



The **Dataverse**[®] Project pour les entrepôts de données

Animé par

- Pascal Aventurier, IRD (pascal.aventurier@ird.fr)
- Pascal Bonnet, CIRAD (pascal.bonnet@cirad.fr)
- Jean-Christophe Desconnets (jean-christophe.desconnets@ird.fr)
- Esther Dzale-Yeumo, INRA (esther.dzale-yeumo@inra.fr)
- Dimitri Szabo, INRA (dimitri.szabo@inra.fr)
- Geneviève Michaud, Science Po (genevieve.michaud@sciencespo.fr)

Motivations

Partager nos expériences en matière de gestion des entrepôts de données Dataverse (*à minima*)

Initier une communauté de pratiques des gestionnaires d'entrepôts autour de Dataverse ou à au niveau des entrepôts pour :

- **Faciliter la création de nouveaux projets dataverse** sur la base des expériences et ressources mis à disposition par les autres instituts
- **Mise en réseau** pour mutualiser les compétences et les développements autour de Dataverse

Objectifs de l'atelier

- Présentation de l'outil Dataverse
- Partager les retours d'expérience des établissements de recherche INRA, CIRAD, IRD et Science Po
- Contenu et forme d'une communauté Dataverse France

Programme détaillé

Introduction Atelier (Esther Dzale-Yeumo, Jean-Christophe Desconnets)

14h40 - 15h20 présentation des fonctionnalités d'un entrepôt de données et de l'outil Dataverse

- Entrepôt de données : définitions (Jean-Christophe Desconnets)
- Présentation de l'outil DataVerse (Dimitri Szabo)

15h20 - 16h20 Retour d'expériences de la mise en place de Dataverse

- À l'INRA (Esther Dzale-Yeumo)
- Au CIRAD (Pascal Bonnet)
- À l'IRD (Pascal Aventurier)
- À Science Po (Geneviève Michaud)

16h20-16h40 : Pause Café

16h40-17h30 Discussion ouverte: Création d'un groupe RDA-France Dataverse, pourquoi faire ?

Fin atelier 17h30

Entrepôt de données, de quoi parle t'on ?

définitions et questions

Quelle donnée ?

