

# Les entrepôt de données

Ou comment rendre les données  
trouvables, accessibles et  
réutilisables ?

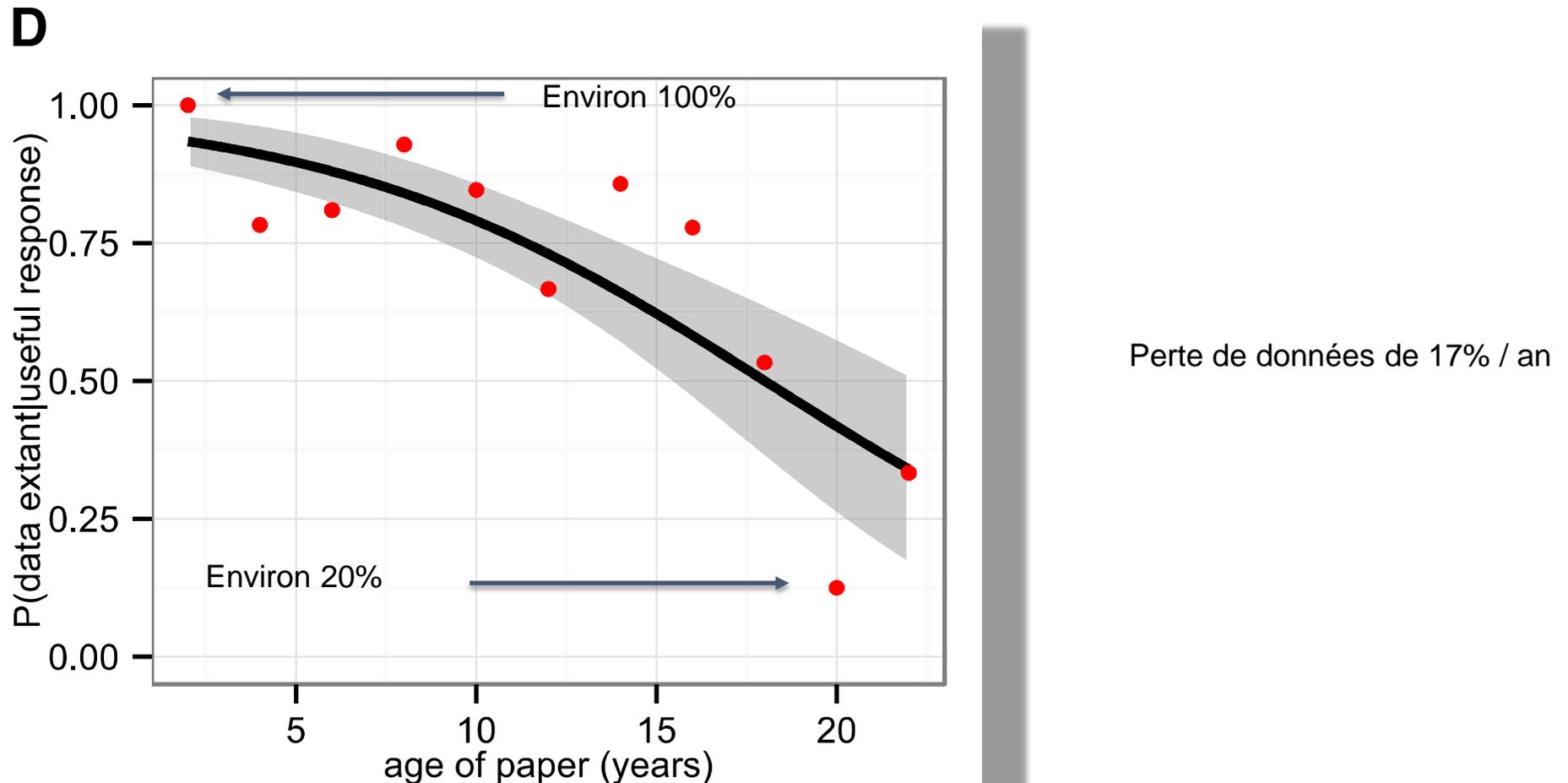
Jean-Christophe Desconnets – Géomaticien  
IRD ESPACE-DEV  
ORCID : <https://orcid.org/0000-0002-4142-5289>



# Plan

- **Introduction**
- **Partie 1 : les données de la recherche en environnement**
  - Quelques définitions autour des données
  - Caractéristiques des données en environnement
- **Partie 2 : notions et fonctionnalités des entrepôts de données**
  - Notions d'entrepôts de données
  - Fonctionnalités
  - Outils
  - Exemples
- **Partie 3 : Discussions**
  - Valeurs ajoutées/Limites des démarches actuelles
  - Complémentarité entrepôts vs dispositif de diffusion
  - Mutualisation des efforts

# Importance de mieux préserver les données et de les rendre accessibles



D) Predicted probability that the data were extant (either “shared” or “exist but unwilling to share”) given that we received a useful response. In all panels, the line indicates the predicted probability from the logistic regression, the gray area shows the 95% CI of this estimate, and the red dots indicate the actual proportions from the data. <http://dx.doi.org/10.1016/j.cub.2013.11.014>

# Exemple de constats dans le domaine océanographique\*

- **Effort d'observation important**

- Une observation du milieu marin très distribuée, malgré un effort de structuration : IR, SO et SNO, ...
- Implication de nombreux laboratoires répartis géographiquement et thématiquement
- Des situations (degrés de maturité) différents

- **Beaucoup de perte en ligne**

- Exemple des campagnes à la mer (étude Européenne) : environ 70% des résultats non accessibles, voire perdus après 10 ans : changement de projet, d'affectation, départ (retraite...)

\* tiré de présentation ODATIS, 14/06/2018, G. Maudire, IFREMER)

# PLAN NATIONAL POUR LA SCIENCE OUVERTE

MERCREDI 4 JUILLET 2018

## DEUXIÈME AXE : STRUCTURER ET OUVRIR LES DONNÉES DE LA RECHERCHE

Les chercheurs seront invités à déposer les données dans des entrepôts de données certifiés, dont la gouvernance et les règles de propriété intellectuelle seront conformes aux bonnes pratiques. À ce titre, les infrastructures nationales et européennes de recherche seront privilégiées, notamment via des centres de données thématiques et disciplinaires. Les plans de gestion des données, instrument de définition des règles de construction, conservation et diffusion des données, seront généralisés. Un prix des données de la recherche sera mis en place afin de mettre en valeur et récompenser les équipes qui réalisent un travail exemplaire dans ce domaine.

#scienceouverte

 esr.gouv.fr





## Partie 1 : les données de la recherche en environnement

ou quelles données cibles pour les entrepôts de données ?

## Quelques définitions autour des données

«**Données de la recherche** [...] enregistrements factuels (chiffres, textes, images et sons), [...] sources principales pour la recherche scientifique [...] reconnus par la communauté scientifique comme nécessaires pour valider des résultats de recherche. [...] » (OCDE, 2007)

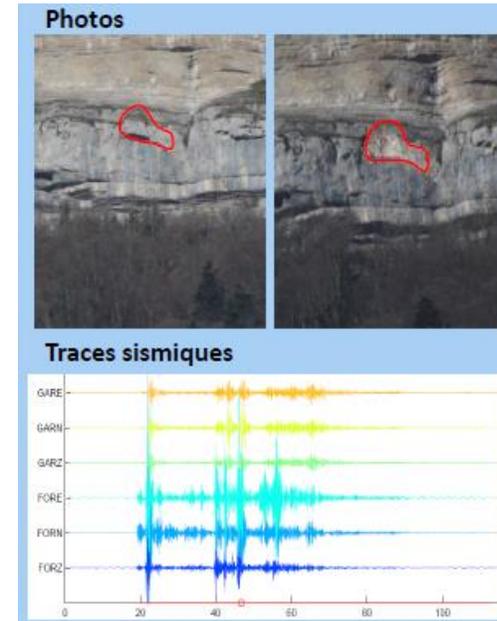


**Données d'intérêt, réutilisable** pour enrichir, compléter, étendre d'autres jeux de données en vue d'améliorer la connaissance de l'objet d'étude...



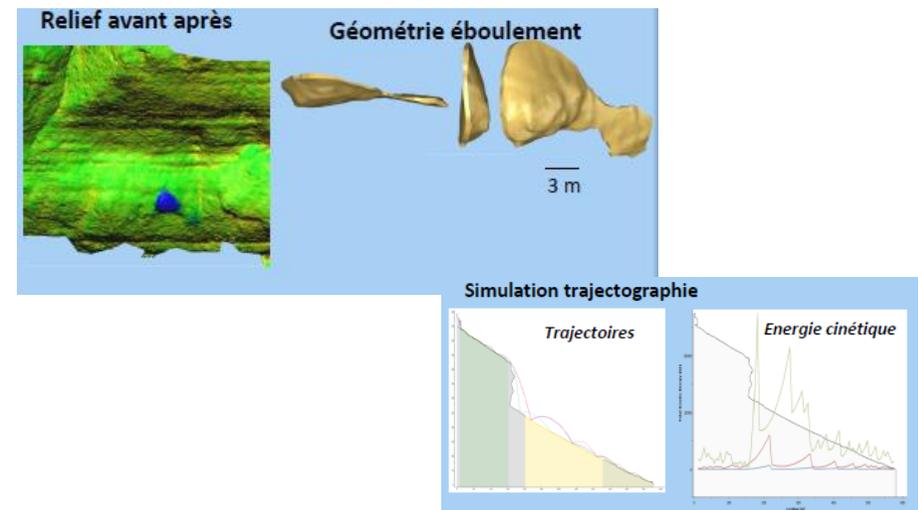
# Quelques définitions autour des données : Données brutes - données élaborées

- « **Données brutes** : recueillies sur un sujet à partir d'observations ou de mesures, et qui n'ont pas encore été traitées, validées ou triées. Sensées être neutres et objectives et ne dépendant pas de leur contexte de création, d'une analyse ou de leur producteur » (*Thessen & Patterson, 2011*)



(Amitrano, Lejeune, & Hobléa, 2017)

- **Données élaborées ou dérivées** : ayant subi des opérations de nettoyage, de sélection, de traitement, d'analyse afin de produire des résultats.



