La gestion des données et l'accompagnement à l'IRD

5 novembre 2020

Pascal Aventurier – Responsable du Service IST
Mission Culture Scientifique et Technologique
@Pascal_AV
pascal.aventurier@ird.fr

Institut de Recherche pour le Développement FRANCE

l'IRD défend un modèle original de partenariat scientifique équitable avec les pays du Sud et une science interdisciplinaire et citoyenne, engagée pour la réalisation des Objectifs de Développement Durable.

75 UNITÉS DE RECHERCHE

62 % DE CO-PUBLICATION AVEC UN PARTENAIRE SUD

5 départements scientifiques

- DYNAMIQUES INTERNES ET DE SURFACE DES CONTINENTS (DISCO)
- ECOLOGIE, BIODIVERSITÉ ET FONCTIONNEMEN DES ECOSYSTÈMES CONTINENTAUX (ECOBIO)
- OCÉANS, CLIMAT ET RESSOURCES (OCEANS)
- SANTÉ ET SOCIÉTÉS (SAS)
- SOCIÉTÉS ET MONDIALISATION (SOC)

2050 agents travaillent à l'IRD

852 chercheurs

30% des agents sont hors métropole

Les données de la recherche à l'IRD

2015 - GT institutionnel observatoires de données / Métadonnées / DOI

2016 - Formation sensibilisation aux données et DMP avec OPEN ACCESS LANGUEDOC-ROUSSILLON

MIDN + IST

2018 - Mission Infrastructure et données numériques (**MIDN**) + IST + D-DUNI informatique : lancement de l'entrepôt de données DataSuds

2018-2019 – Datathon 5 ateliers : les chercheurs viennent avec leurs données

Sept 2019 - Entrepôt DataSuds et data.ird.fr + recrutement d'un administrateur (poste interne)

- Séminaire de lancement de DataSuds (+ 100 participants)

Depuis octobre 2019 - Formations à la carte dans les unités

2019 - Création d'un réseau de référents dans chaque unité

2020 - Lancement du groupe DMP (Plan de Gestion de Données)

- Formation labellisée par le service Formation RH . Projet e-learning
- La MIDN devient **Mission pour la Science Ouverte**

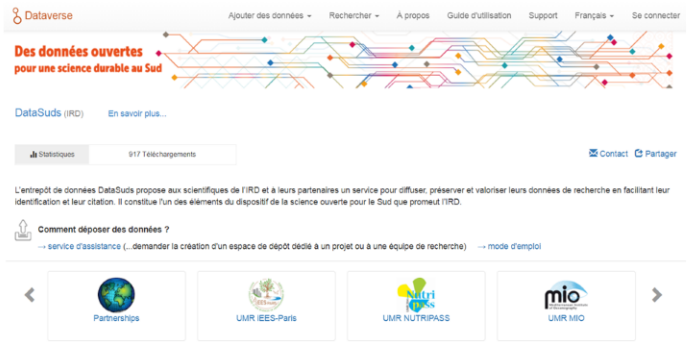
Données de recherche et services de documentation : où en est-on ?

5 nov. 2020 Conférence en ligne (France)

couperin.org

L'entrepôt DataSuds

Resp : Luc Decker



- Données en danger
- Données de longue traîne (lien)
- Données de l'IRD (UMR) et partenaires Sud
- Le choix de l'entrepôt de dépôt n'est pas imposé
- Fonction annuaire en projet : moissonnage des métadonnées d'autres entrepôts (sans copie des données)

Principe du dépôt unique

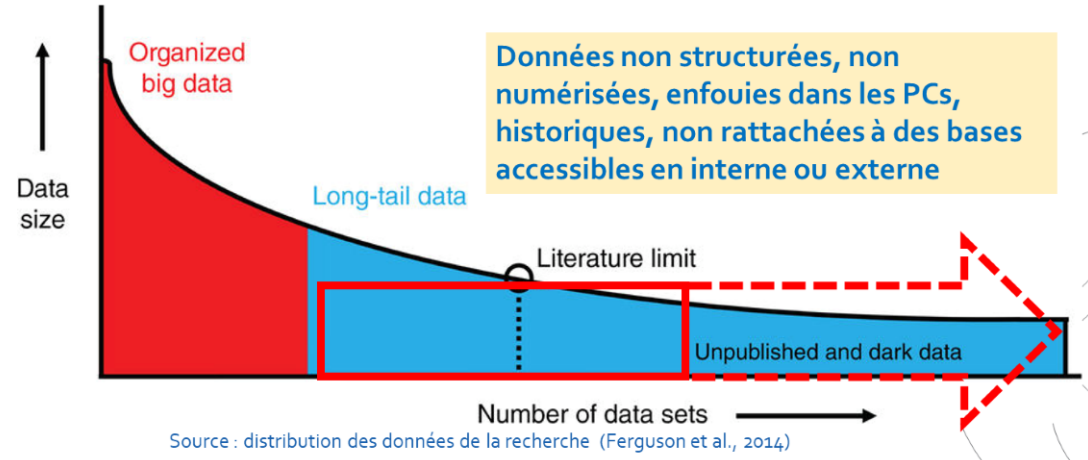
Données de recherche et services de documentation : où en est-on ?
5 nov. 2020 Conférence en ligne (France)

