Open data, enjeux actuels et futurs

Laurent Burnel

MQDS 2023
Measurement - Quality - Data Science

INRAE

Open data, enjeux actuels et futurs

Illustration par l'ouverture des données de la recherche publique française

Bordeaux – Laurent Burnel - 6 juin 2023

Plan

- Définition
- Historique
- Enjeux actuels et futurs
- Illustration à l'Institut National de Recherche pour l'Agriculture, l'Alimentation et l'Environnement (INRAE)
- Perspectives

Définition
Open science / Open data

Un mouvement inclusif

Objectif de rendre la science plus ouverte, accessible, efficace, démocratique, transparente

Une facette

Mouvement de mise à disposition des données publiques
Dont données de la recherche

Historique
Communiquer la science au fil du temps

SCIENCE « FERMEE »

Entre soi scientifique

Avant 17e siècle
Echanges interpersonnels, Correspondances

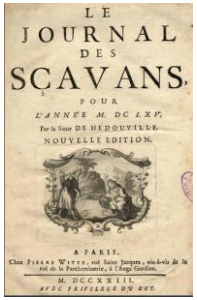
17éme - 18e siècles
Sociétés savantes, discussions en assemblée, comptes-rendus et mémoires

SCIENCE « OUVERTE »

Volonté sociale, scientifique et politique
Sous l'effet du numérique (web, big data)

Années 1990	Années 2000	Années 2010	Années 2020
NAISSANCE	JEUNESSE	ADULTE	MATURITE ?

1665 — 1850 — **1990** — **1995** — 2016 — 2021



1ére revue scientifique

1000 revues scientifiques

OPEN SCIENCE

Open Access (publications)

OPEN DATA

120 000 revues scientifiques

4 éditeurs pour 40 % du marché de l'édition scientifique

La science ouverte, une utopie qui gagne du terrain

- 2 Plans nationaux pour la science ouverte
  2018-2021 et 2021-2024 (France)

Années 2020 : la maturité ?

- Mars 2016 : Principes FAIR (Europe)
- Avril 2016 : RGPD CEE
- Octobre 2016 : la loi pour une République numérique

Années 2010 : l'entrée dans l'âge adulte

- Janvier 2004 : L'OCDE dit oui à l'accès ouvert aux données
- Janvier 2003 : Déclaration de Berlin (Allemagne)
  Qui élargit la notion du libre accès aux données de recherche

Années 2000 : la fougue de la jeunesse

- 2000 : Accès libre droit d'auteur, licence CC
  Aaron Swartz (USA)

Années 1990 : une naissance pleine de promesses

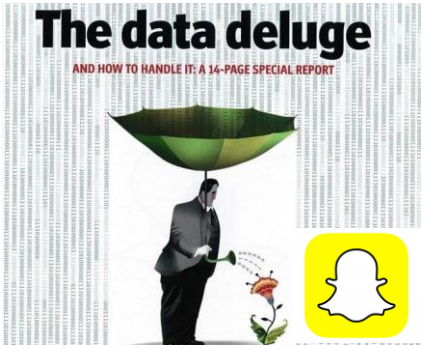
- Janvier 1991 :
  Archive ouverte arXiv
  Paul Ginsparg (USA)

Enjeux globaux
Dans un monde qui change, qui s'accélère

Complexité des défis mondiaux

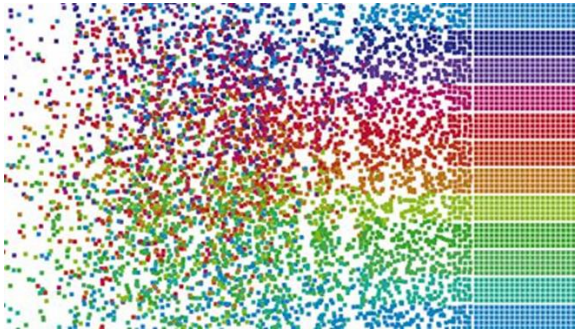
Disruption numérique

Déluge de données



Besoin de structurer des données hétérogènes

Open data
une clé pour aborder
cette complexité ?



