Entrepôt de données IRD: un service en ligne pour l'ouverture et le partage des données scientifiques au Sud

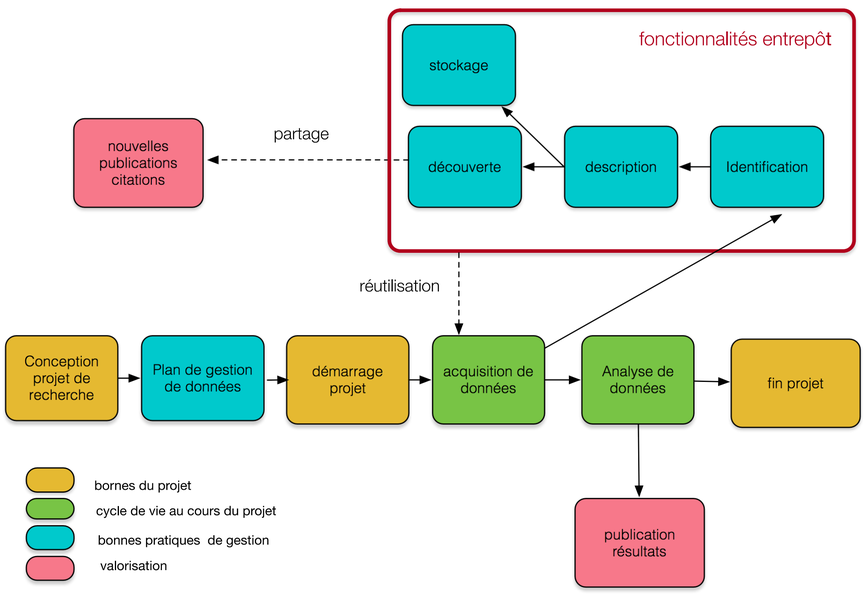
Desconnets J-C. [(1)], **Aventurier P.** [(2)], **Banon S.**[(3)], **Doucouré C.**[(2)], **Coupin T.**[(4)], **Hensens H.**[(2)], **Soumaré A.** [(1)]
(1) Mission Infrastructures et données Numériques – IRD,
(2) Mission Culture Scientifique et Technique/Service Information Scientifique et Technique – IRD,
(3) Direction pour le Développement des Usages Numériques et Innovants/Service CRM – IRD
(4) UMR Gouvernance, Risques, Environnement, Développement (GRED)

Un enjeu pour le partage des résultats de la recherche avec le Sud

Pour faire face à :

- **des défis scientifiques de plus en plus complexes** : changement climatique, santé, migrations,…
- **de forts enjeux dans les zones intertropicales :** croissance démographique, sécurité alimentaire, pression accrue sur les ressources,…
- **la nécessité de l'interdisciplinarité**, de **l'ouverture** et de la **mutualisation** des données et des compétences.

Changer les pratiques de gestion des données pour mieux maîtriser leur cycle de vie



La mise à disposition d'un service assurant **l'identification**, le **stockage** et la **description** selon les **principes FAIR** au sein d'un entrepôt permet d'introduire de nouvelles pratiques dans la diffusion et la **valorisation** des données au delà du projet et favorise ainsi leurs **réutilisation et citation** par d'autres équipes de recherche.

Pérenniser et valoriser le patrimoine de données de l'IRD et de ses partenaires

Objectifs :

- **Assurer un partage ouvert, pérenne** des jeux de données d'intérêt récoltées dans les zones intertropicales par **l'IRD et ses partenaires**.
- **A court terme :** Fournir un service de **d'identification**, de **dépôts**, de **sauvegarde** et de **découverte** des données
- **A plus long terme :** Permettre la découverte des **ressources numériques produites par l'IRD** et déposées dans d'autres entrepôts et bases de données.

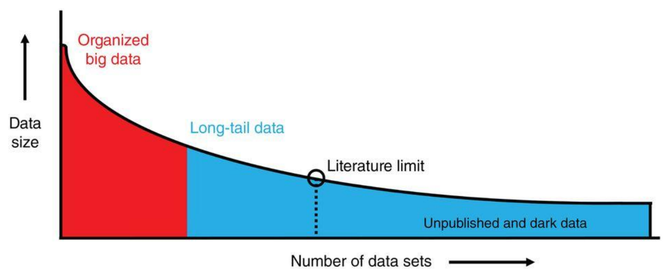
Prendre en compte la diversité des besoins

Des données venant de différents domaines avec des caractéristiques diverses (exemples)

- Génomique
- Ecosystèmes marins exploités
- Agro biodiversité
- Santé
- Géosciences
- Sciences sociales

Des attentes très variées

- Données en danger
- Reproduction d'expériences
- Demande de DOI
- Création de data paper
- Partage des données avec d'autres équipes



Distribution des données de la recherche
(Ferguson et al.,2014 :https://www.nature.com/articles/nn.3838)

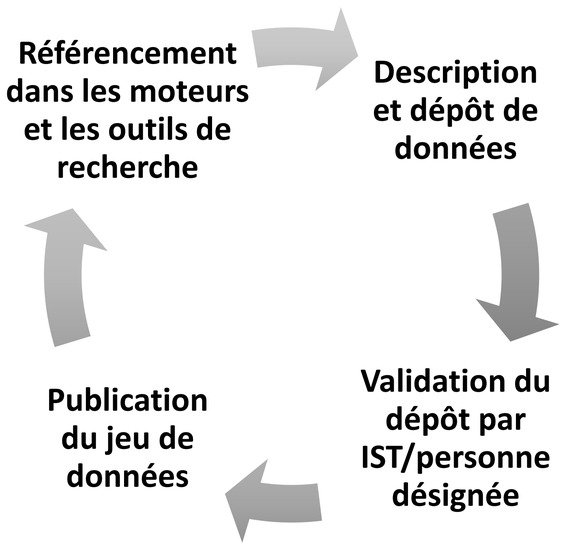
Les **données ciblées** sont celles de la **longue queue des données** (*long tail of data*) : des données qui ne sont actuellement pas mises à disposition au sein de portails de données adossés aux observatoires ou pôles de données.

Principales exigences fonctionnelles

- Attribution de DOI
- Gestion des dépôts par le chercheur
- Lien sécurisé pour la relecture
  Historique des versions
- Métriques
- Moissonnage par les moteurs de recherche

Le chercheur dépose et décrit ses données

- **Création d'un espace personnalisable et formation d'une personne référente** pour chaque unité et projets qui le désire



Un administrateur des données l'accompagne

Basé sur un outil open source et largement répandu

The Dataverse® Project

Ouverture prévue au printemps 2019

Contacts :
Jean-Christophe.Desconnets@ird.fr, Pascal.Aventurier@ird.fr

Premières Journées Nationales de la Science Ouverte: de la stratégie à l'action. Paris, 4-6 Décembre 2018