couperin.org

https://dataverse.ird.fr/
http://e-irg.eu/documents/10920/238968/LongTailOfData2016.pdf
Application *Dataverse*
Données en danger, données de longue traîne (lien)
Ouvert aux données des structures dont l'IRD est cotutelle (UMR, US, UMI, LMI) et des partenaires de l'IRD au Sud
Le choix d'un l'entrepôt de dépôt n'est pas imposé
Fonction annuaire en projet : moissonnage des métadonnées d'autres entrepôts (sans copie des données)

Positionnement de l'entrepôt DataSuds
Données cibles

Observatoires, pôles de données, bases de données accessibles sur le web



Données non structurées, non numérisées, enfouies dans les PCs, historiques, non rattachées à des bases accessibles en interne ou externe

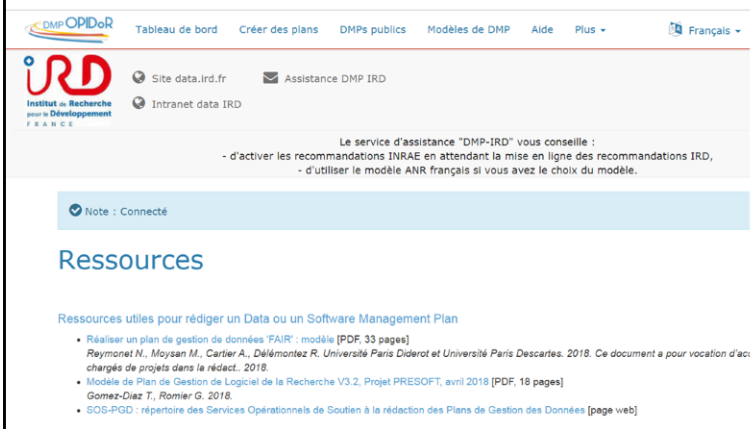
Source : distribution des données de la recherche (Ferguson et al., 2014)

L'accompagnement / Formation

Resp. Hanka Hensens

- Pour informer sur les bonnes pratiques
- S'appuyer sur l'existant (DORANUM , FOSTER, INRAE, CIRAD…)
- Constats
  - peu de connaissances de la part des chercheurs
  - **Manque de temps** pour la mise en forme des données
  - Confusion entre données pendant le projet et l'archivage dans un entrepôt
  - Choix de l'entrepôt

Plan de gestion de données



Resp : Caroline Doucouré

- Outil DMP Opidor
- Obligatoire pour chaque projet financé
- groupe multi-compétences
  - Services Juridiques
  - Valorisation
  - IST
  - Informatique
- Services
  - Rédaction de recommandations en cours
  - « relecture » des plans de gestion de données

Données de recherche et services de documentation : où en est-on ?
5 nov. 2020 Conférence en ligne (France)

couperin.org

Obligatoire pour chaque projet financé : exigences des bailleurs de fonds , voir aussi le Plan National de la Science Ouverte

Quelles compétences ?

4 personnes « Données » (2 ETP) dans le service IST de l'IRD de 13 personnes :

- 1 administrateur de l'entrepôt (BAP E), 1 chef de projet (BAP E)
- 2 documentalistes

au sein du groupe de travail IRD aux compétences complémentaires (informatique, juridique, valorisation..)

Compétences de base :

Communication scientifique / Science ouverte
Gestion de projet
Métadonnées et rôle des standards

À acquérir

Principes FAIR
Connaissances juridiques
Connaissances techniques
Plan de gestion de données

