Les entrepôts de données

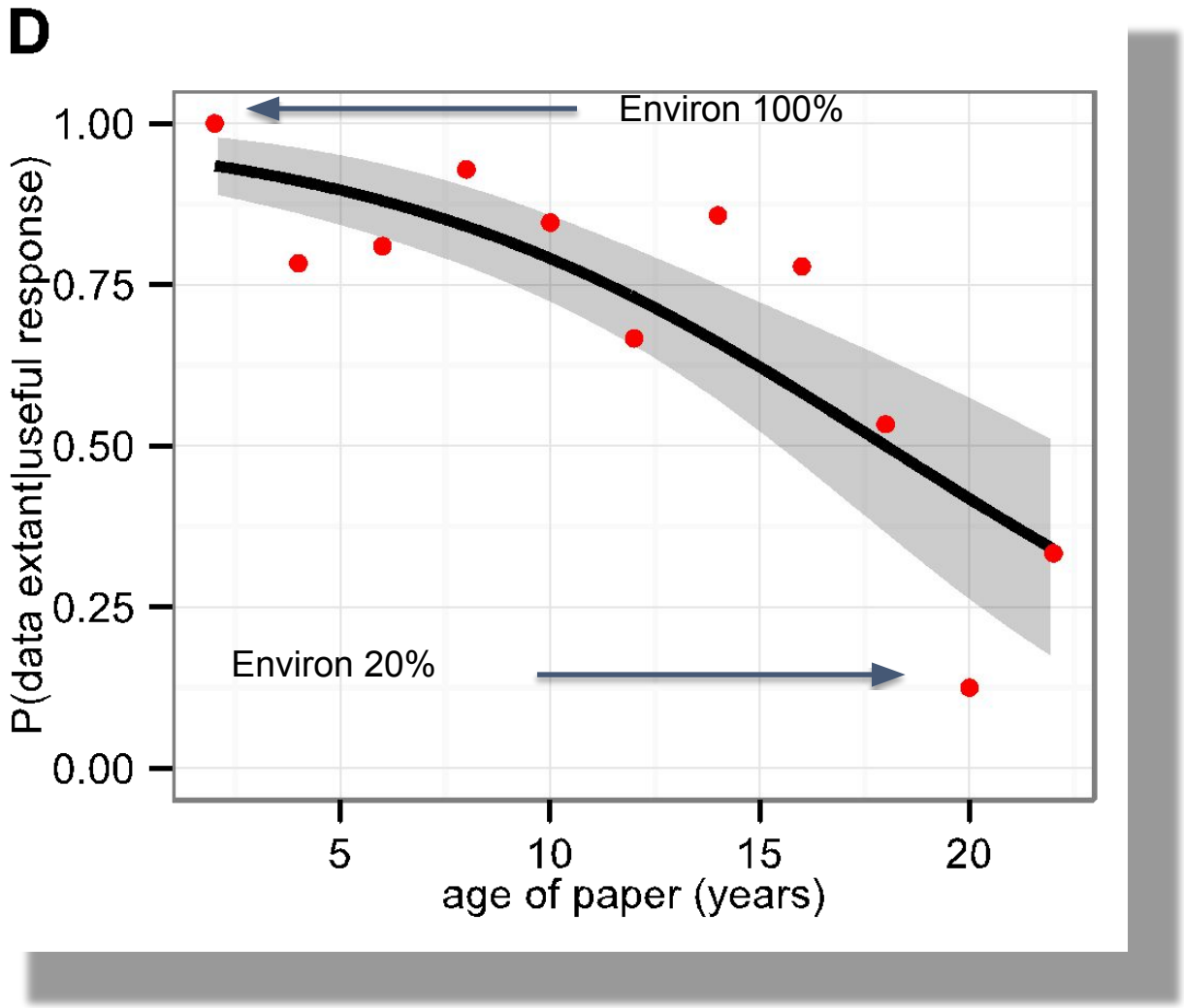
Ou comment rendre les données trouvables, accessibles et réutilisables ?

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



Perte de données de 17% / an

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'affection, départ (retraite...)

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

ODATIS

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
Liberté • Égalité • Fraternité
RÉPUBLIQUE FRANÇAISE
MINISTÈRE DE L'ENSEIGNEMENT SUPÉRIEUR, DE LA RECHERCHE ET DE L'INNOVATION

Partie 1 : les données de la recherche en environnement

ou quelles données cibles pour les entrepôts de données ?

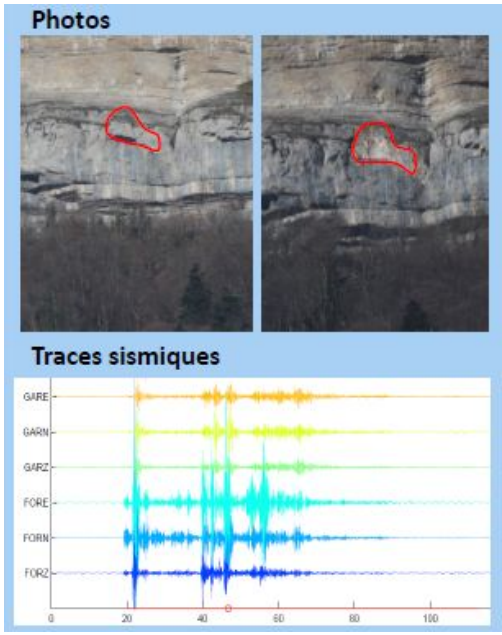
Quelques définitions autour des données

«**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, réutilisables pour enrichir, compléter, étendre d'autres jeux de données en vue d'améliorer la connaissance de l'objet d'étude…

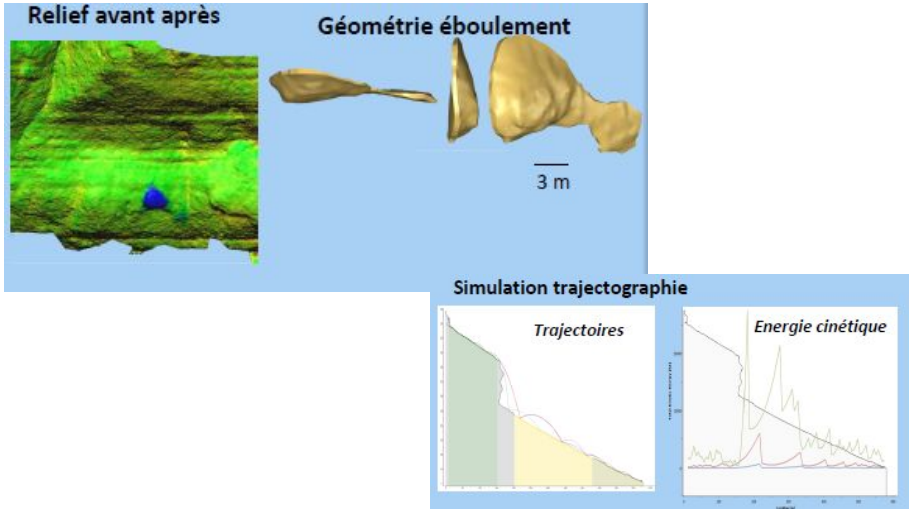
Quelques définitions autour des données : **Données brutes - données élaborées**

- « **Données brutes** : recueillies sur un sujet à partir d'observations ou de mesures, et qui n'ont pas encore été traitées, validées ou triées. Censées 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.



D'après Sylvie Cocaud. Cocaud et Aventurier 2017 : http://dx.doi.org/10.15454/1.4993537478868977E12