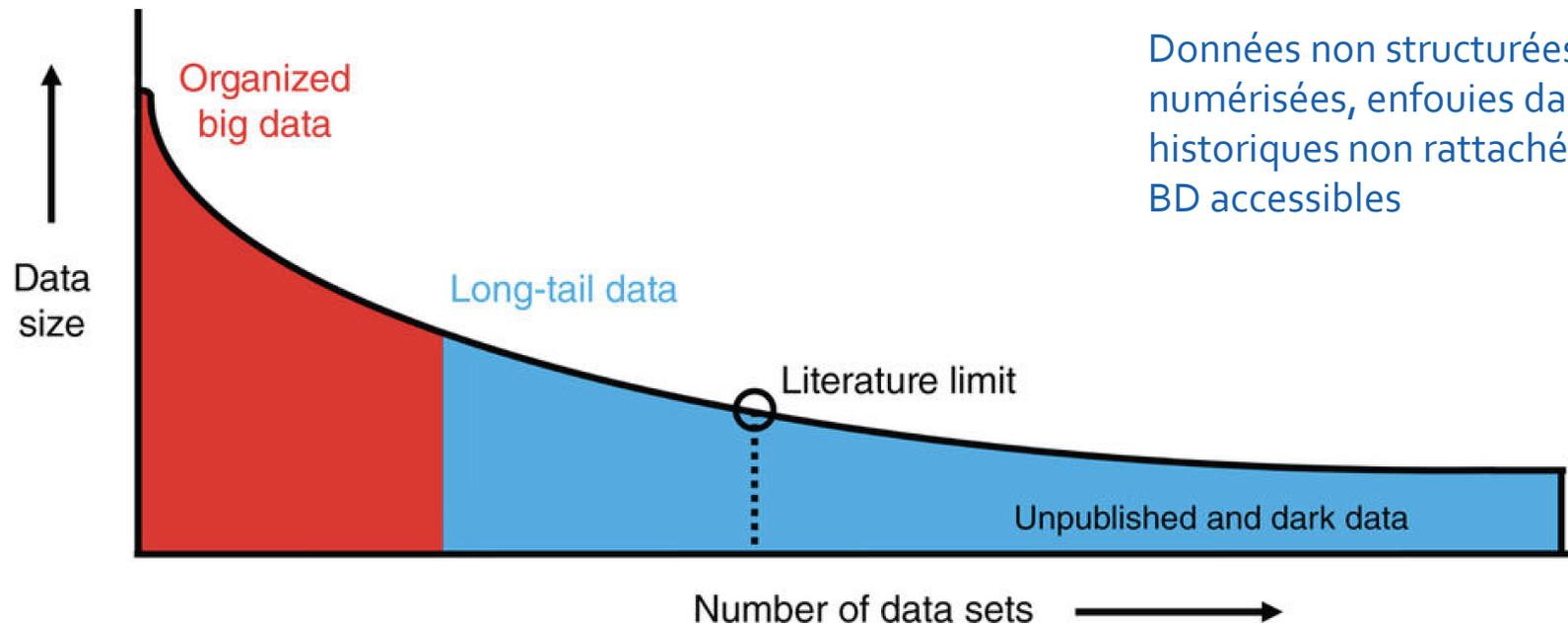
# Big data vs Long tail of data

## Big data

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

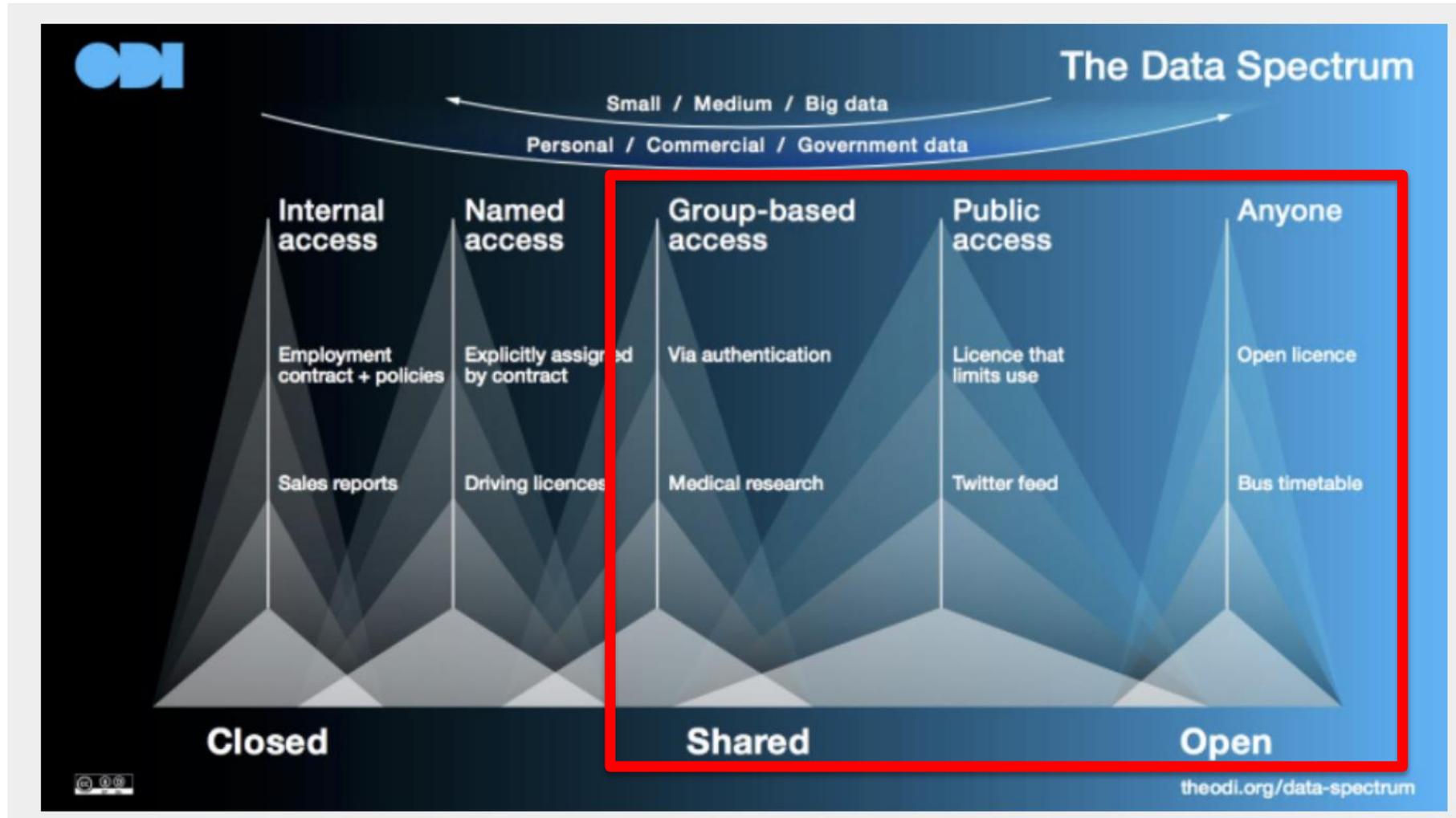
## La longue queue des données

Données non structurées, non numérisées, enfouies dans les PC, historiques non rattachées à des BD accessibles



Distribution des données de la recherche  
(Ferguson et al., 2014)

# Données fermées/données partagées/données ouvertes



(d'après l'Open data institute : <https://theodi.org/about-the-odi/the-data-spectrum/> )

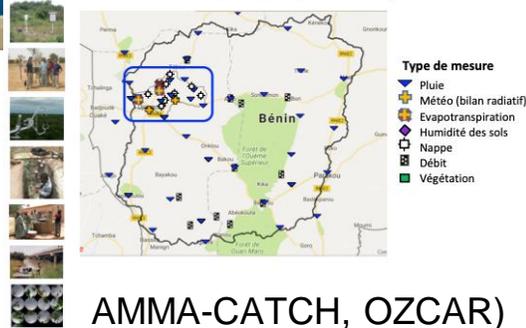
# Les données en environnement

- Spécificités : « données dynamiques », « longues séries »
  - ◆ issues de dispositifs d'observation de nature diverses qui **produisent en permanence** et sur le long terme



SISMER, ODATIS)

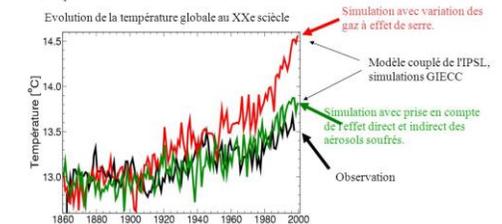
100 stations de mesure (565 capteurs) : 1300 mesures/heure



SPOT 6

## Modélisation du climat et changement climatique (Pôle modélisation de l'IPSL)

Evolution de la composition de l'atmosphère, source d'incertitude sur le changement climatique

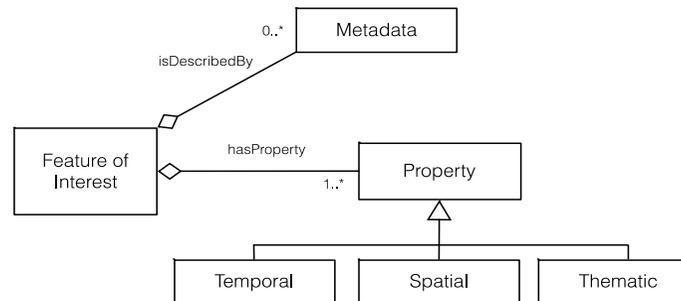


ations réalisées par Sébastien Denvil et le pôle modélisation de l'IPSL)  
Important aussi pour rétroaction vapeur d'eau, nuages, etc ...

