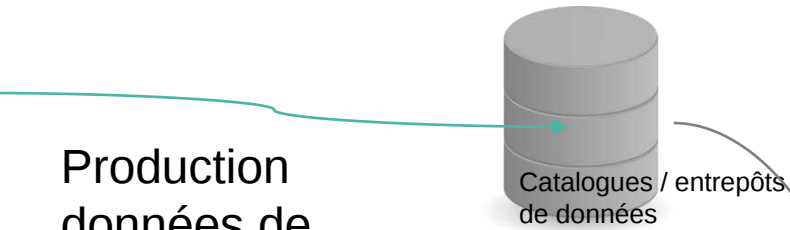
Où chercher ?
Quels catalogues ?
Où est la donnée source ?
Quel schéma de métadonnées ?



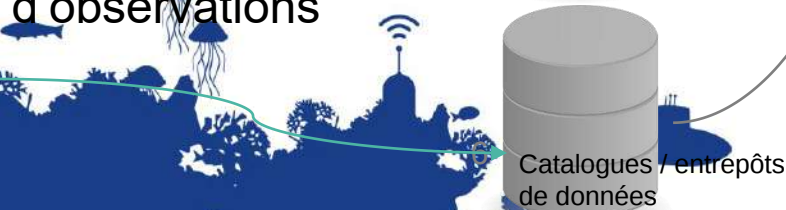
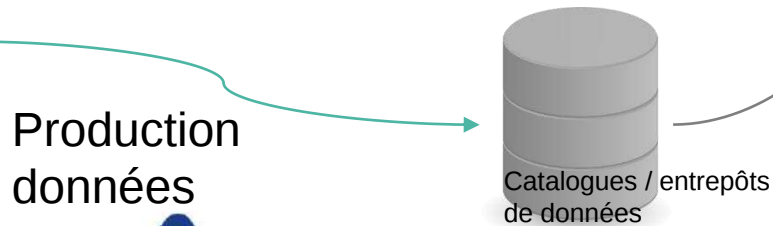
Utilisations diverses des (meta)données



Production
données de
recherche



Production
données
d'observations



IR Data Terra



Méta catalogue –
Catalogue unifié

Recherche facilitée
Recherche facilitée

Chercheurs



- Choix du modèle pivot facilitant la connaissance et la recherche

Rappel : les pôles de données, SNO, CDOS produisent des données d'observations

- Un modèle pivot basé sur la description d'un jeu de données
- Un modèle pivot généraliste et non thématique

➡ Un modèle pivot basé sur le standard **DCAT V2.0**

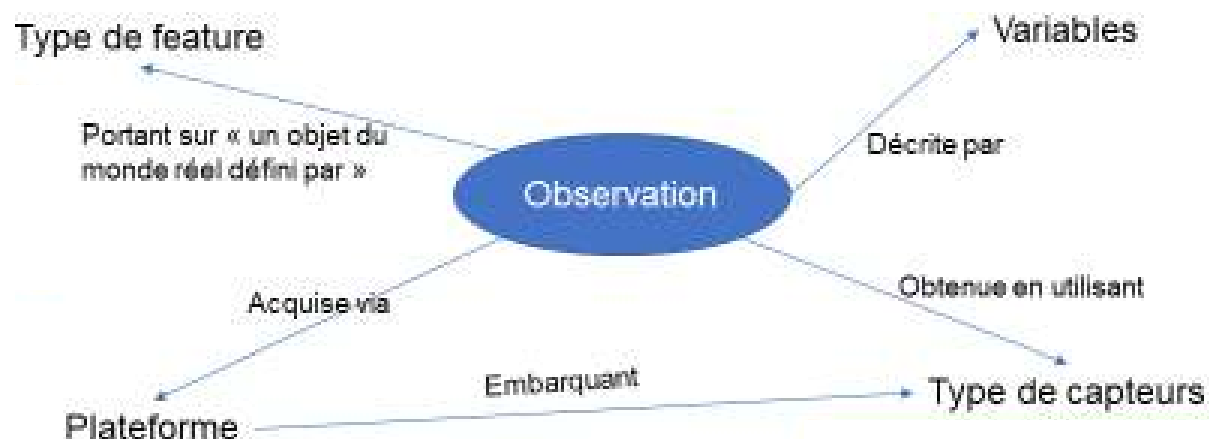


- Mise en place d'un vocabulaire commun

Rappel : chaque pôle, thématique possède ses propres définitions, parfois son propre vocabulaire pour des données d'observation