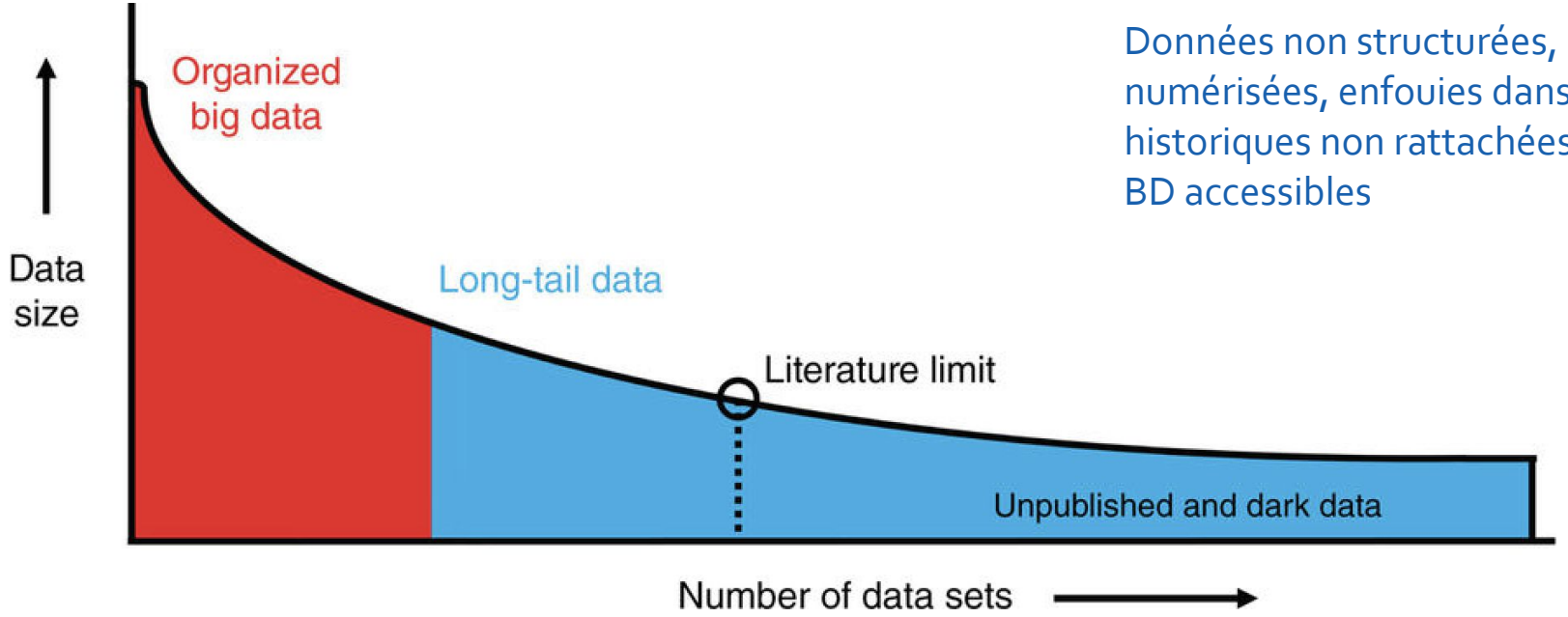
Big data vs Long tail of data

Big data

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

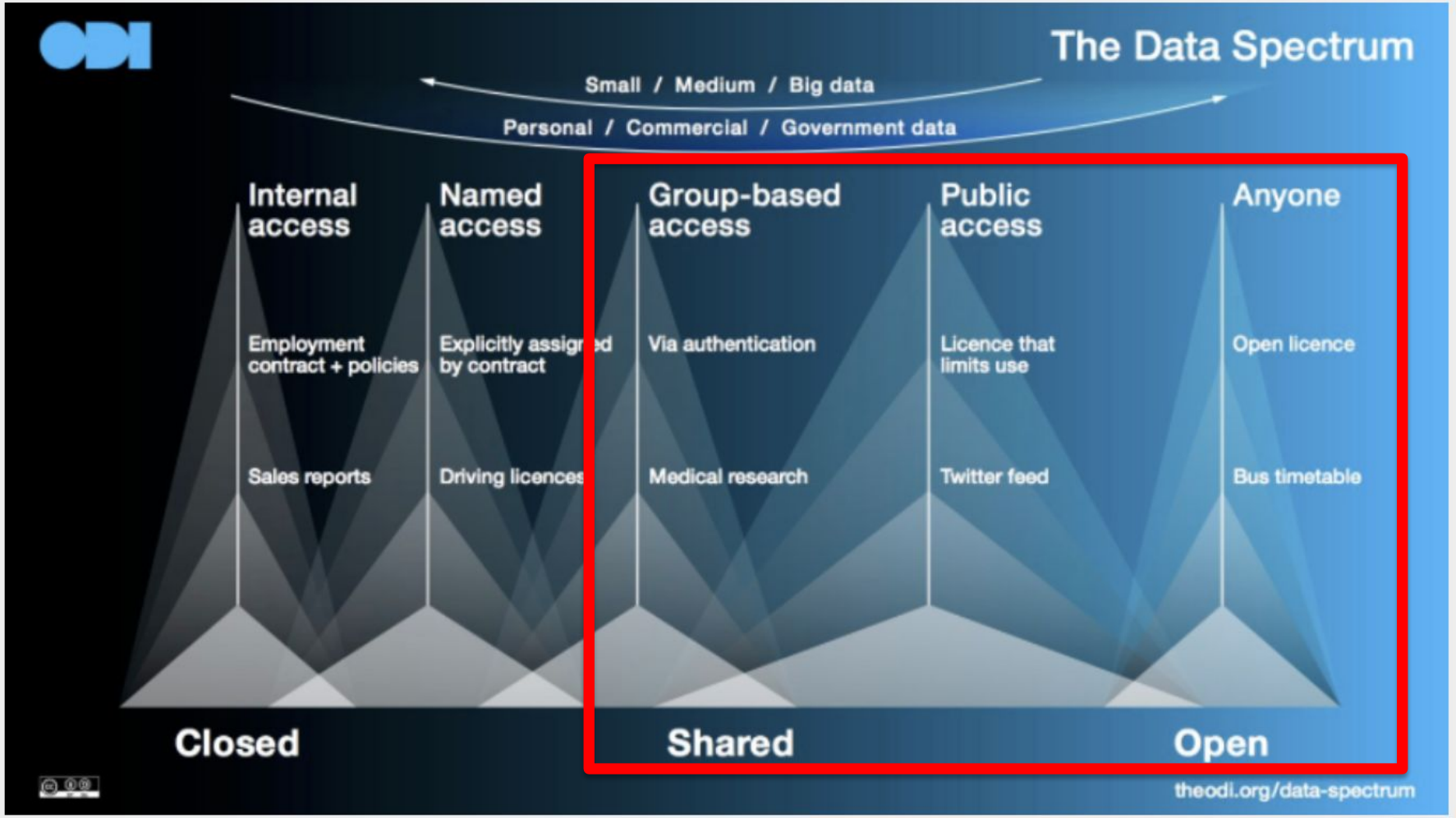
La longue queue des données

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)

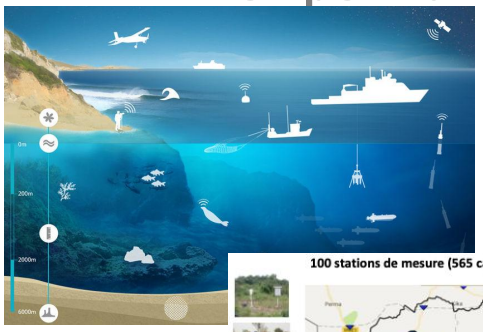
Données fermées/données partagées/données ouvertes



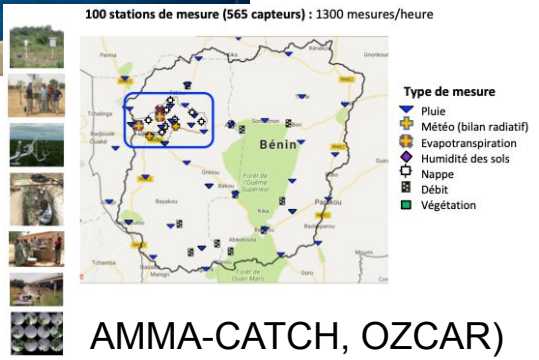
(d'après l'Open data institute : https://theodi.org/about-the-odi/the-data-spectrum/)

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

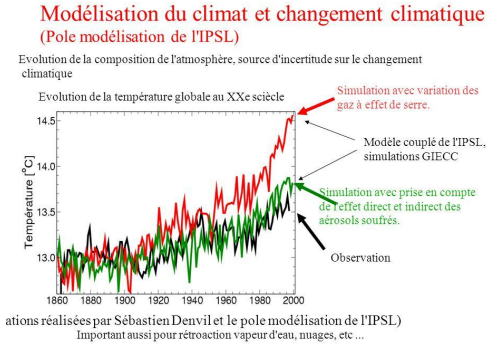


SISMER, ODATIS)



AMMA-CATCH, OZCAR)