- ◆ Nécessité d'avoir de **longues séries de données** pour étudier les phénomènes sur de longues périodes

# Les données en environnement

- Données spatio-temporelles
  - ◆ Nature et représentation de la donnée : Temporalité et spatialité

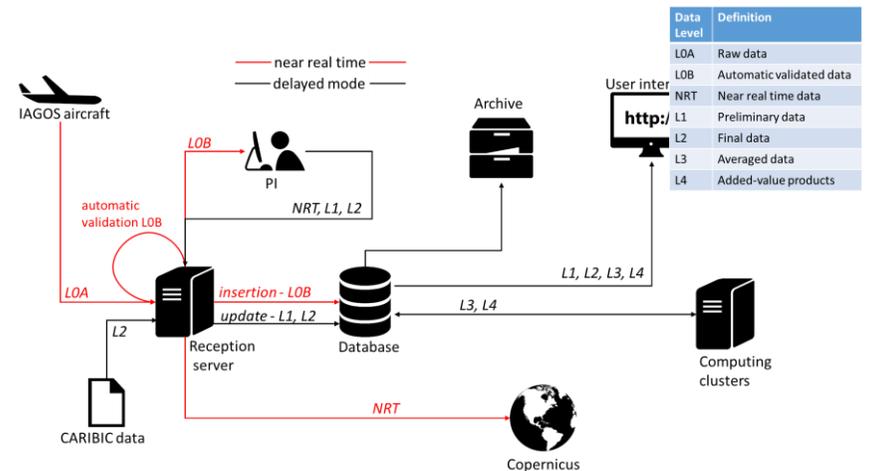
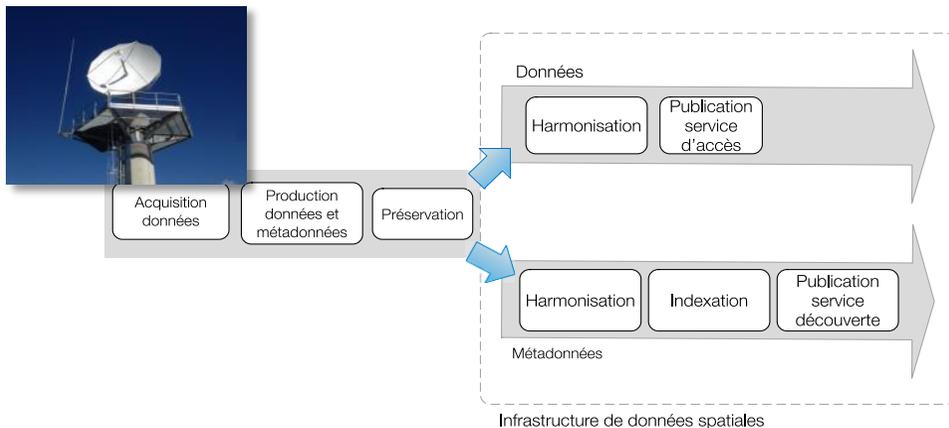


→ induit le besoin d'outils et de services spécifiques pour valoriser, réutiliser les données (visualisation, analyse, traitements...)

- ◆ Dans la gestion des données
  - **Maturité** des outils, méthodologies et standards
  - Formats de données, standards de métadonnées, services sur les données
  - Cadre d'application : Directives INSPIRE
  - **Nombreuses implémentations existantes:** plateformes d'accès, de traitements

# Les données en environnement : différences de production des données

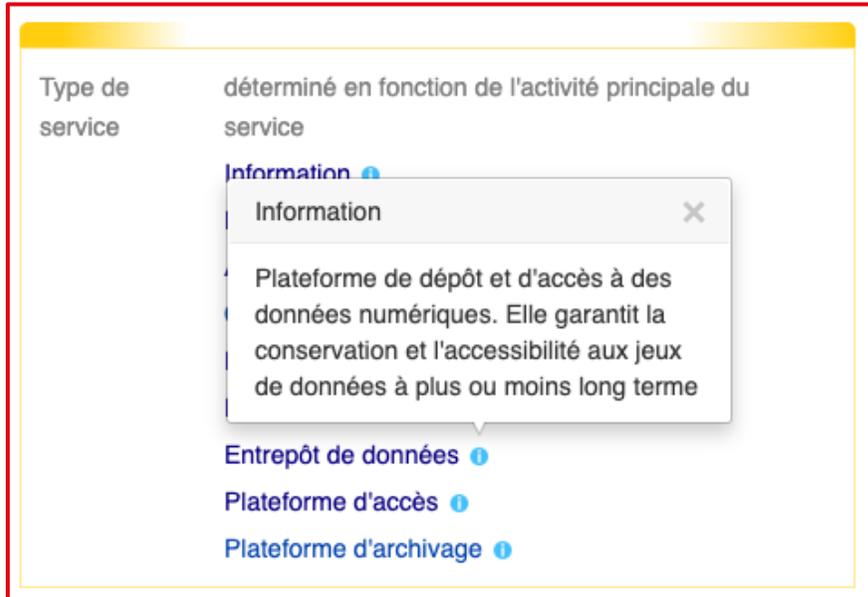
- Différents modes d'acquisition induisent divers degrés d'hétérogénéités sur :
  - ◆ organisation tout au long du cycle de vie
  - ◆ les règles d'ouverture, les conditions d'utilisation
  - ◆ chaîne de production « pipeline » des données :
    - manuelle
    - Formalisée, automatisée, voire industrialisée





## Partie 2 : notions et fonctionnalités des entrepôts de données

# Entrepôt de données : définition



Type de service déterminé en fonction de l'activité principale du service

Information ⓘ

Information

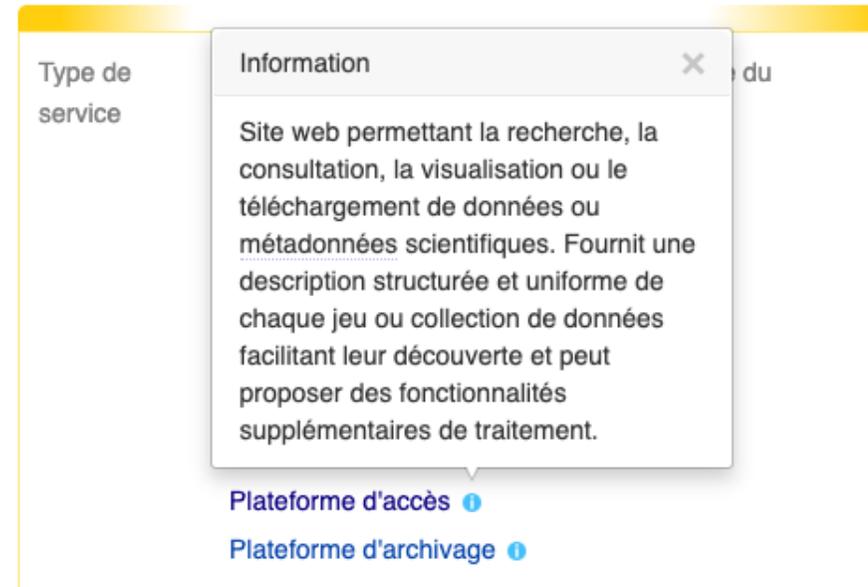
Plateforme de dépôt et d'accès à des données numériques. Elle garantit la conservation et l'accessibilité aux jeux de données à plus ou moins long terme

Entrepôt de données ⓘ

Plateforme d'accès ⓘ

Plateforme d'archivage ⓘ

≠



Type de service

Information ⓘ

Site web permettant la recherche, la consultation, la visualisation ou le téléchargement de données ou métadonnées scientifiques. Fournit une description structurée et uniforme de chaque jeu ou collection de données facilitant leur découverte et peut proposer des fonctionnalités supplémentaires de traitement.

Plateforme d'accès ⓘ

Plateforme d'archivage ⓘ

≠

**Entrepôt de données** ou EDD (ou **base de données décisionnelle** ; en anglais, **data warehouse** ou DWH) désigne une base de données utilisée pour collecter, ordonner, journaliser et stocker des informations provenant de base de données opérationnelles et fournir ainsi **un socle à l'aide à la décision** en entreprise.