- Un vocabulaire commun basé sur la structuration des données d'observations

⇒ **Mise en place de la colonne vertébrale des vocabulaires de DataTerra**



- Connaissance des schémas et granularité des datasets des différents pôles de données

Des schémas de métadonnées variés et des implémentations / utilisations différentes :

- **ISO 19115-1**
- **ISO 19115-3**
- **DCAT**
- **SOSA**
- **STAC ?**



➤ Réalisation du mapping

Des schémas de métadonnées variés et des implémentations / utilisations différentes :

- ISO 19115-1 → modèle pivot basé sur DCAT
- ISO 19115-3 → modèle pivot basé sur DCAT
- DCAT → Modèle pivot basé sur DCAT
- SOSA
- EML ?
- STAC ?



- Restitution des informations selon le modèle pivot

Déploiement de composants permettant :

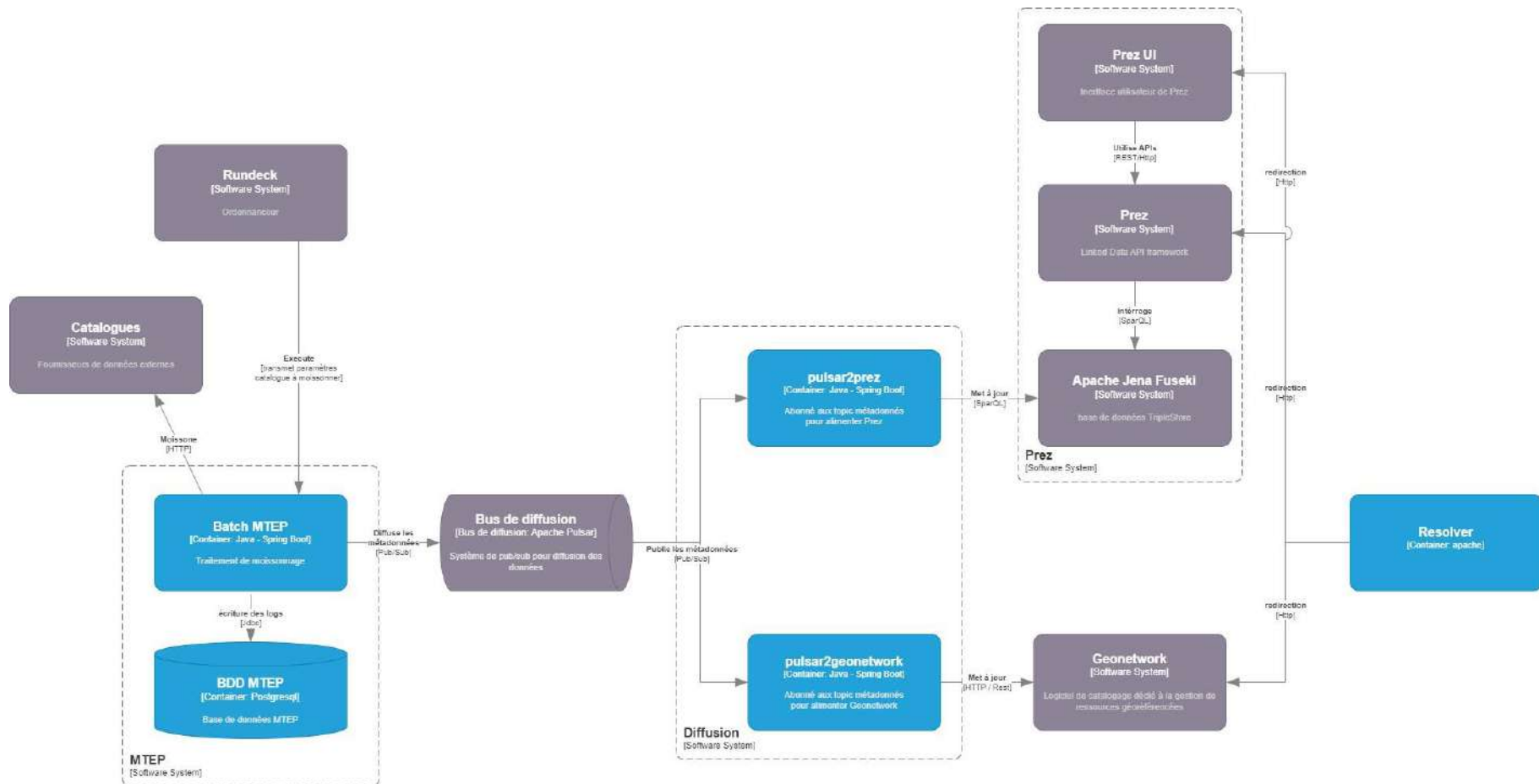
- Le moissonnage des catalogues des pôles de données
- L'implémentation des mappings réalisés avec les pôles
- La restitution des informations sous forme lisible pour un humain ET pour des machines