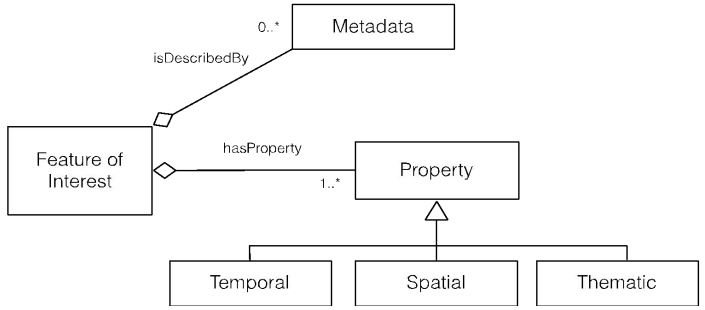
SPOT 6



  - 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

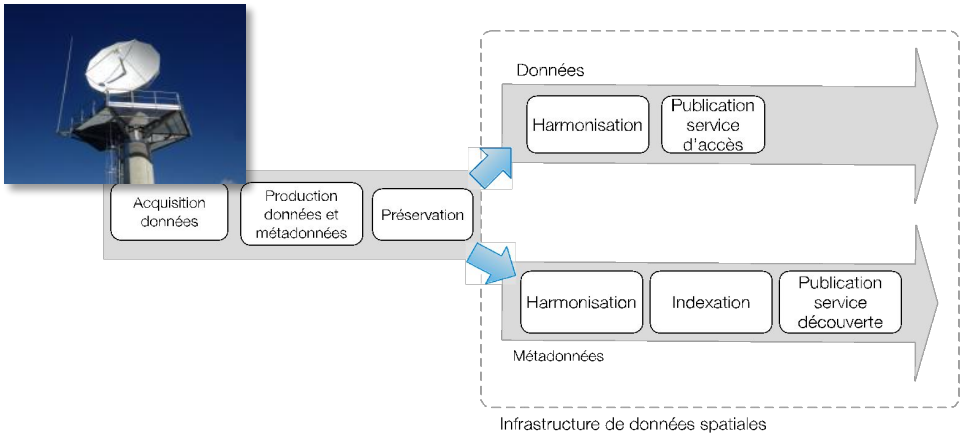


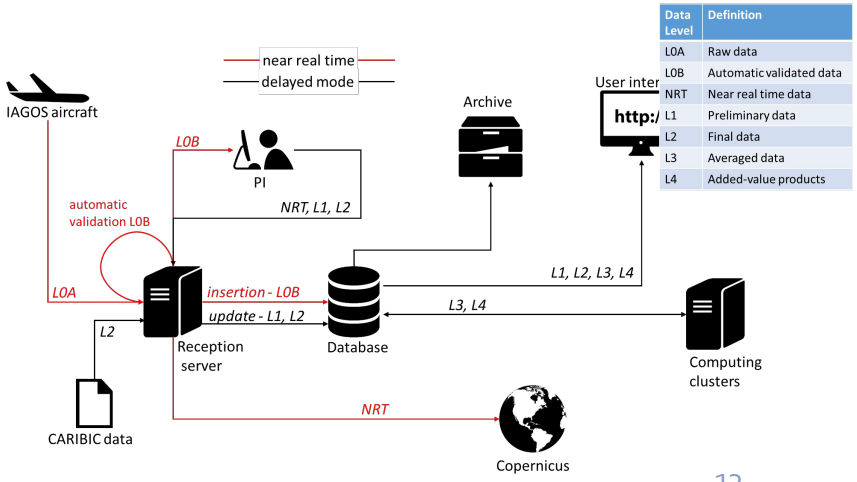
--> induit le besoin d'outils et de services spécifiques pour valoriser, réutiliser les données (visualisation, analyse, traitements…)

  - 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'hétérogénéité 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

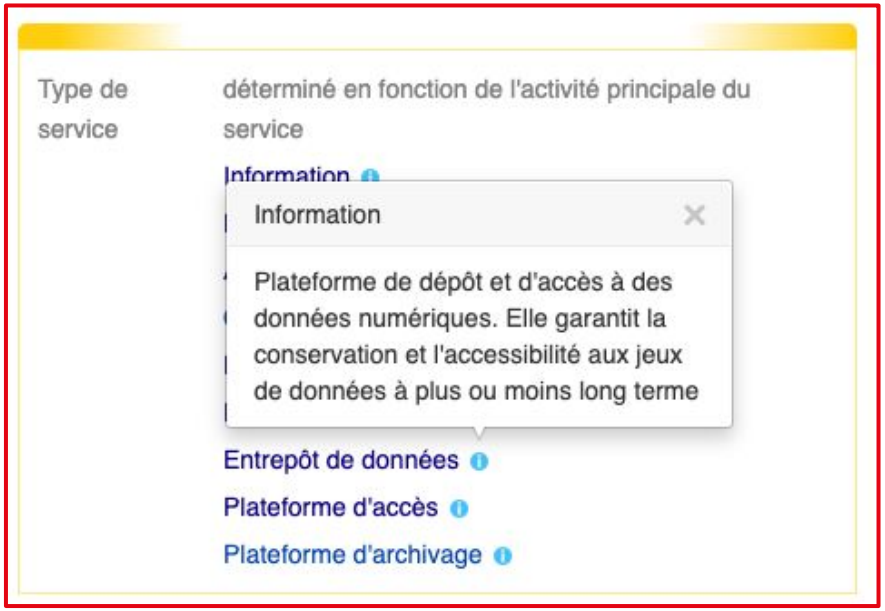




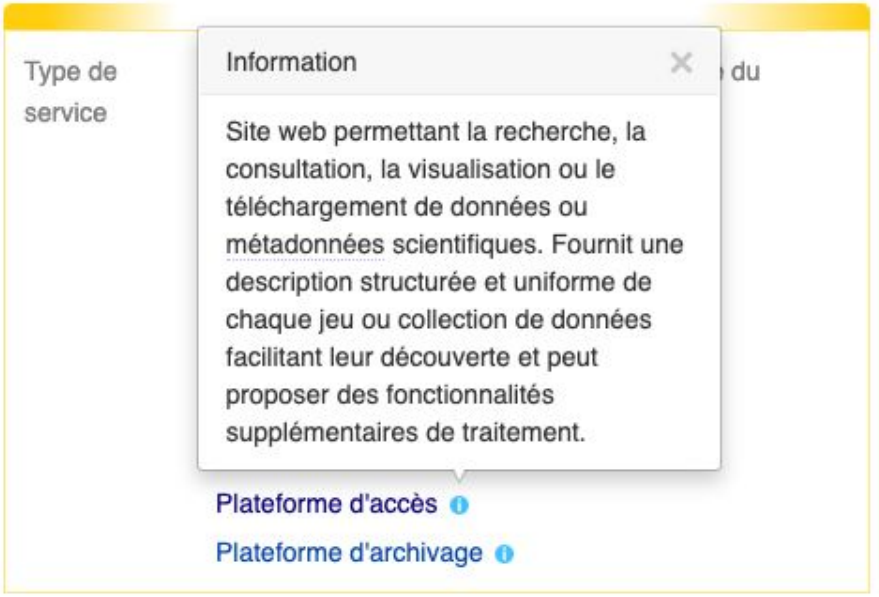
Data Level	Definition
L0A	Raw data
L0B	Automatic validated data
NRT	Near real time data
L1	Preliminary data
L2	Final data
L3	Averaged data
L4	Added-value products

Partie 2 : notions et fonctionnalités des entrepôts de données

Entrepôt de données : définition



≠



≠

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

https://cat.opidor.fr/index.php/Description_d%27un_service : Cat OPIDoR, wiki des services dédiés aux données de la recherche)

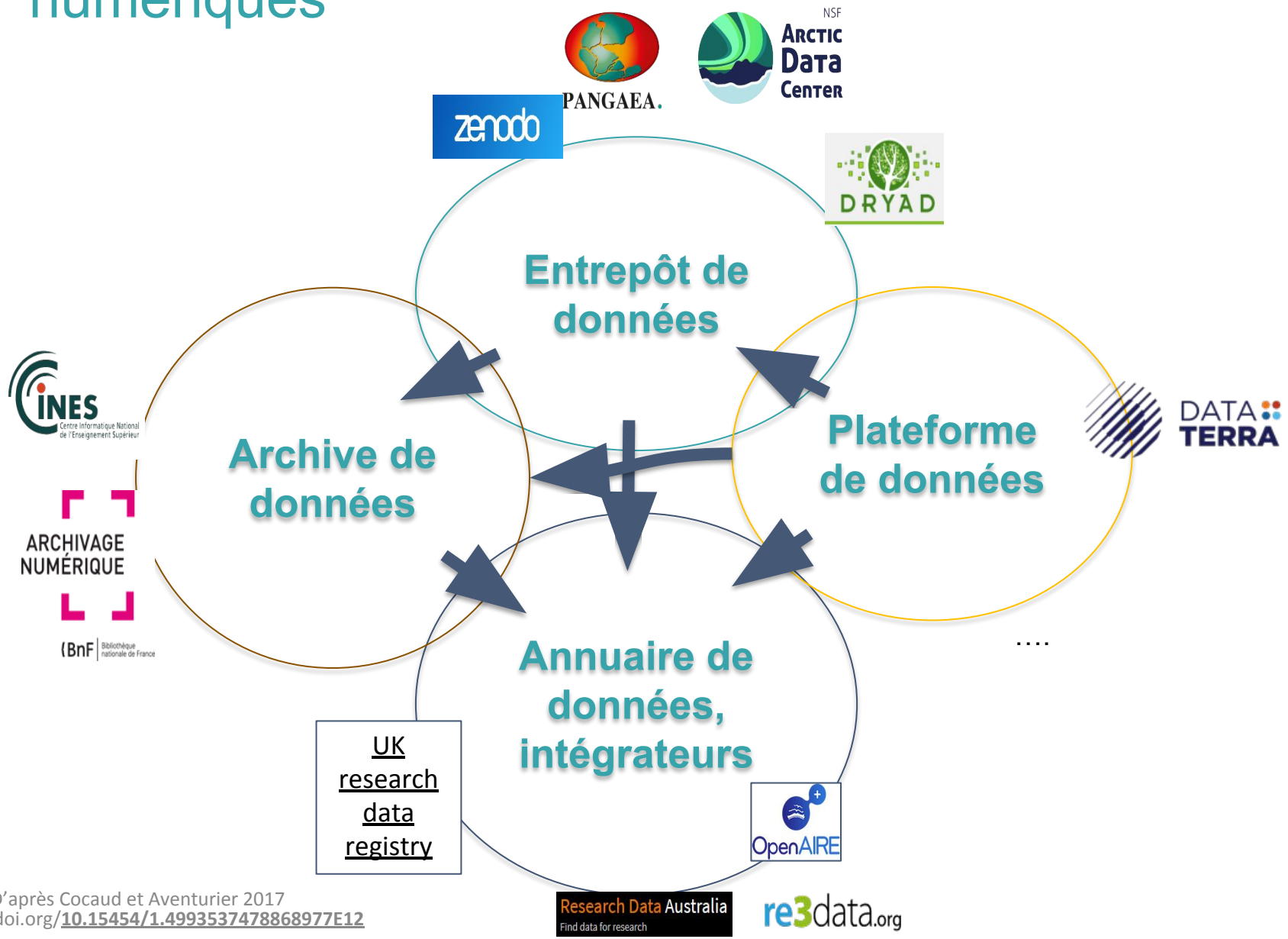
Entrepôt de données : définition

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

D'après Cocaud et Aventurier 2017
http://dx.doi.org/**10.15454/1.4993537478868977E12**