*Wikipédia*

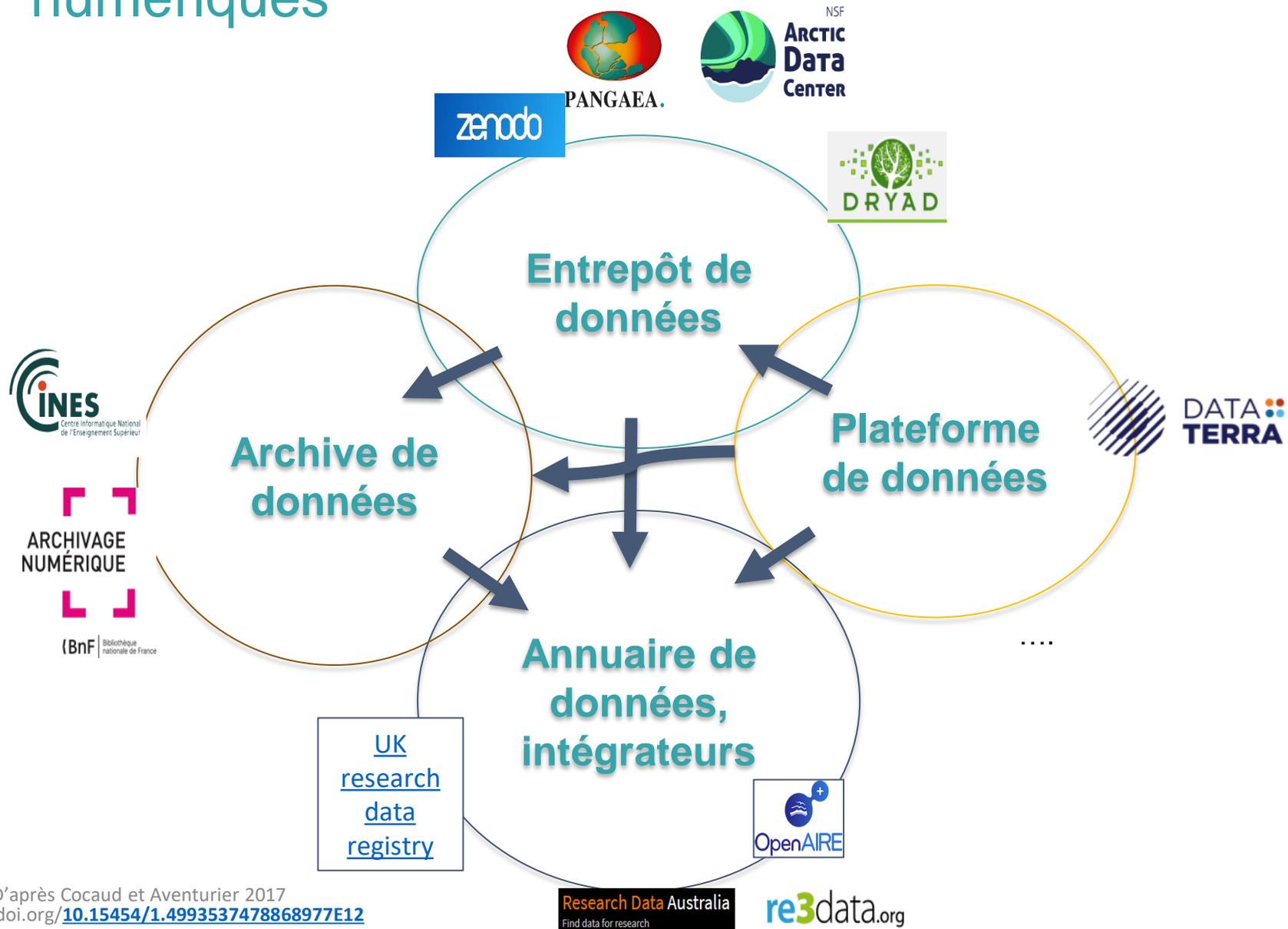
# Entrepôt de données : définition

Service en ligne permettant le dépôt, la description, la conservation, la recherche et la diffusion des jeux de données.



**Entrepôt disciplinaire / institutionnel / ouvert à toutes disciplines**

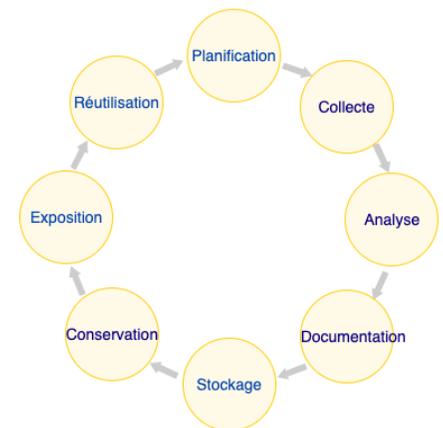
# Les entrepôts en lien avec les autres dispositifs numériques



# Entrepôt de données dans le cycle de vie de la donnée



Stades du cycle de vie dans lesquels un entrepôt de données intervient



Cycle de vie de la donnée (CatOpidor)

# Fonctionnalités des entrepôts : dépôt et conservation des données

- Téléchargement / import de fichiers :
  - ◆ données + documents complémentaires
  - ◆ Limites / taille des fichiers, formats acceptés, données publiées...
  - ◆ Via IU ou API
- Organisation des données en collections

- Conservation (stockage sécurisé + archivage à long terme)

## Intégrité

- sauvegarde
- empreinte numérique

## Lisibilité

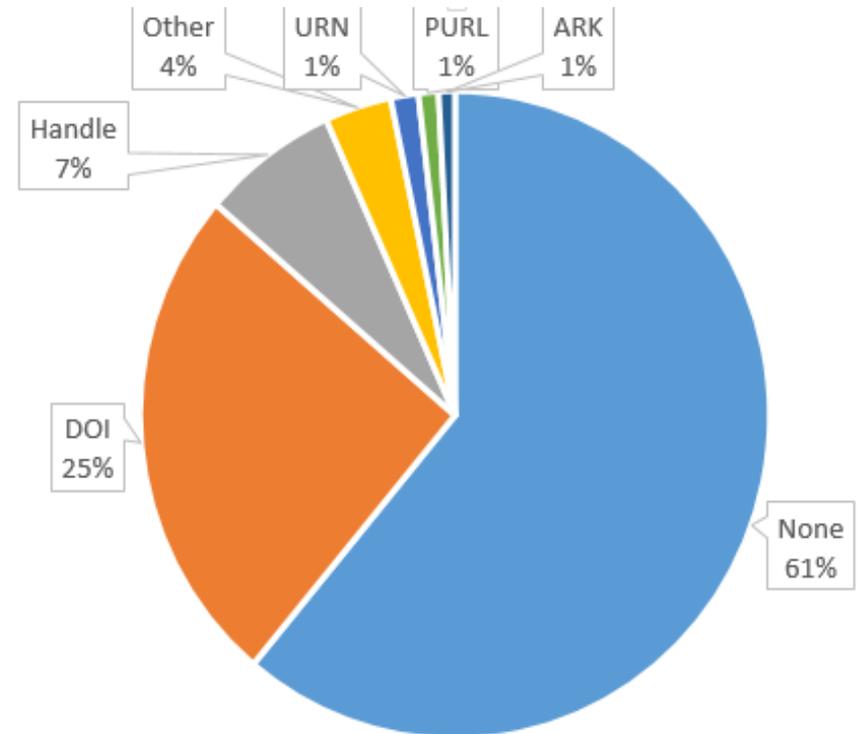
- veille sur l'évolution des technologies,
- utilisation de formats durables (ou conversion), standards

## Intelligibilité

- métadonnées riches, standards
- identification pérenne
- contrôle des versions de jeux de données.