Bénéfices :

- Société
- Economie
- Recherche

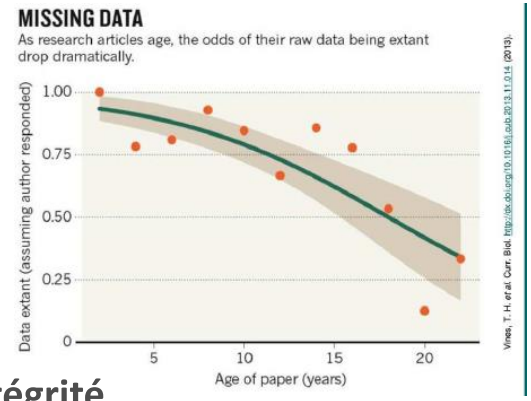
« Le pouvoir sera à ceux qui sauront traiter les données »

Données de la recherche :
Un vrai bien commun?

Enjeux scientifiques
Défis de l'Open data dans la recherche publique française

Obsolescence actuelle des données

20 ans après publication,
80% des données seraient perdues !




Le Monde S'abonner

SCIENCES • BIOLOGIE

Deux semaines de mise à pied pour l'ex-présidente du CNRS

Mise en cause pour des manipulations de données, la biologiste Anne Peyroche admet des fautes, mais pas de fraude. Son employeur, le CEA, a mis en ligne l'essentiel des documents qui ont motivé sa sanction.

Par Hervé Morin • Publié hier à 13h56, mis à jour à 05h55

Remise en question de l'intégrité et de la transparence du processus scientifique

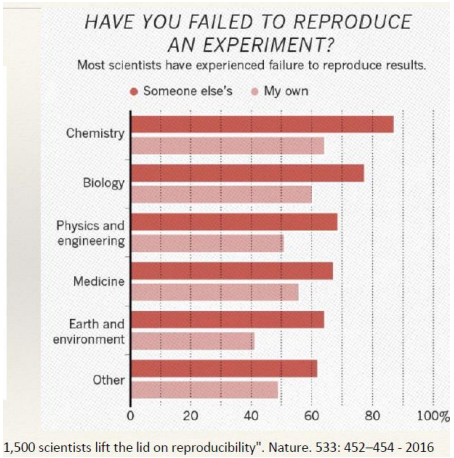
54 % des données utilisés par les scientifiques sont encore « invérifiables »

Source:
https://archive.wikiwix.com/cache/index2.php?url=https%3A%2F%2Fprojects.ac%2Fblog%2Fwp-content%2Fuploads%2F2014%2F01%2FLove-your-data_Projects_s.jpg

Crise de la reproductibilité

Plus de 70% des chercheurs ont échoué à reproduire l'expérience d'un collègue ; et plus de 50% n'ont pas réussi à reproduire leur propre expérience…