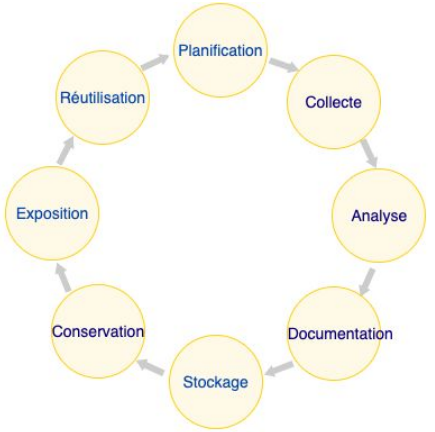
Les entrepôts en lien avec les autres dispositifs numériques



D'après Cocaud et Aventurier 2017
http://dx.doi.org/10.15454/1.4993537478868977E12

Entrepôt de données dans le cycle de vie de la donnée

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

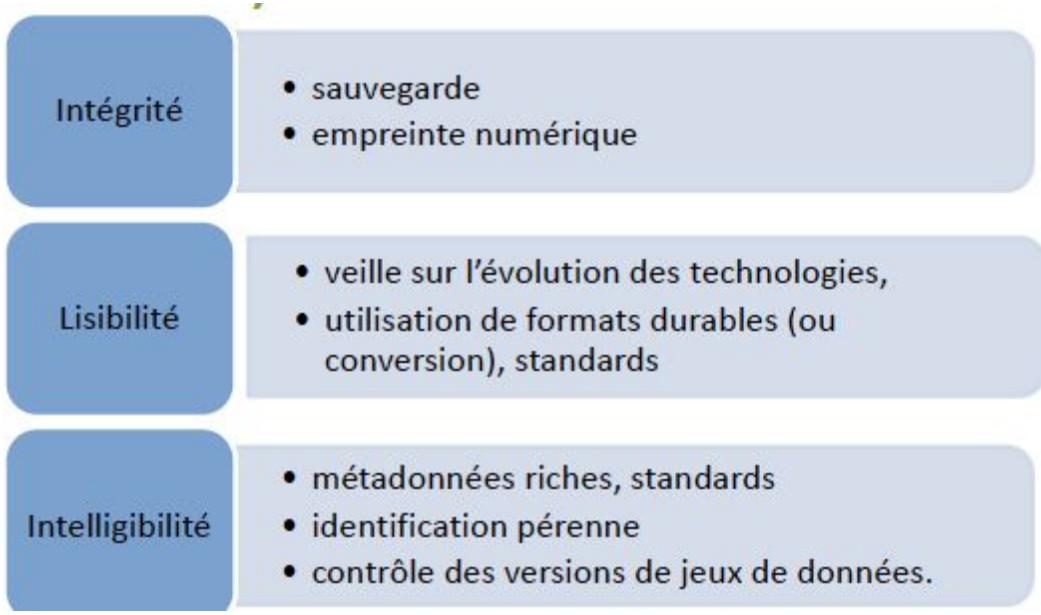


Cycle de vie de la donnée (CatOpidor)

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

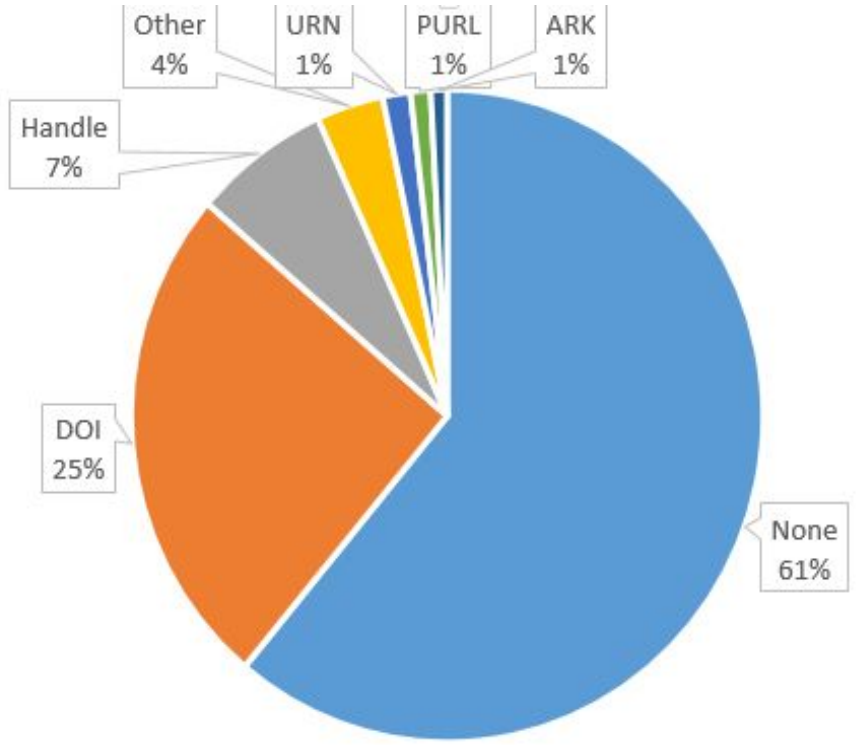
- Conservation (stockage sécurisé + archivage à long terme)



D'après Cocaud et Aventurier 2017

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)