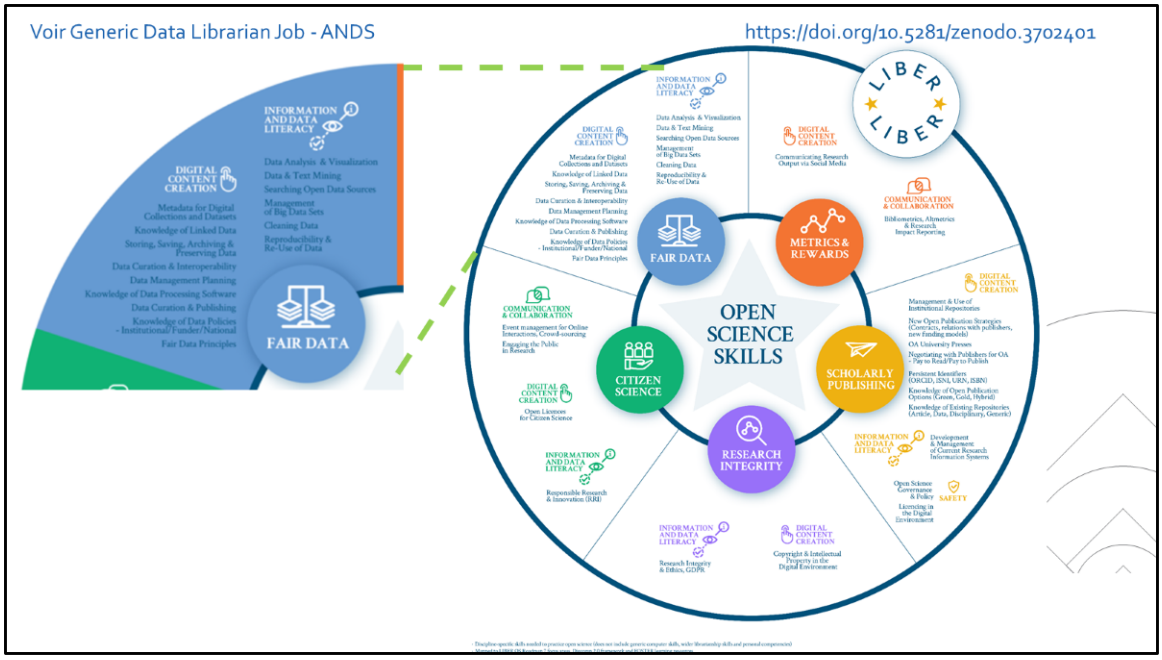
GTSO -Couperin

Depuis 2020 : DU Scientific Data Management à l'Univ. de Montpellier

Données de recherche et services de documentation : où en est-on ?

5 nov. 2020 Conférence en ligne (France)

https://www.fosteropenscience.eu/node/2091
https://doranum.fr/
https://www.usine-digitale.fr/article/a-montpellier-un-diplome-d-universite-creuse-la-gestion-des-donnees-de-la-science.N997229
https://www.ands.org.au/working-with-data/data-management/overview/data-management-skills/information-specialists-and-data-librarian-skills
https://www.ands.org.au/__data/assets/pdf_file/0006/406437/generic_datalibrarian_jobdescription.pdf



McCaffrey, Ciara, Thorsten Meyer, Clara Riera Quintero, Cecile Swiatek, Nathalie Marcerou-Ramel, Camilla Gillén, Karin Clavel, et al. 2020. Open Science Skills Visualisation. Zenodo. https://doi.org/10.5281/zenodo.3702401.

Le contexte en France

- Les projets sont souvent en multi-partenariats ou dans le cadre d'une unité Mixte de recherche (plusieurs instituts tutelles) → où déposer les données ?
- Des infrastructures de recherche qui peuvent gérer des données
- Etude en cours Service générique d'accueil et de diffusion de données simples (à lire sur ouvrir la science)
- Projet ANR BRIDGE : harmoniser les pratiques entre 3 partenaires (IRD, CIRAD et INRAE)

Principe du dépôt unique

Données de recherche et services de documentation : où en est-on ?
5 nov. 2020 Conférence en ligne (France)

Voir ouvrir la science : https://www.ouvrirlascience.fr/service-generique-daccueil-et-de-diffusion-de-donnees-simples/
https://ist.blogs.inrae.fr/bridge-science-ouverte/
https://www.enseignementsup-recherche.gouv.fr/pid25384/strategie-nationale-des-infrastructures-de-recherche.html

Les principes définissant une grande infrastructure de recherche peuvent s'énoncer ainsi :
elle doit être un outil ou un dispositif possédant des caractéristiques uniques identifiées par la communauté scientifique utilisatrice comme requises pour la conduite d'activités de recherche de haut niveau. Les communautés scientifiques visées peuvent être nationales, européennes ou internationales, selon le cas ;
elle doit disposer d'une gouvernance identifiée, unifiée et effective, et d'instancesde pilotage stratégique et scientifique ;
elle doit être ouverte à toute communauté de recherche souhaitant l'utiliser, accessible sur la base de l'excellence scientifique évaluée par les pairs ; elle doit donc disposer d'instances d'évaluation adéquates ;
elle peut conduire une recherche propre, et/ou fournir des services à une (ou plusieurs) communauté(s) d'utilisateurs intégrant les acteurs du secteur

économique.Ces communautés peuvent être présentes sur le site, venir ponctuellement y réaliser des travaux, ou interagir à distance.

Conclusion

Permet de redynamiser d'autres services IST + science ouverte
Ne pas rater le train en marche.

- Coordination entre différents services
- Légitimité des services IST pour développer des services autour des données
- À l'IRD, projet DR s'inscrit dans la dynamique d'établissement (Création de la Mission Science Ouverte)
- Implication dans des projets scientifiques
- En cours : charte de la Science Ouverte