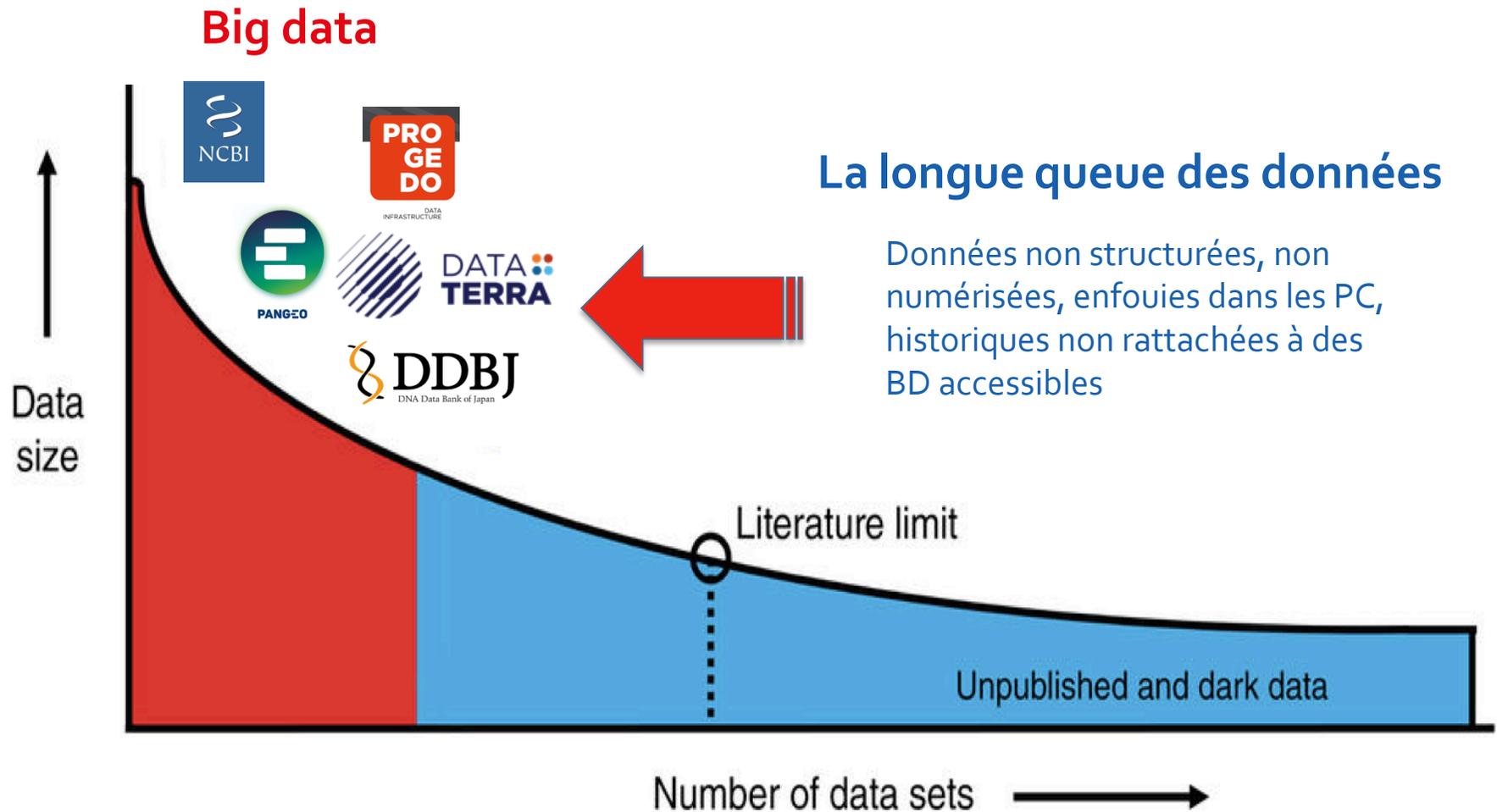
«**Données de la recherche** [...] enregistrements factuels (chiffres, textes, images et sons), [...] sources principales pour la recherche scientifique [...] reconnus par la communauté scientifique comme nécessaires pour valider des résultats de recherche. [...] » (OCDE, 2007)



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



Quelle données cibles pour les entrepôts ?



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

A repository is an online database service, an archive that manages the long-term storage and preservation of digital resources and provides a catalogue for discovery and access.

<https://www.sheffield.ac.uk/library/rdm/repositories>

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

: Fonctionnalités

Dépôt et conservation



- Stockage, conservation
- Organisation de l'entrepôt (thématique institution, structuration en collection de données)
- Dépôt par les utilisateurs autorisés ou programme

Description



- Attribution d'un identifiant
- Description avec des métadonnées standards contrôlées

Publication



- Authentification
- Workflow de publication
- Vérification du dépôt
- Attribution d'une licence d'utilisation
- Plusieurs niveaux de publication (ouvert, restreint, embargo....)

Valorisation via l'entrepôt



- Recherche par facettes (thématique, localisation, auteur...)
- Derniers jeux de données saisies
- Statistiques de consultation
- Citation du jeu de données
- Export des métadonnées

Echanges/Analyse



- Scannées par des catalogues de données internationaux
- interconnexion automatique avec d'autres entrepôts
- Import/export automatique via programme (API)
- Visualisation des données