Schéma d'archi des composants mis en place

(MTEP : Moissonneur – Transformateur – Enrichissement – Persistance)



MTEP Data Terra – Schéma d'architecture



Les formations prévues pour l'aide à la FAIRISATION

Ateliers FAIR (Gaia Data)



- Utilisation de ces standards sont préconisés pour se diriger vers les standards « FAIR » (alignement simple avec le MTEP):
- Le W3C et le web sémantique
 - OGC/ISO Observations, measurements and sampling
 - OGC/WMO-OMM WaterML2
 - OGC GeoSciML
 - EarthResource ML
 - WMS / OGC API Maps (Geoserver, PyGeoAPI, Examind)
 - WFS / OGC API Features (Geoserver, PyGeoAPI, Examind)
 - WPS / OGC API Process (Geoserver, 52N, Examind, PyWPS)
 - CSW / OGC API Records (Geonetwork)
 - SensorThings API (FROST, Examind)
 - WCS / OGC API Coverage
 - THREDDS / ERDAPP
 - **STAC**
 - EDR



Les formats:

<https://www.odatis-ocean.fr/donnees-et-services/principes-de-gestion-des-donnees/formats-attributs-conventions>

- Le NetCDF4 est un format de fichier fréquemment rencontré dans le domaine scientifique et permet d'accéder à de nombreuses ressources informatiques
- Le TSV
- Les formats type ARCO - Analysis Ready Cloud Optimized : ZARR, COG, etc.

Les Vocabulaires et Référentiels:

<https://www.odatis-ocean.fr/donnees-et-services/principes-de-gestion-des-donnees/referentiels-et-vocabulaires>

Déploiement du SSO Data terra (lorsqu'il sera validé par la cybersécurité)



Quelques conseils et informations complémentaires!



Ecrire un plan de gestion de donnée permet de se poser les bonnes questions!

- Permet de planifier l'arrivée de nouvelle donnée
- De structurer une offre de service
- D'avoir une visibilité à long terme
- Rassurer le producteur de donnée
- D'assurer la protection de la donnée (propriété intellectuelle)
- D'avoir un retour sur l'usage de la donnée

- ✓ Data Terra : Centre de de référence thématique de Recherche Data Gouv pour les données du Système Terre: <https://recherche.data.gouv.fr/fr/page/centres-de-reference-thematiques-expertises-par-domaine-scientifique>
 - ✓ Recensement des services accompagnant la rédaction des PGD au sein des établissements d'enseignement supérieur et de la recherche: <https://scienceouverte.couperin.org/sos-pgd/>
 - ✓ Rédaction de votre PGD (fortement conseillé): <https://dmp.opidor.fr/>
- Utilisation du MADMP (PDG machine actionnable - <https://dmp.opidor.fr/help#content>)
- ✓ Aide à l'amélioration du contenu de votre PGD : [DoRANum – Grille de relecture de PGD](#)
 - ✓ Ressources pédagogiques DoRANum: <https://doranum.fr>





Merci de votre attention !