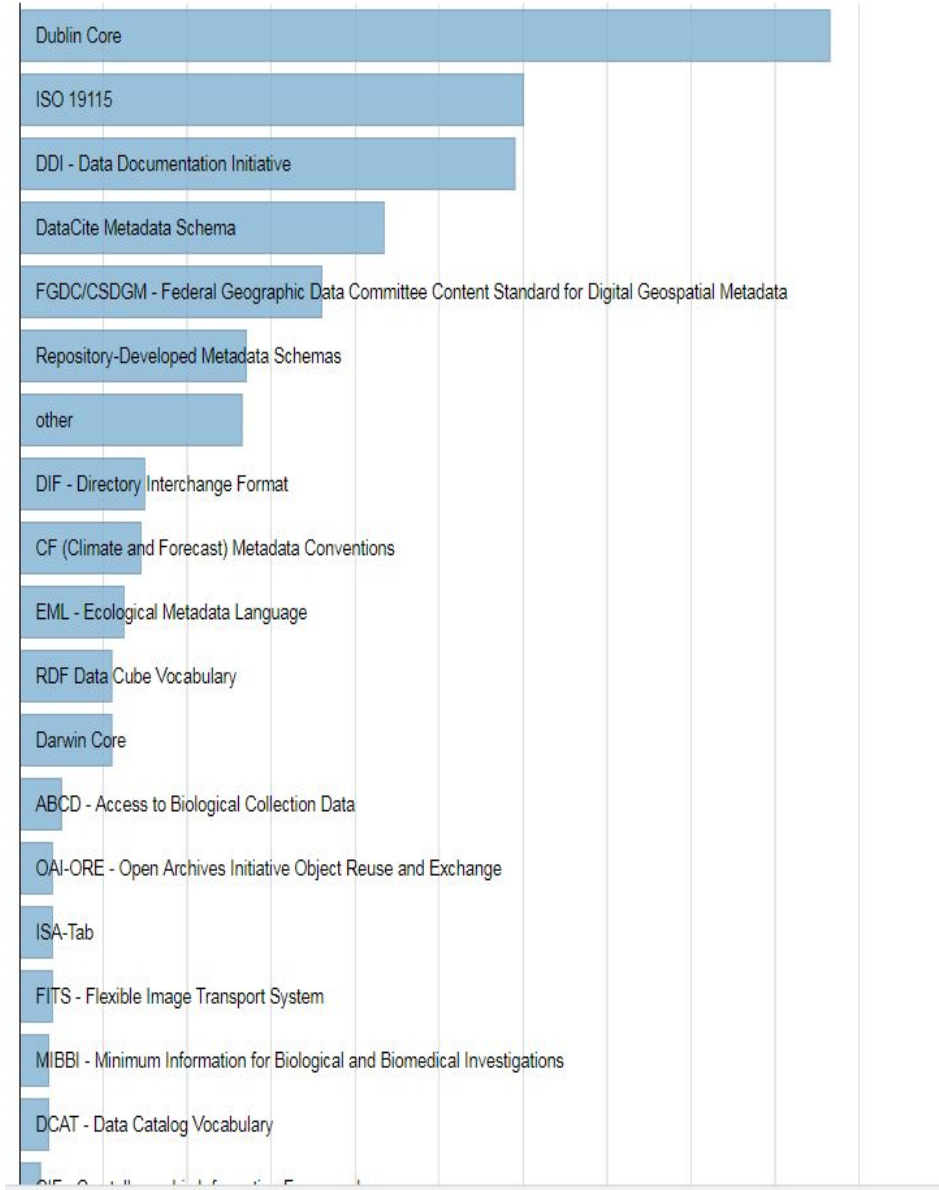


Données extraites de re3data.org le 19/05/2017 : total = 2042 entrepôts

D'après Cocaud et Aventurier 2017

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



Données extraites de re3data.org le 29/04/2018 : total = 2042 entrepôts

D'après Cocaud et Aventurier 2017

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

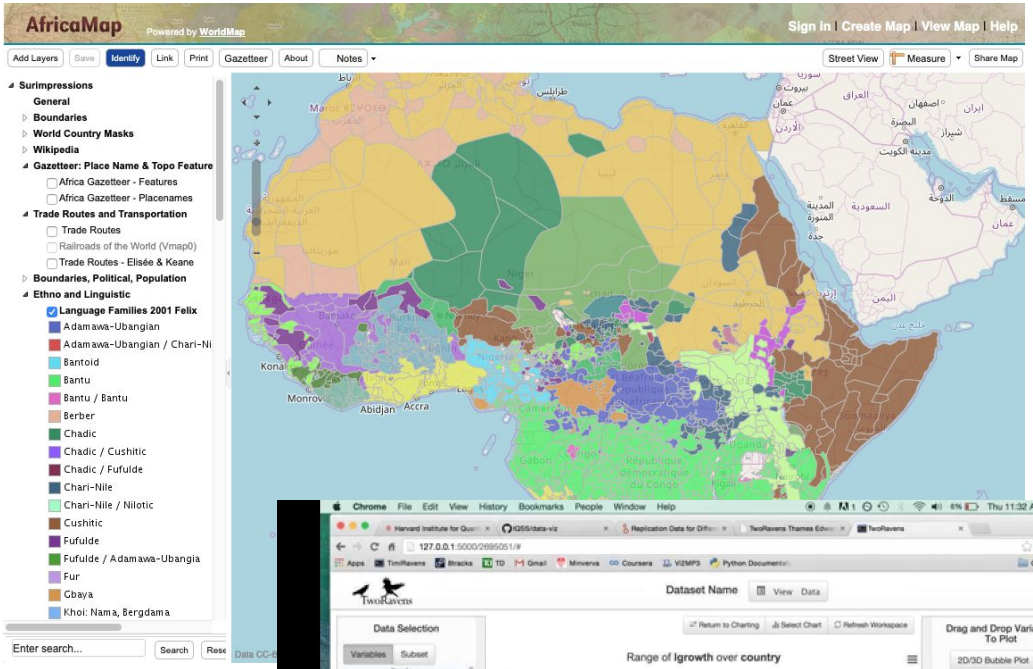
- Droit d'accès aux données :
  - téléchargement libre (open)
  - Embargo
  - demande d'accès (restricted), guestbook
  - pas de téléchargement possible (closed)
- Génération d'URL privé parfois proposé
- Attribution d'une licence à chaque jeu de données
  - Saisie libre vs liste fermée
- Attention aux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

D'après Cocaud et Aventurier 2017

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/, Norvegian 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

- 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	figsh
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 (seperate 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 Dat[PANGAEA] Europ (PDBe)…etc
- **De l'institution**

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

  DataSuds

  SEANOE
- **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

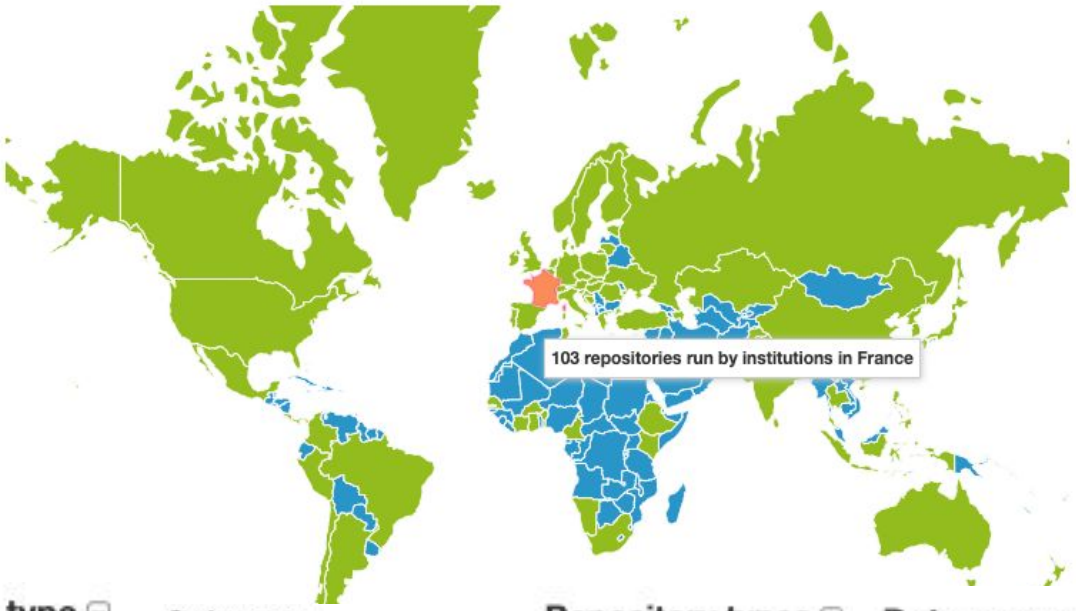
- Selon l'annuaire Re3data.org

re3data.org
REGISTRY OF RESEARCH DATA REPOSITORIES

Search... Search

Browse by country

Graphical | Text



Institution type
- commercial (3)
- non-profit (103)

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

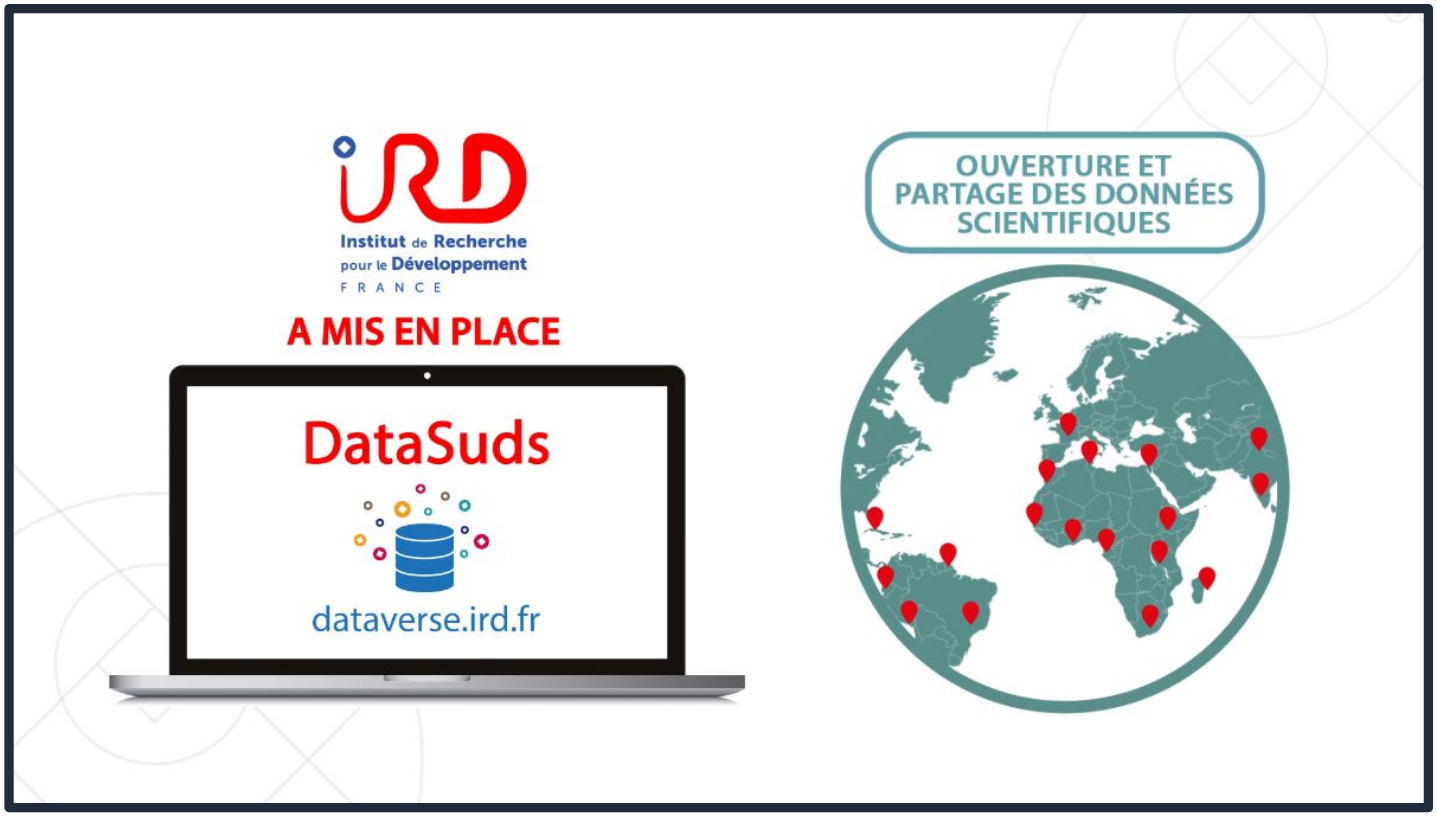
Repository types
- disciplinary (95)
- institutional (13)
- other (5)