# Fonctionnalités des entrepôts : identification pérenne des données

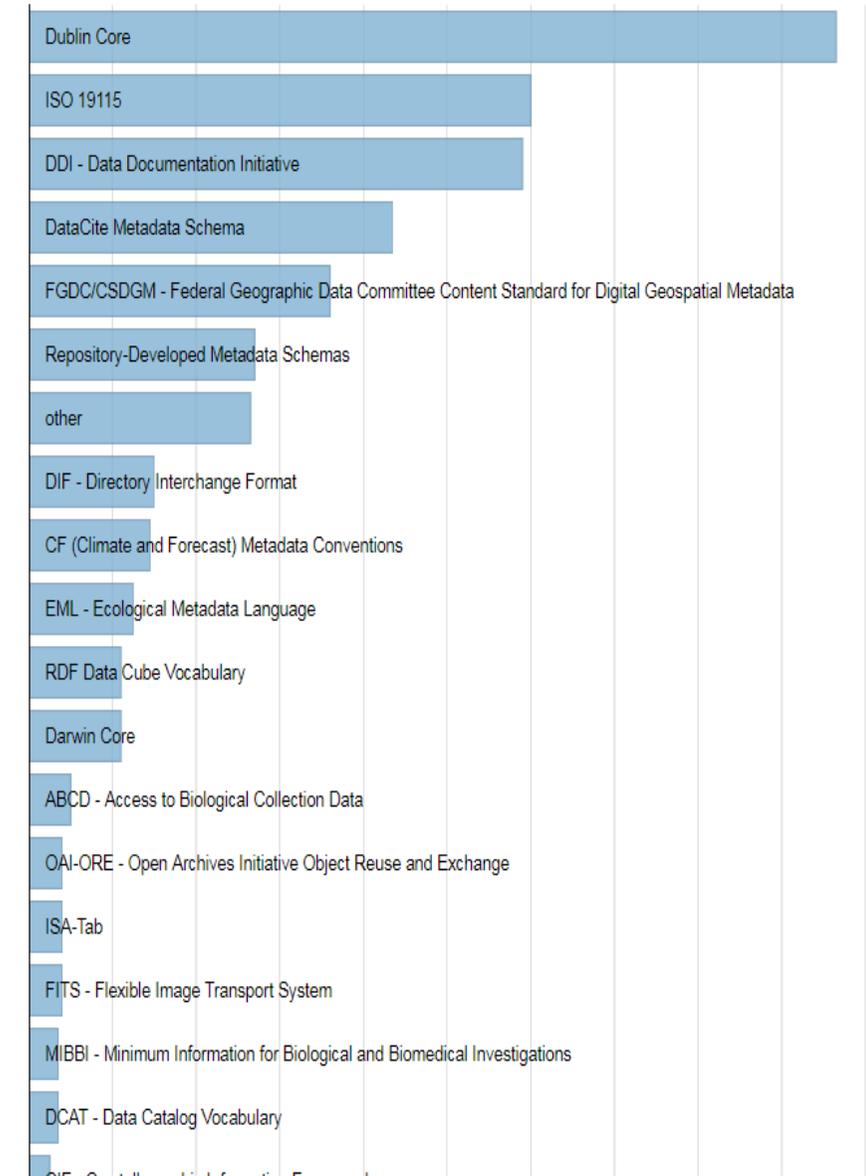
- Attribution d'un identifiant au moment du dépôt
  - ◆ Au jeu dans son ensemble,
  - ◆ +/- à chaque version du jeu de données
- Intégration possible d'identifiant préexistant chez certains entrepôts (ex : Zenodo)



Données extraites de re3data.org le 19/05/2017 : total = 2042 entrepôts

# Fonctionnalités des entrepôts : description des données

- **Métadonnées**
  - ◆ Schémas de métadonnées
    - génériques
    - spécifiques à un domaine
  - ◆ Qualité des métadonnées
    - vocabulaires contrôlés (import ou accès à des vocabulaires externes)
  - ◆ Saisie, import via IU vs API
- **Documentation complémentaire**
- **Possibilité de listes contrôlées**



# Fonctionnalités des entrepôts : contrôle des droits d'accès aux données, conditions d'utilisation et licences

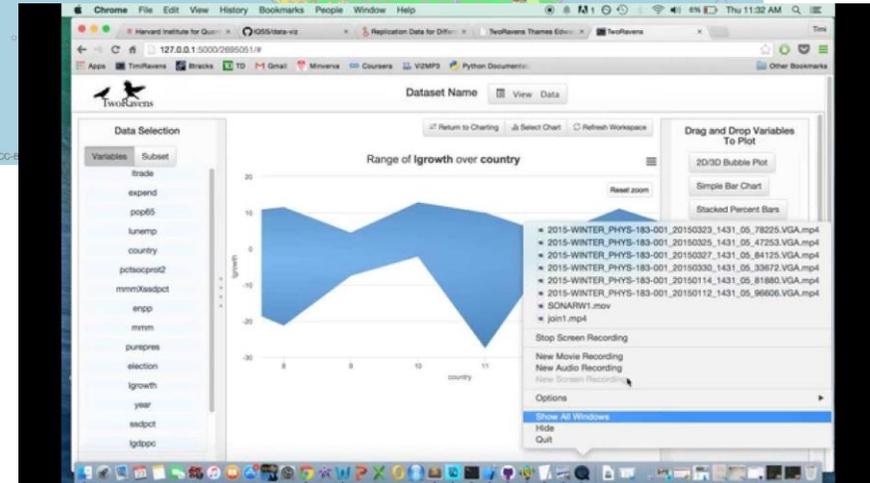
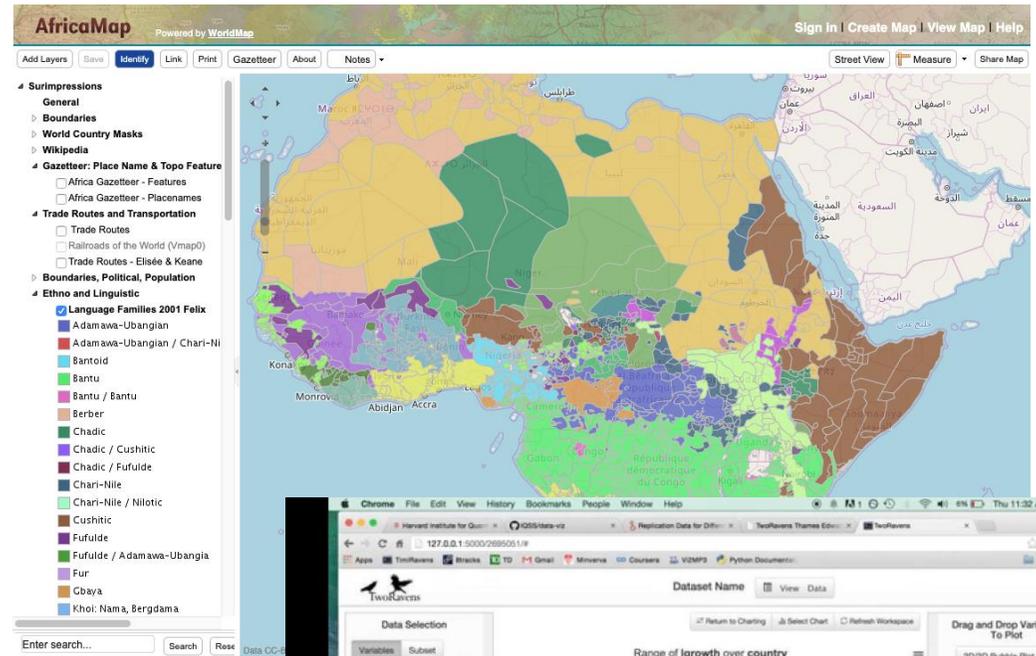
- Droit d'accès aux données :
  - ◆ téléchargement libre (open)
  - ◆ Embargo
  - ◆ demande d'accès (restricted), guestbook
  - ◆ pas de téléchargement possible (closed)
- Génération d'URL privé parfois proposé
- Attribution d'une licence à chaque jeu de données
  - ◆ Saisie libre vs liste fermée
- Attention aux entrepôts qui imposent une licence unique

# Fonctionnalités des entrepôts : recherche, affichage, export des (méta)données

- Recherche
  - ◆ simple et/ou avancée : dans les métadonnées, parfois dans les données,
  - ◆ Graphique (zone géographique, structure molécule...)
  - ◆ Navigation et affinement par facettes
- Affichage
  - ◆ des métadonnées
    - Liens avec d'autres datasets, avec des publications
    - Génération de la citation
  - ◆ des données (prévisualisation, selon format)
- Exports
  - ◆ métadonnées (≠ formats),
  - ◆ téléchargement des données

# Fonctionnalités des entrepôts : exploration et visualisation des données

- Outils d'analyse ou de visualisation
  - ◆ TwoRavens (Analyse de données tabulaires)
  - ◆ WorldMap (visualisation de données à référence spatiale)
  - ◆ ...



# Fonctionnalités des entrepôts : découverte, visibilité

- Exposition des métadonnées via **OAI-PMH** ou dans un triple store RDF (Linked data)
- API pour la recherche et l'accès (ex. SWORD sur dataverse)
- Liaison avec d'autres ressources :
  - ◆ Ex : IsCitedBy, Cites, ... qui sont les valeurs possibles de la propriété relationType de la métadonnée relatedIdentifier
- Statistiques d'usage de ses données
  - ◆ Nombre d'affichages et de téléchargements, datasets les + téléchargés, altmetrics, ...