Entrepôt de données vs Portail de données

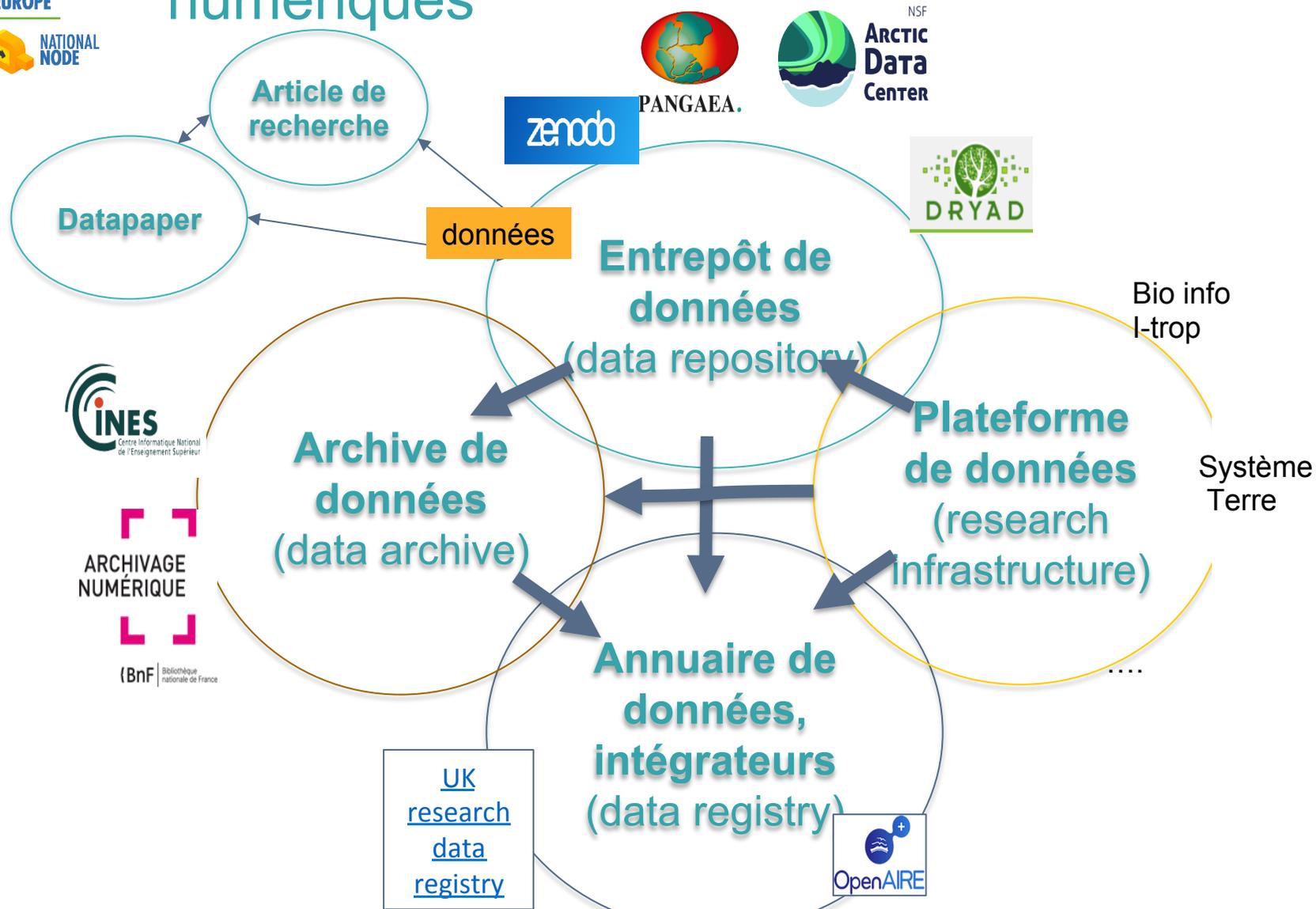
	Entrepôt de données	Portail de données
Nature des données (format, représentation, étendue)	Hétérogène, hétéroclite	Homogène (étendue, représentation)
Temporalité	longue	Réduite à un projet, à un dispositif de recherche (Obs, Infrastructure...)
Fonctionnalités utilisateur final	Réduite : Recherche, téléchargement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



RESEARCH DATA ALLIANCE EUROPE



Les entrepôts en lien avec les autres dispositifs numériques



D'après Cocard et Aventurier 2017



Entrepôt de données

Avant tout, un dispositif numérique pour

- collecter et structurer des métadonnées
- Préserver
- Partager
- Valoriser

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