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 ??!!

dataverse.ird.fr

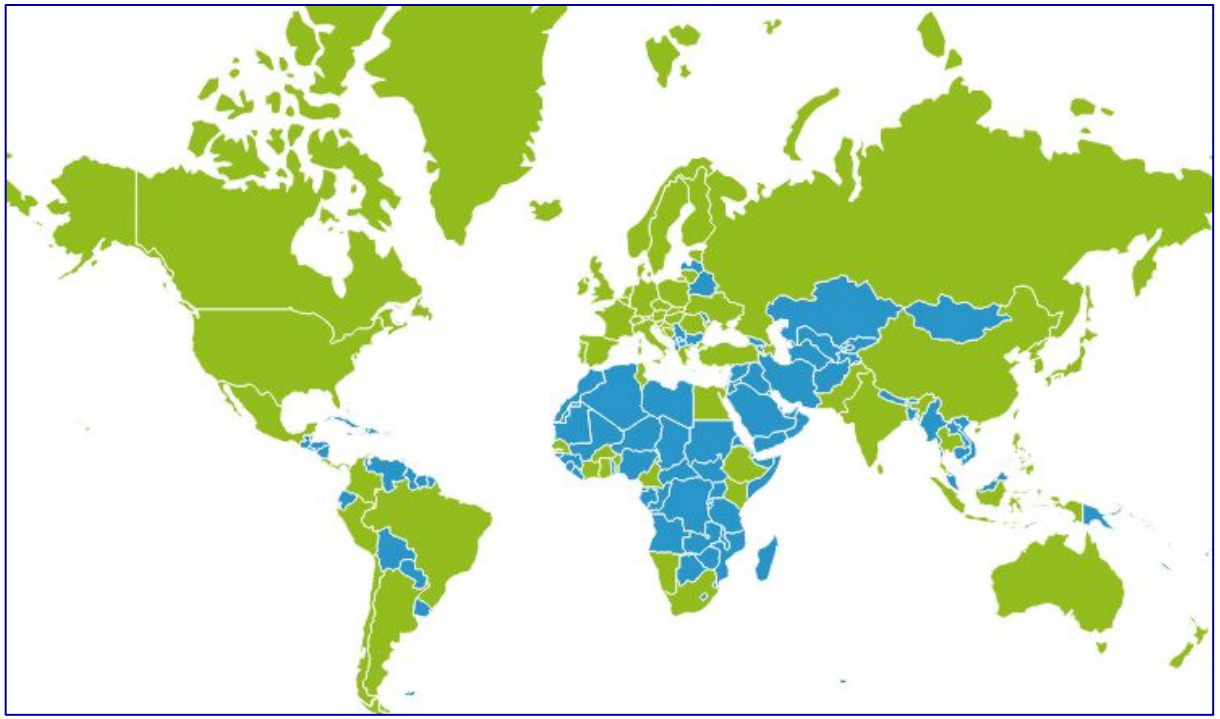
Enjeux pour l'IRD et la science au Sud

- **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

dataverse.ird.fr

Une science ouverte au Sud

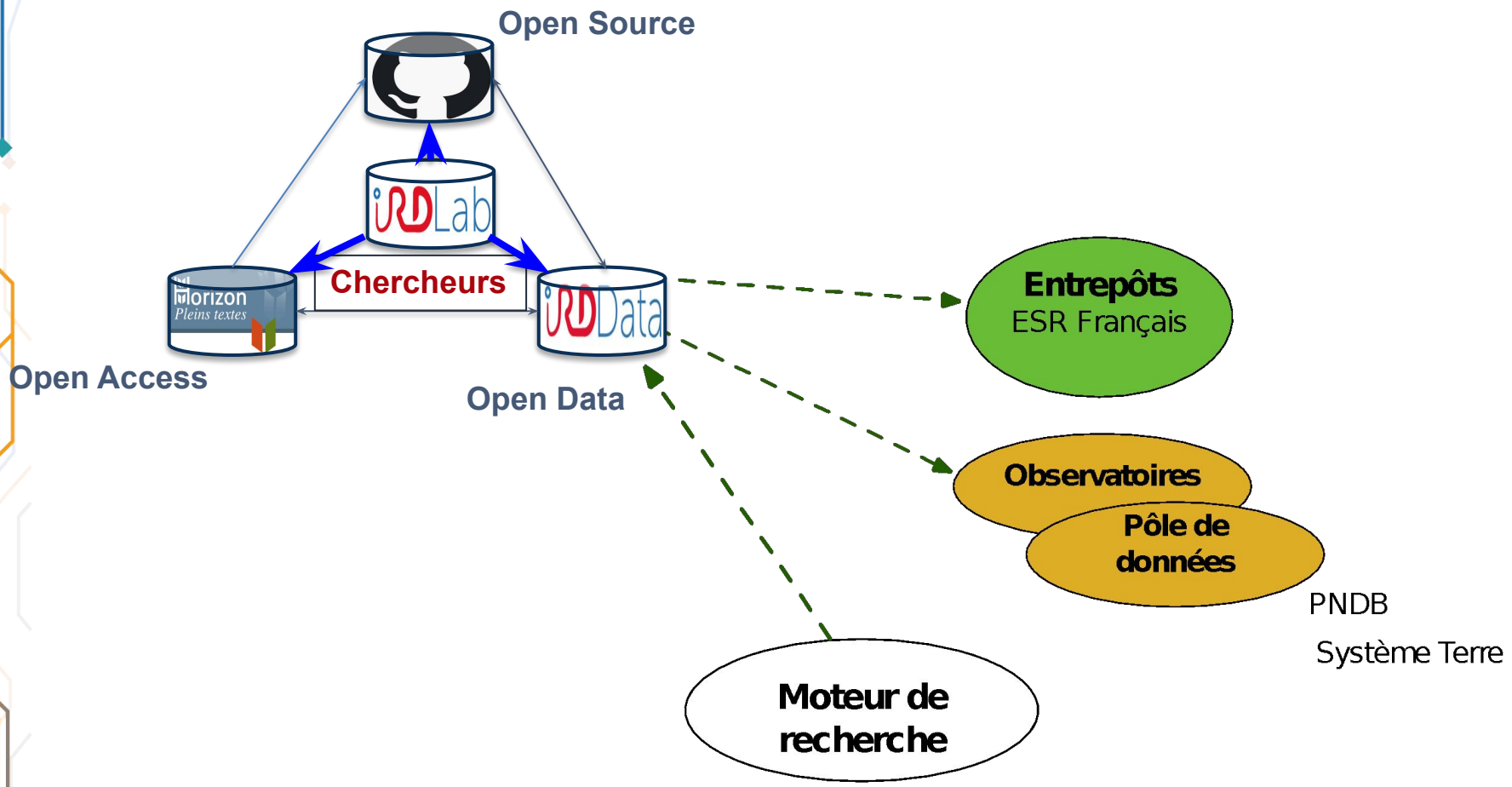
- Enjeux scientifiques et sociétaux
- Ethique et maîtrise de la diffusion des données



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

re3data.org - Registry of Research Data Repositories. https://doi.org/10.17616/R3D last accessed: 2019-09-04