# Intérêt des entrepôts pour la visibilité des données

- Les entrepôts sont scannés par des outils de recherche spécifiques

- ◆ [Data Cite search](#)

- ◆ [Data Citation Index](#) (Thomson Reuters)

- ◆ [Google Dataset Search](#)

- ◆ [Data Search](#) (Elsevier)



- ◆ et moissonnés par des catalogues, intégrateurs, infrastructures européennes de données... de plus en plus nombreux



- Peuvent diffuser leurs données via le protocole d'échange standard OAI-PMH

# Création ou utilisation d'un entrepôt de données (existant)



- Outils open source sur étagère

- ◆ Dataverse (<https://dataverse.org/>, Harvard University)
- ◆ Nesstar (<http://www.nesstar.com/>, Norwegian centre for research Data )
- ◆ CKAN (<https://ckan.org/>, Open Knowledge Foundation)

Présentation Dataverse (Dimitri Szabo, INRA)

[:https://drive.google.com/file/d/13k6SLYEzWpM2lcVndvxfBWO67rc86bJc/view?usp=sharing](https://drive.google.com/file/d/13k6SLYEzWpM2lcVndvxfBWO67rc86bJc/view?usp=sharing)

- Revue et comparatif d'entrepôts de données\*

- ◆ Étude qui analyse les caractéristiques fonctionnelles d'entrepôts et leur gouvernance dont le modèle économique

Comparative review of data repositories		Analyze Boston (CKAN)	data.world	Dryad	figshare
Categories	Software Features				
Publishing & Versions	Ability to access older versions of files	No	No	Yes - on data package page	Yes
Customization	Ability to create a "Dataverse" or Repository bringing together multiple datasets	NA	No	No	Yes
Customization	Ability to customize the look of your "Dataverse" or collection	NA	No	No	No
File Upload & Handling	Ability to embargo files	NA	No	Yes	Yes
Account & User Info	Ability to list your ORCID in your profile (separate from login/auth)	NA	No	No	Yes
Customization	Allowing user to select a "featured dataset"	Yes	No	No	No
	Analyzing tabular data				

\* *A Comparative Review of Various Data Repositories*

<https://dataverse.org/blog/comparative-review-various-data-repositories>

# Choix d'un entrepôt pour déposer ses données

- **de la discipline** dont relève la recherche

- ♦ [GBIF](#), [DRYAD](#), [Genbank](#), [UniProtKB](#), [Protein Data Bank](#), [PANGAEA](#), [Europe](#), [PDBe](#)...etc



DataSuds

- **De l'institution**

données n'ayant pas vocation à aller dans un entrepôt disciplinaire

- **Des bailleurs**

- ♦ Wellcome Trust Data repositories ([13 entrepôts recommandés](#))
- ♦ H2020 : **entrepôt garantissant la gratuité de l'accès**, de l'extraction, de l'exploitation, de la reproduction et de la dissémination.

- **des revues pour les données qui seront publiées**

- ♦ Recommandent de + en + le dépôt dans un entrepôt disciplinaire (ou à défaut généraliste), voire intègrent un entrepôt dans le processus de publication (Nature + Figshare, Open Journal System + Dataverse...)
  - Nature : [+90 entrepôts recommandés](#)
  - CellPress : [liste par type de données](#)



# Entrepôts de données en France: état des lieux



- Selon **CatOpidor** : wiki des services dédiés aux données de la recherche (géré et hébergé par Inist-CNRS)
  - ◆ 53 entrepôts de données dont
  - ◆ 32 indexés « Science & technologie » (physique, chimie, informartique, sc. Univers, Sc. Terre,...)
  - ◆ 2 indexés « Science de la terre »

# Entrepôts de données en France: état des lieux

- Selon l'annuaire [Re3data.org](http://Re3data.org)  
**Browse by country**

**re3data.org**  
REGISTRY OF RESEARCH DATA REPOSITORIES

Search...



## Institution type

commercial (3)  
non-profit (103)

## Software

DataVerse (5)  
Fedora (1)  
MySQL (4)  
other (19)  
unknown (62)

## Repository types

disciplinary (95)  
institutional (13)  
other (5)

## Data access

closed (9)  
embargoed (15)  
open (82)  
restricted (44)

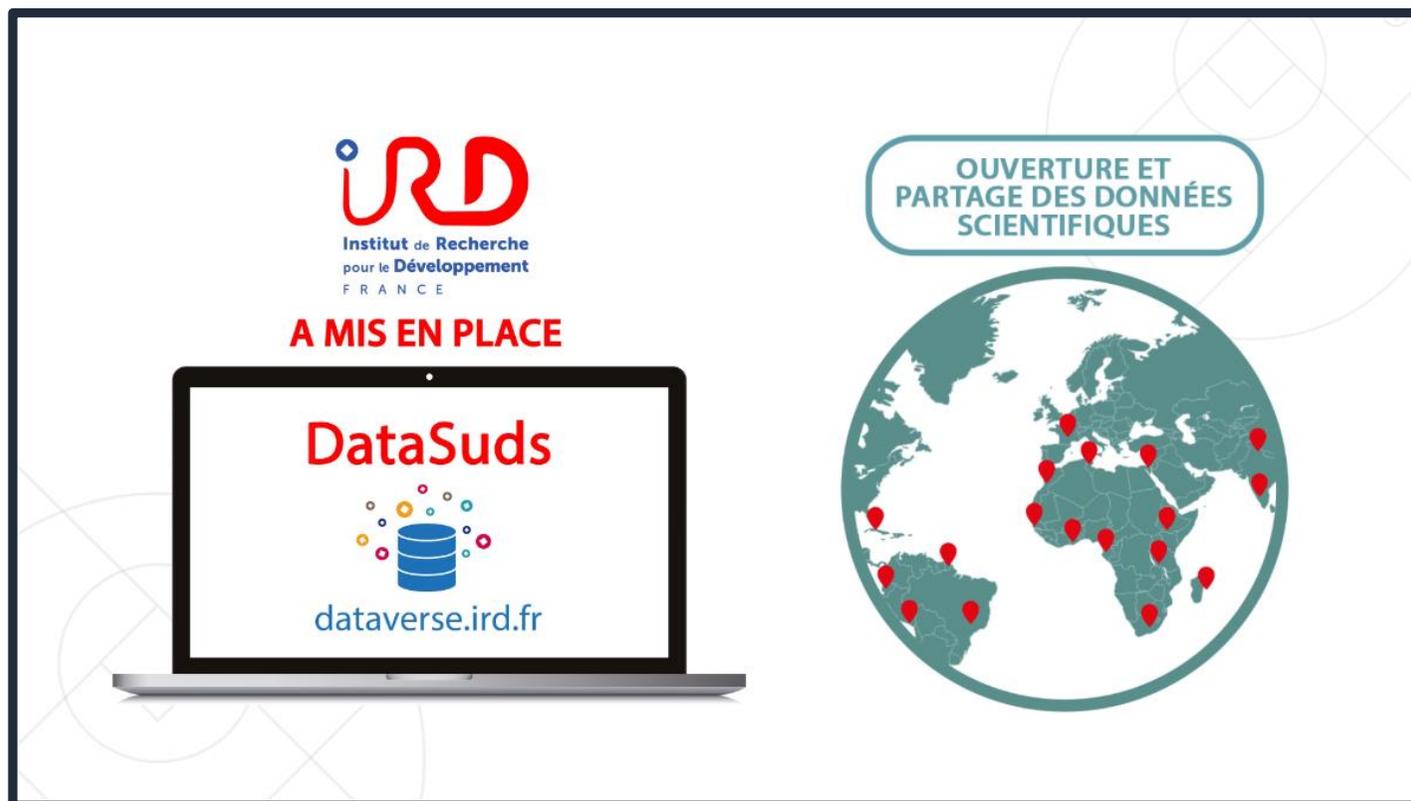
## PID systems

ARK (3)  
DOI (29)  
PURL (1)  
URN (1)  
hdl (5)  
none (60)  
other (4)

## Data licenses

CC (31)  
CC0 (2)  
Copyrights (33)  
ODC (4)  
OGL (1)  
Public Domain (7)  
other (49)

# DataSuds : l'entrepôt Dataverse de l'IRD



Youtube : <https://www.youtube.com/watch?v=oGT81upsITY>

**Des données ouvertes  
pour une science durable au Sud**

# Pourquoi l'entrepôt de données DataSuds ?

Un entrepôt de plus ???!



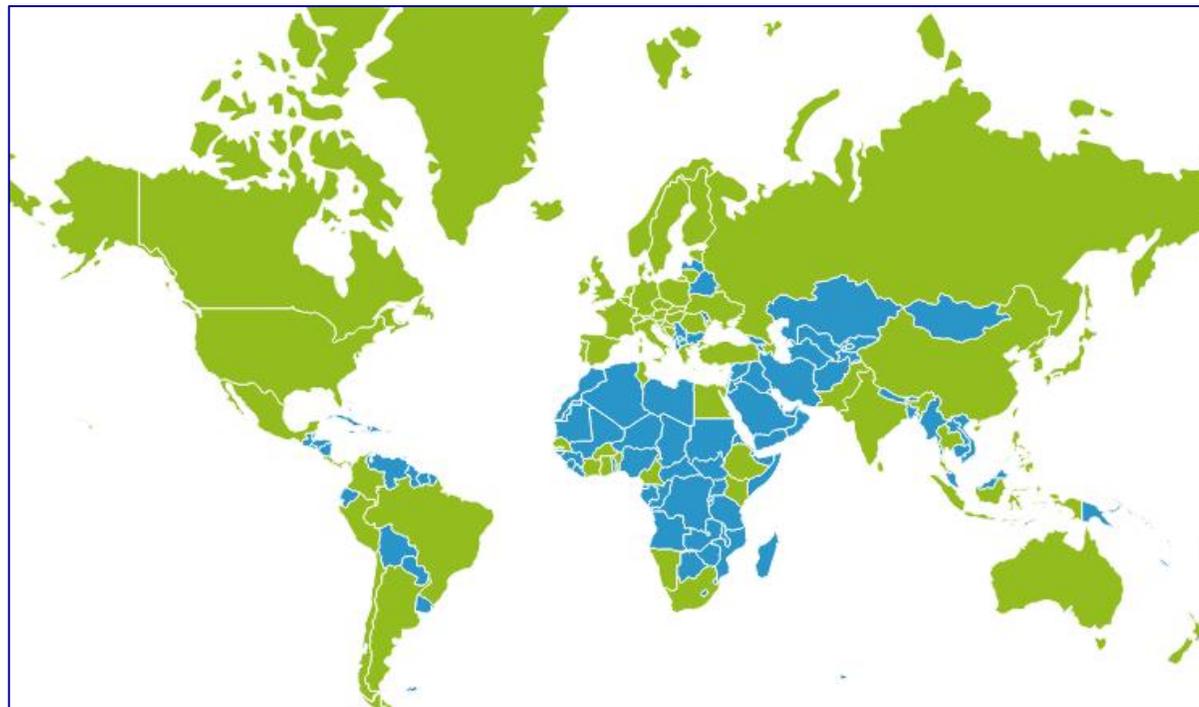
# Enjeux pour l'IRD et la science au Sud