Coordonner les actions et interconnecter les dispositifs de données



Fonctionnalités de l'entrepôt DataSuds

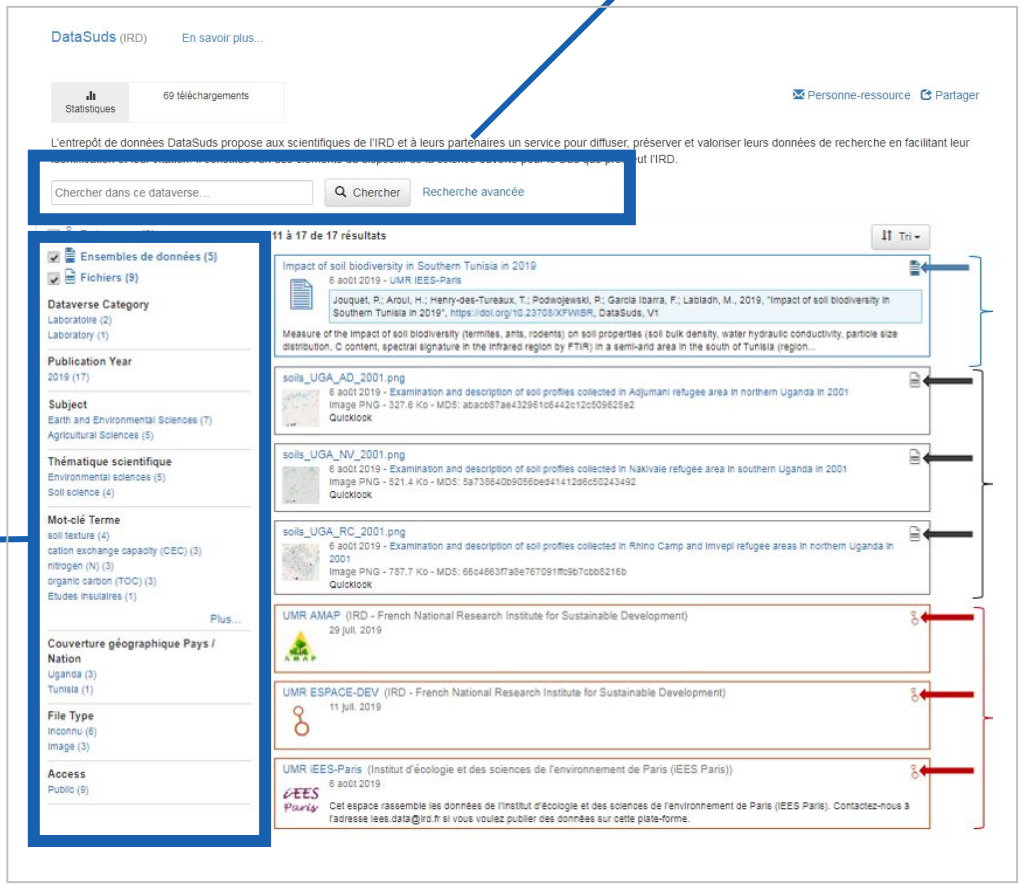
Agropolis International - Montpellier, 6 septembre 2019

Des données ouvertes
pour une science durable au Sud

Consulter les données

Recherche libre

Filtres dynamiques



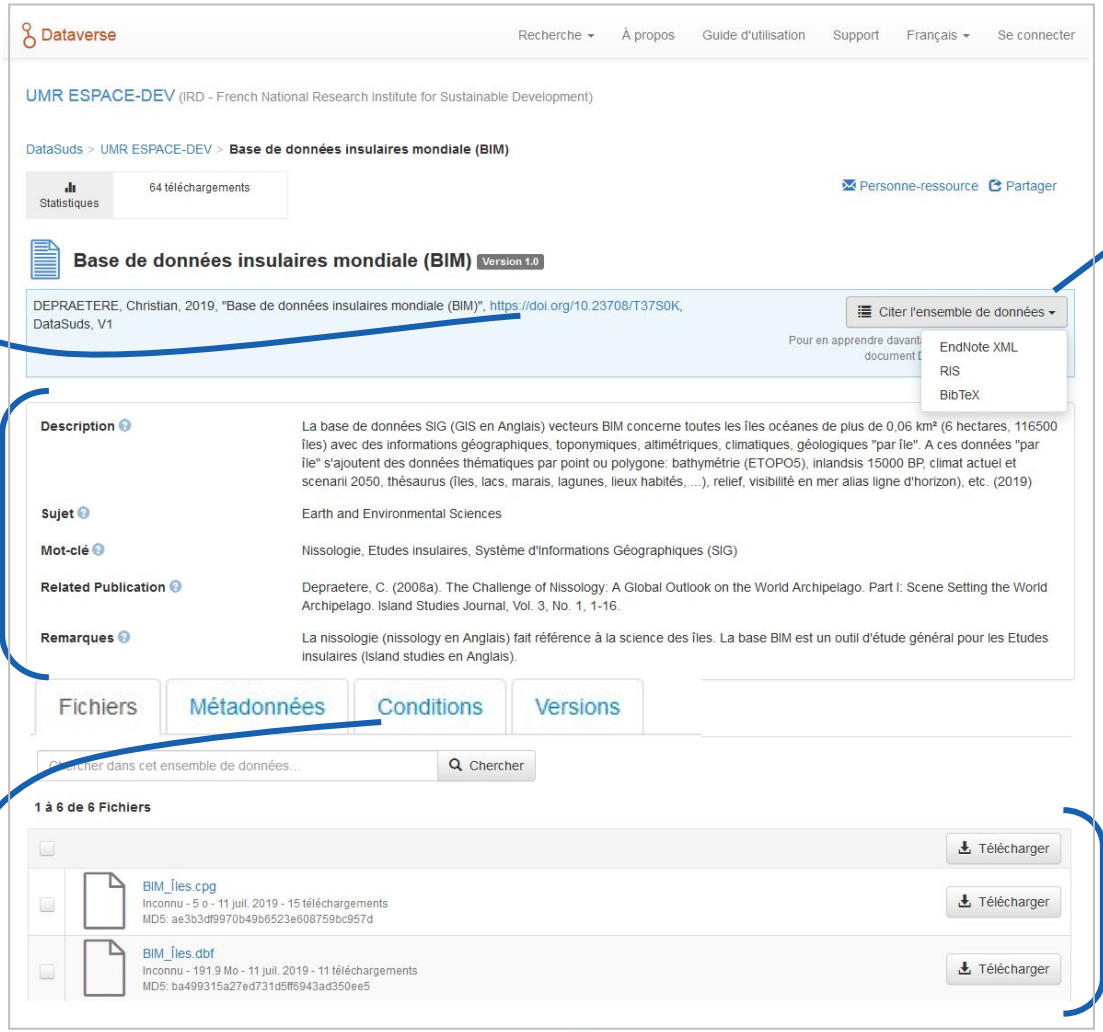
dataverse.ird.fr

Ensemble de données, attribution de DOI et citation

Identifiant

Métadonnées

Licence



Export citation

Accès fichiers

dataverse.ird.fr

Processus de dépôt

0

- Un administrateur des données forme et donne les droits au scientifique

1

- Le scientifique dépose et décrit ses données
- Support en ligne sur data.ird.fr

2

- Le jeu de données est vérifié, validé et publié

3

- Le jeu de données est présenté sur l'entrepôt DataSuds

dataverse.ird.fr

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

dataverse.ird.fr

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
- ✔ **1 site support, IRD Data : data.ird.fr**
- ✔ **1 entrepôt, DataSuds** : dataverse.ird.fr

Une organisation participative

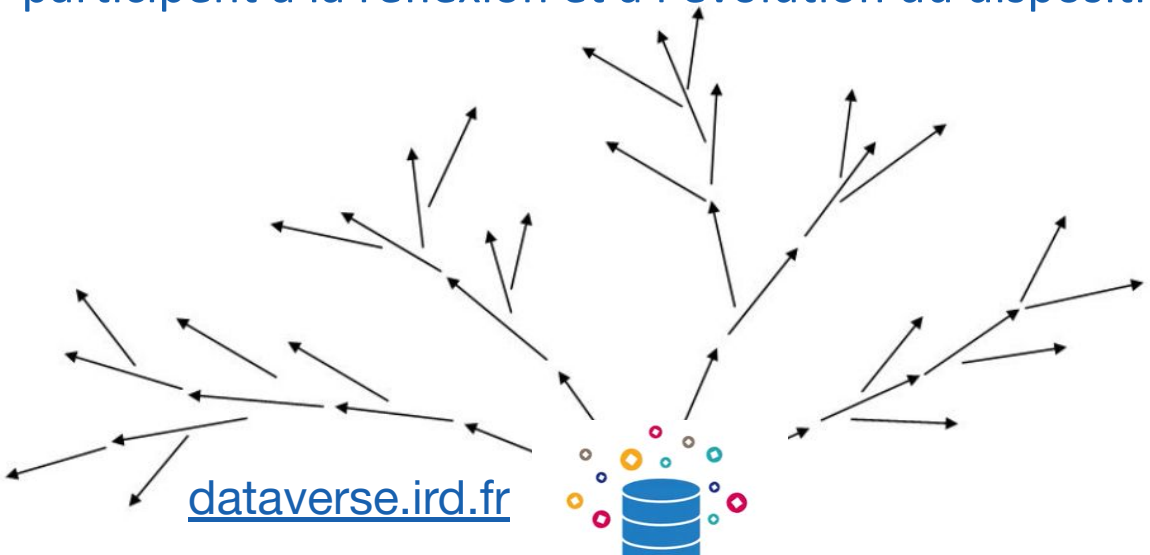
- **4 réunions de recueil de besoins**
- **3 datathons sur 2 Délégations Régionales** (dépôt de données par les scientifiques après sensibilisation)
- **21 administrateurs de 8 unités** formés (programme en cours)
- **53 inscrits à la liste de diffusion** des administrateurs de DataSuds : dataverse-referents@listes.ird.fr

Tous acteurs, tous responsables

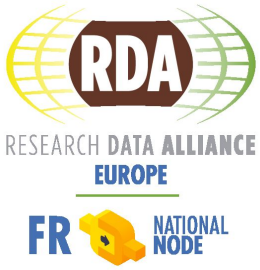
Les décisions sont déléguées au plus près de la recherche.

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
Nature des données (format, représentation, étendue)	Hétérogène, hétéroclite	Homogène (étendue, représentation)
Temporalité	Longue	Réduite à un projet, à un dispositif de recherche (Obs, Infrastructure…)
Fonctionnalités utilisateur final	Réduite : Recherche, téléchargement	Spécifiques et pointues (requêtes, analyse, traitements…)
Administration	Avancée : rôles, workflow de publication	Plus frustre
Objectifs	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

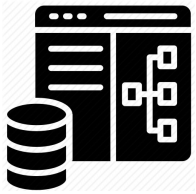
- positionnement des entrepôts de données en fin de cycle de vie
- vision objet de la donnée (fichiers..) rend difficile son exploitation par d'autres plateformes, pour des services à valeur ajoutée (visualisation, fouille de données, traitements, analyse statistiques)
- Réelle réutilisation, valorisation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

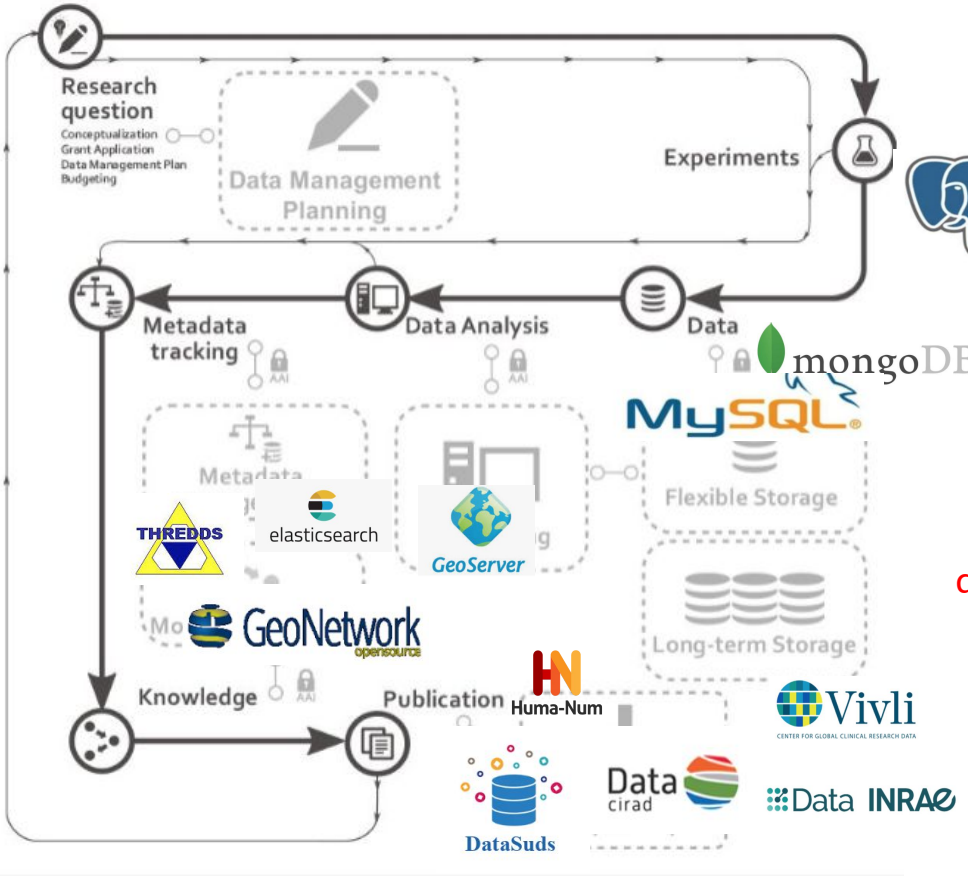
Identification et workflow adaptées à des données statiques

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

Cycle de vie des données dans les projets de recherche

gestion des données ?





dépôt des données ?

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
- Maîtrise 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 fortes 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

- ANR Flash BRIDGE
  - Gouvernance inter instituts
  - Fairisation des entrepôts de données

- Groupe entrepôt de données RDA – France

Jean-Christophe Desconnets (IRD) - pilote du GT

Objectifs

Évaluer la faisabilité d'un service générique d'accueil et de diffusion de données simples et aboutir à des **résultats activables** pour préciser concrètement le périmètre, les solutions techniques, organisationnelles et leurs coûts à travers différents scenarii ;

Proposer des scenarii de mise en œuvre, portant sur des solutions techniques, l'accompagnement, la modération, la curation des données, la gouvernance et le déploiement du dispositif.

Composition du GT

Équipe de coordination

- **Cécile ARÈNES** (Sorbonne Université), Conservatrice des bibliothèques, Chargée de mission Données de la recherche et Humanités numériques
- **Pascal AVENTURIER** (IRD), Ingénieur de recherche, Responsable IST
- **Michel BAMOUNI** (INRAE), Ingénieur d'études, Architecte Logiciel / expert technique
- **Nicolas CAZENAVE** (CINES), Ingénieur de recherche, Coordinateur du pôle "Projets Nationaux et Innovation"
- **Paolo LAI** (INIST, CNRS), Ingénieur de recherche, Responsable du département Valoriser les données de recherche
- **Dimitri SZABO** (INRAE), Ingénieur de recherche, Chef de projet en ingénierie des systèmes d'information

Les membres de l'équipe sont accompagnés par le **cabinet Datactivist**.

Membres du groupe

- **Cécile CALLOU** (INEE, CNRS)
- **Joachim DORNBUSCH** (EHESS)
- **Romain FÉRET** (Université de Lille)
- **Sophie GIRAUD** (CEA)
- **Yannick HOARAU** (Université de Strasbourg)
- **Thomas JOUNEAU** (Université de Lorraine)
- **Anne LAURENT** (Université de Montpellier)
- **Elise LEHOUX** (Université de Paris)
- **Violaine LOUVET** (CNRS)
- **Gilles OHANESSIAN** (CNRS)
- **Laurent ROMARY** (INRIA)
- **Julien SICOT** (Université de Rennes)

Volets de l'étude (1/2)

1 EXPRESSION DES BESOINS ET DES CONTRAINTES

Objectifs :

- analyser des retours d'expérience et rencontre des différents acteurs français et européens pour proposer le périmètre le plus pertinent
- identifier les besoins fonctionnels.

Contenu du livrable:

- périmètre du dispositif
- synthèse des besoins fonctionnels
- cahier des charges pour les fonctionnalités primaires et additionnelles du service

2 BANC DE TESTS LOGICIELS

Objectifs :

- analyser les offres logicielles et assurer les benchmarks pour argumenter les choix
- analyser et proposer les ergonomies et les services à valeur ajoutée indispensable à l'appropriation en lien avec les besoins fonctionnels identifiés

Contenu du livrable:

- synthèse des benchmarks sur les outils
- préconisations pour le déploiement au regard des infrastructures numériques possibles aux échelles régionales et nationales
- (optionnel) pré-requis et actions à envisager pour la certification Core Trust Seal de l'infrastructure de données associée au service.

3 PROPOSITION DE SCÉNARII DE MISE EN ŒUVRE

Objectifs :

- étudier les différents degrés de mutualisation et leur impact sur l'appropriation, la faisabilité technique au regard des contraintes d'interopérabilité, d'architecture,
- mesurer les coûts en s'appuyant sur un nombre restreint de scénarii,
- évaluer le besoin et les coûts en matière de déploiement

Contenu du livrable:

- description des différents scénarii de mutualisation
- description des coûts des différents scénarii retenus

Volets de l'étude (2/2)

4 GOUVERNANCE

Objectifs :

- définir les rôles et responsabilités du dispositif, présenter des éléments sur le niveau et les procédures de curation
- proposer une démarche réfléchie et progressive pour mettre en place un entrepôt de qualité qui remporte l'adhésion

Contenu du livrable:

- tableau des rôles et des responsabilités, schéma des différentes couches, cartographie, schéma de workflow sur le cycle de vie et disposition des acteurs dans les différentes étapes
- description de la démarche et de la montée en charge progressive pour aller vers un service adopté donnant accès à des données de qualité et des fonctionnalités à valeur ajoutée

5 COMMUNICATION

Objectifs :

- prise de connaissance et partage des travaux vers la communauté de gestion des données, les établissements

Contenu du livrable:

- mise en forme de l'étude pour leur diffusion et publication sur ouvrirlascience.fr , diffusion des rapports du GT CoSO, présentation aux JNSO, diffusion à l'échelle internationale (traduction EN)