- **Visibilité, partage et accès** aux données des UMR, LMI,...
- **Maîtrise** de la diffusion des données (licence, niveau d'accès...)
- **Ethique** : rendre les données plus facilement accessible à vos partenaires du sud, obtenir leur accord pour la diffusion
- **Valorisation** : être visible pour susciter des collaborations domaine de la recherche et secteur privé



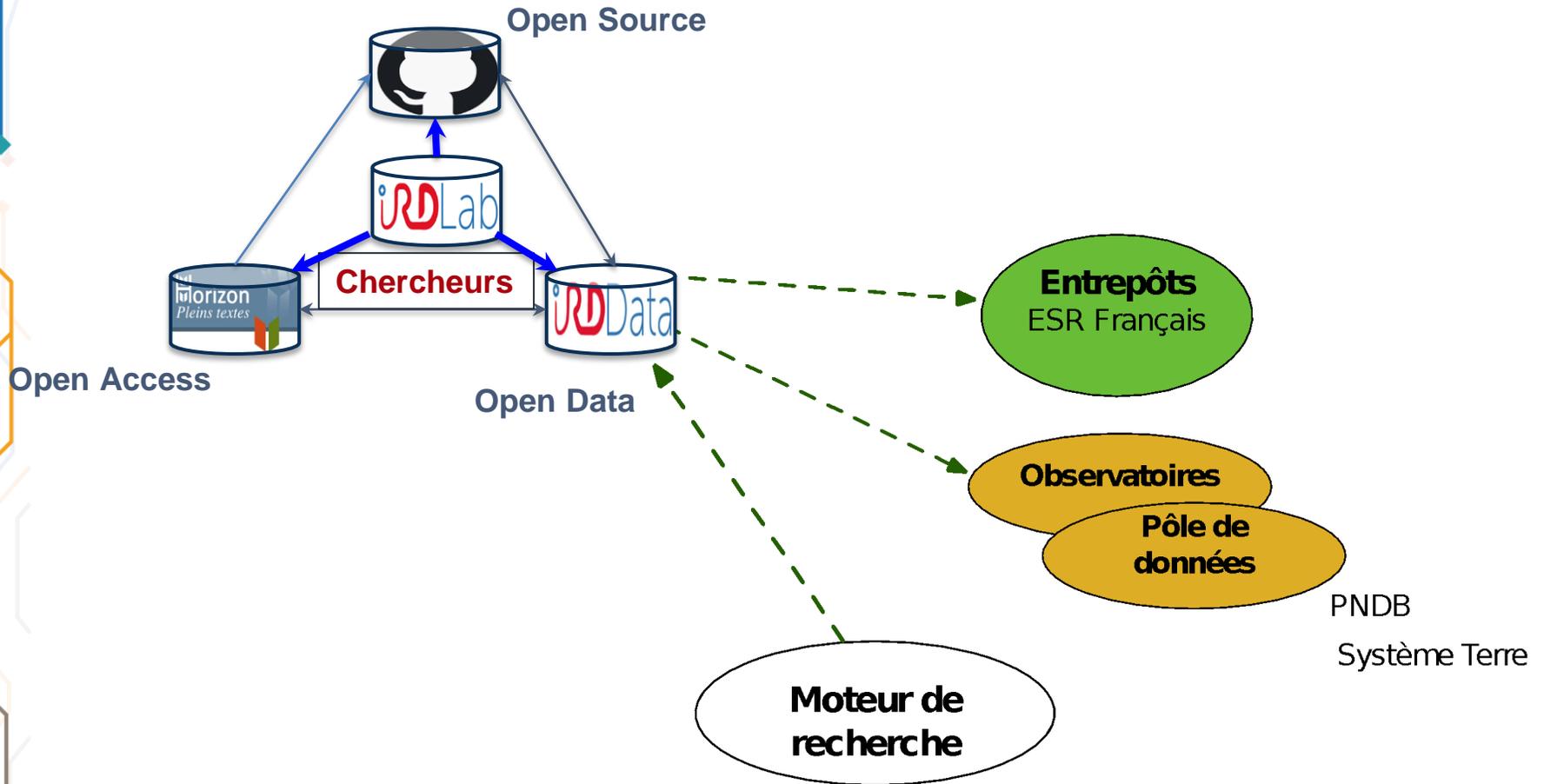
# Une science ouverte au Sud

- Enjeux scientifiques et sociétaux
- Ethique et maîtrise de la diffusion des données



Cartographie des entrepôts de données selon re3data.org  
(09/2019)

# Coordonner les actions et interconnecter les dispositifs de données



# Fonctionnalités de l'entrepôt DataSuds

Agropolis International - Montpellier, 6 septembre 2019

**Des données ouvertes  
pour une science durable au Sud**



# Consulter les données

# Recherche libre



Filtres dynamiques

The screenshot shows the DataSuds (IRD) search interface. At the top, it says "DataSuds (IRD) En savoir plus...". Below that, there are statistics: "69 téléchargements" and options for "Personne-ressource" and "Partager". A search bar contains the text "Chercher dans ce dataverse..." and a "Chercher" button. Below the search bar, there are 11 results out of 17. The results are listed in a table with columns for document type, title, date, and source. The first result is "Impact of soil biodiversity in Southern Tunisia in 2019" by Jouquet, P.; Aroul, H.; Henry-des-Tureaux, T.; Podwojewski, P.; Garcia Ibarra, F.; Labladi, M., 2019. The second result is "soils\_UGA\_AD\_2001.png" from 2019. The third is "soils\_UGA\_NV\_2001.png" and the fourth is "soils\_UGA\_RC\_2001.png". The fifth result is "UMR AMAP (IRD - French National Research Institute for Sustainable Development)" from 2019. The sixth is "UMR ESPACE-DEV (IRD - French National Research Institute for Sustainable Development)" from 2019. The seventh is "UMR IEES-Paris (Institut d'écologie et des sciences de l'environnement de Paris (IEES Paris))" from 2019. On the left side of the interface, there are dynamic filters for "Ensembles de données (5)", "Fichiers (8)", "Date de publication", "Sujet", "Thématique scientifique", "Mots-clés", "Couverture géographique", "Type de fichier", and "Accès".



# Ensemble de données, attribution de DOI et citation



Identifiant



Métadonnées



Licence

**Dataverse** Recherche À propos Guide d'utilisation Support Français Se connecter

UMR ESPACE-DEV (IRD - French National Research Institute for Sustainable Development)

DataSuds > UMR ESPACE-DEV > **Base de données insulaires mondiale (BIM)**

Statistiques 64 téléchargements Personne-ressource Partager

**Base de données insulaires mondiale (BIM)** Version 1.0

DEPRAETERE, Christian, 2019, "Base de données insulaires mondiale (BIM)", <https://doi.org/10.23708/T37S0K>, DataSuds, V1

Citer l'ensemble de données

Pour en apprendre davantage sur ce document :

- EndNote XML
- RIS
- BibTeX

**Description** La base de données SIG (GIS en Anglais) vecteurs BIM concerne toutes les îles océanes de plus de 0,06 km<sup>2</sup> (6 hectares, 116500 îles) avec des informations géographiques, toponymiques, altimétriques, climatiques, géologiques "par île". A ces données "par île" s'ajoutent des données thématiques par point ou polygone: bathymétrie (ETOPO5), inlandsis 15000 BP, climat actuel et scénarii 2050, thésaurus (îles, lacs, marais, lagunes, lieux habités, ...), relief, visibilité en mer alias ligne d'horizon, etc. (2019)

**Sujet** Earth and Environmental Sciences

**Mot-clé** Nissologie, Etudes insulaires, Système d'Informations Géographiques (SIG)

**Related Publication** Depraetere, C. (2008a). The Challenge of Nissology: A Global Outlook on the World Archipelago. Part I: Scene Setting the World Archipelago. Island Studies Journal, Vol. 3, No. 1, 1-16.

**Remarques** La nissologie (nissology en Anglais) fait référence à la science des îles. La base BIM est un outil d'étude général pour les Etudes insulaires (Island studies en Anglais).

Fichiers Métadonnées Conditions Versions

Chercher dans cet ensemble de données... Chercher

1 à 6 de 6 Fichiers

<input type="checkbox"/>		BIM_îles.cpg	Inconnu - 5 o - 11 juil. 2019 - 15 téléchargements MD5: ae3b3df9970b49b6523e08759bc957d	Télécharger
<input type="checkbox"/>		BIM_îles.dbf	Inconnu - 191.9 Mo - 11 juil. 2019 - 11 téléchargements MD5: ba499315a27ed731d5ff6943ad350ee5	Télécharger



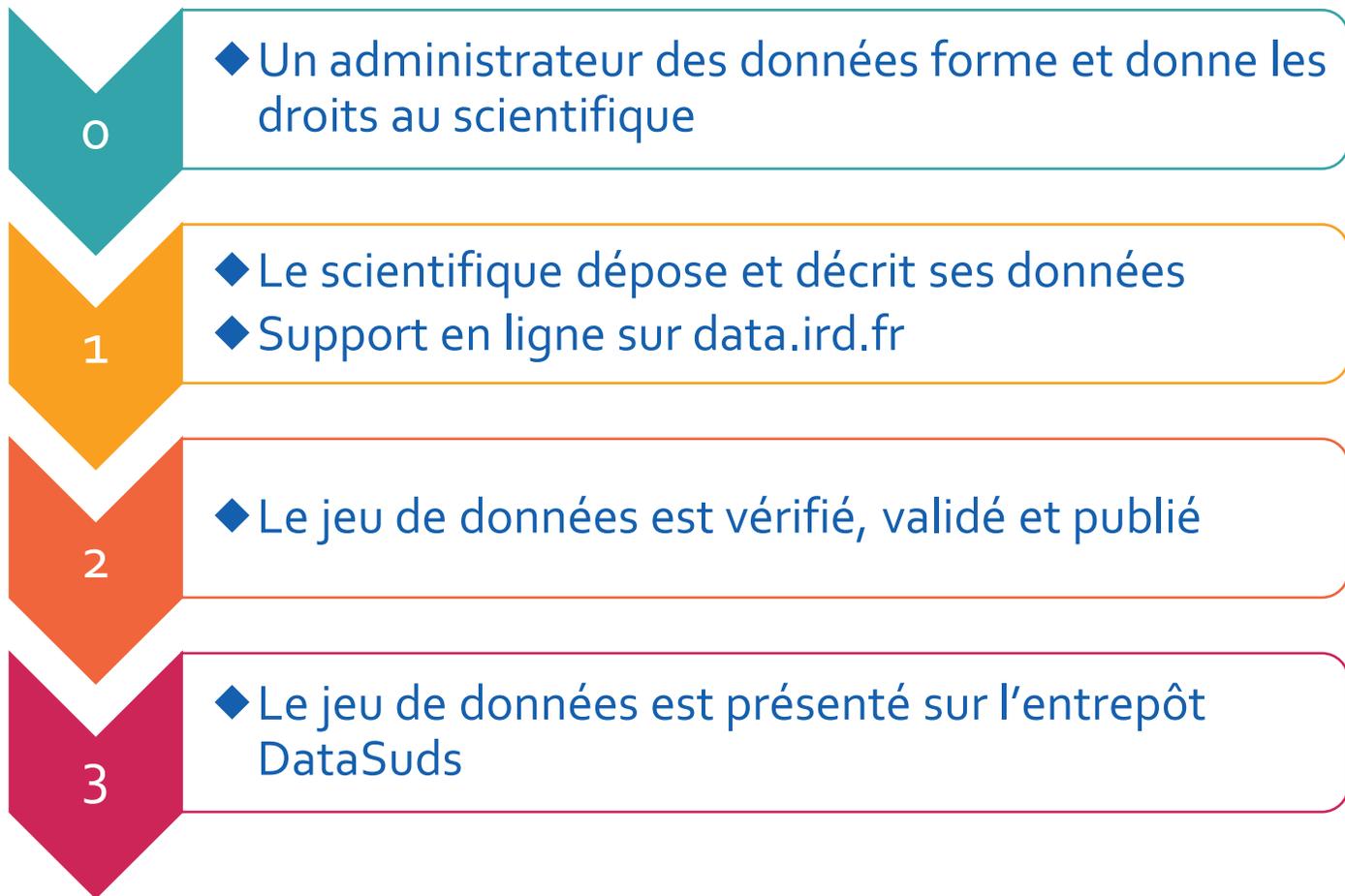
Export citation



Accès fichiers



# Processus de dépôt



# Les plus de Dataverse

Attribution de DOI

Gestion des dépôts  
par le scientifique

Métadonnées  
adaptées à la  
discipline

Autorisation  
Restriction d'accès  
si nécessaire

Organisation en  
arborescence

Lien provisoire  
sécurisé pour les  
reviewers

Fédération  
d'identité de l'ESR  
+ Partenaires

Moissonnage

Gestion fine des  
droits utilisateur





# L'organisation et l'accompagnement autour de DataSuds

**Des données ouvertes  
pour une science durable au Sud**



# Un projet collectif et transversal

Projet **MIDN –IST/MCST – DDUNI**, avec l'appui de :  
DAJ, DMOB (Service Innovation et Valorisation)  
et DRH (Service Développement des talents)

- ✓ **1 administrateur de données** : Luc Decker, appuyé par un réseau de compétences
  - Animation du réseau des référents et administrateurs
  - Appui et formation des irdiens et des partenaires au Sud
  - Qualité, maintien et évolution de DataSuds
- ✓ **1 adresse mail [data@ird.fr](mailto:data@ird.fr)**
- ✓ **1 site support, IRD Data : [data.ird.fr](http://data.ird.fr)**
- ✓ **1 entrepôt, DataSuds : [dataverse.ird.fr](http://dataverse.ird.fr)**



# Une organisation participative



- **4 réunions de recueil de besoins**
- **3 datathons sur 2 Délégations Régionales** (dépôt de données par les scientifiques après sensibilisation)
- **21 administrateurs de 8 unités formés** (programme en cours)
- **53 inscrits à la liste de diffusion des administrateurs de DataSuds** : [dataverse-referents@listes.ird.fr](mailto:dataverse-referents@listes.ird.fr)

# Tous acteurs, tous responsables

Les décisions sont déléguées au plus près de la recherche.

Les **référents** nommés dans chaque structure par leur responsable et formés par l'équipe Data IRD :

- décident du workflow des données,
- forment si besoin des administrateurs de projet ou d'équipe
- conseillent les scientifiques de leur unité
- participent à la réflexion et à l'évolution du dispositif



# Entrepôts de données : Quelques exemples en plus



- Atelier Dataverse pour les entrepôts de données
  - ◆ "Expérience DATAVERSE INRA »
  - ◆ "Expérience DATAVERSE CIRAD »
  - ◆ "Expérience DATAVERSE Sciences Po »

<https://rdafrance2019.sciencesconf.org/>



## Partie 3 : Discussions

# Entrepôts vs plateforme d'accès

	Entrepôt de données	Plateforme d'accès aux données
<b>Nature des données</b> (format, représentation, étendue)	Hétérogène, hétéroclite	Homogène (étendue, représentation)
<b>Temporalité</b>	longue	Réduite à un projet, à un dispositif de recherche (Obs, Infrastructure...)
<b>Fonctionnalités utilisateur final</b>	Réduite : Recherche, téléchargement	Spécifiques et pointues (requêtes, analyse, traitements...)
<b>Administration</b>	Avancée : rôles, workflow de publication	Plus frustre
<b>Objectifs</b>	Préservation, valorisation sur de larges champs disciplinaires	Diffusion, partage sur un champ thématique ou disciplinaire restreint

# En résumé : valeurs ajoutées d'un entrepôt de données

## **Avant tout, un dispositif numérique pour**

- collecter et structurer des métadonnées
- Préserver
- Partager au delà de la discipline
- Publier (DOI, Licence, ..)
- Valoriser les travaux des chercheurs, notamment citer les données

**Apportant une maîtrise du workflow de publication des données au delà des administrateurs ou *data manager***

# En résumé : faiblesses au regard de la spécificité de nos données et nos besoins

## Structurer sur une vision documentaire des données

- positionnement des entrepôts de données en fin de cycle de vie
- vision objet de la donnée (fichiers..) rend difficile son exploitation par d'autres plateformes, pour des services à valeur ajoutée (visualisation, fouille de données, traitements, analyse statistiques)
- Réelle réutilisation, valorisation des données (pas que du chercheur) ???

## Identification et workflow adaptées à des données statiques

- Pas suffisante pour des données spatio-temporelles dynamiques, voire temps réel
- Créer DOI sur des séries temporelles pose problème et doit être investi

# Complémentarité entrepôts vs plateforme d'accès

**Ou comment améliorer nos plateforme d'accès aux données ?**

**en les dotant des fonctionnalités des entrepôts**

- ◆ Citabilité des données via les DOI
- ◆ Maitrise des aspects réglementaires de diffusion
- ◆ Workflows de publication mieux formalisés, implémentés
- ◆ Potentiel de diffusion au delà de la discipline (api vers schema.org, datacite, crossref...)
- ◆ Préconisations plus forte pour utiliser les systèmes d'identification pour mettre les données avec les publications, les acteurs

# Structuration et mutualisation des forces autour des entrepôts de données

- Animation de la communauté
  - ◆ Atelier DataVerse RDA France
  - ◆ Atelier aux Journées Nationales Science Ouverte
- ANR Flash BRIDGE
  - ◆ Gouvernance inter instituts
  - ◆ Fairisation des entrepôts de données
- Groupe entrepôt de données RDA – France
- Etude de faisabilité pour le MESRI – service mutualisé d'un service générique d'accueil et de diffusion des données simples