Source : https://doi.org/10.1038/533452a



1,500 scientists lift the lid on reproducibility". Nature. 533: 452–454 - 2016

Enjeux scientifiques
Open data dans la recherche publique française

Le cadre est à présent posé

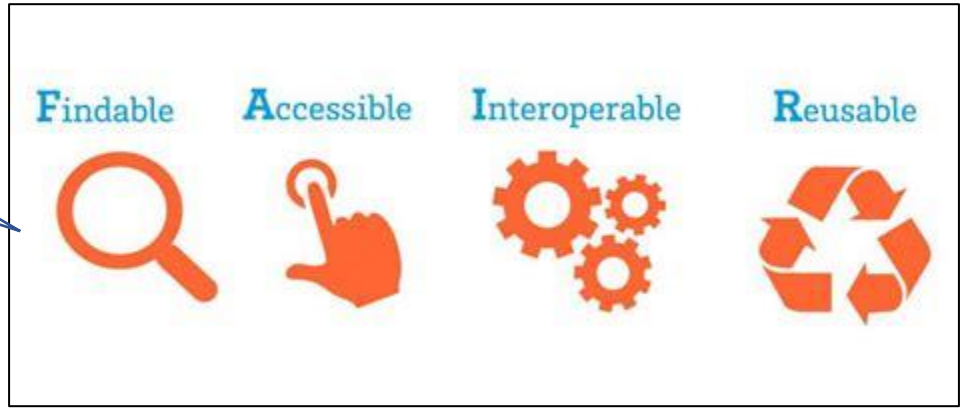
Les données de la recherche doivent être

«ouvertes autant que possible,

aussi fermées que nécessaire»

Une Directive européenne du 20 juin 2019

Selon des principes FAIR pour nous guider



Objectifs : Réutilisation des données et Reproductibilité de la science

Enjeux actuels de l'open data
2 focus : les métadonnées

Données sans métadonnées =

Enjeux actuels de l'open data
2 focus : les principes FAIR

Utilisation d'identificateur unique et pérenne (DOI)

Fourniture de métadonnées persistantes

Indexation dans des catalogues

Stockage durable

Facilitation de l'accès et du téléchargement

Utilisation de licences adéquates

Sémantique (référentiels) et syntaxique (codes, formats)

Usage de normes et standards

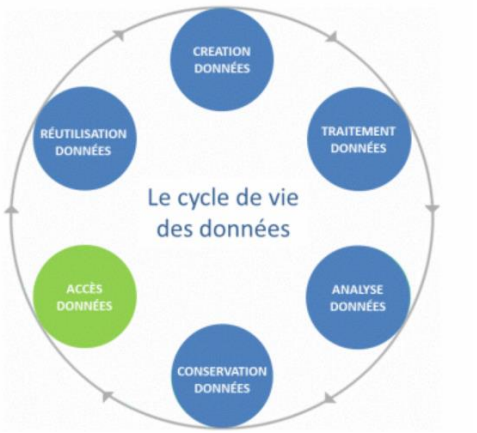
Résultante de **F-A-I** :

Licence claire et accessible,

Métadonnées riches (dont provenance)

Enjeux actuels

Réussir l'ouverture des données comme l'ouverture des publications scientifiques



DMP OPIDoR

Plan de gestion de données

GeoFlow Editor

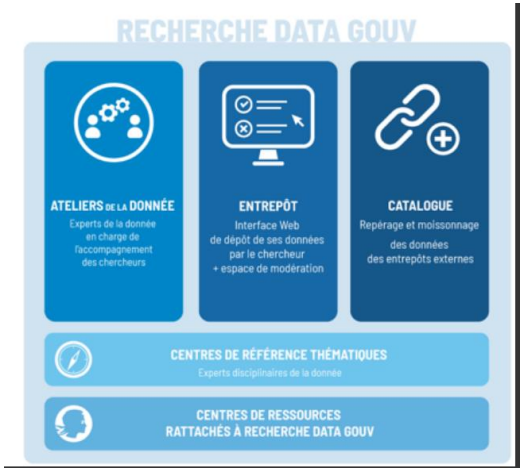
Editeur de métadonnées

Les outils sont là !

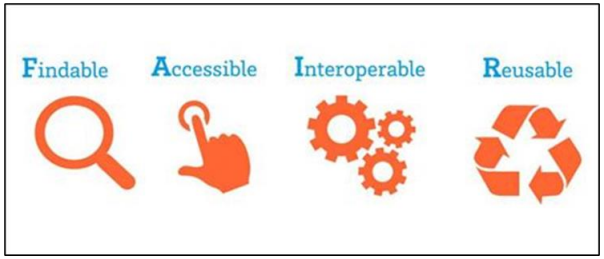
Software Heritage

GitLab

Entrepôt de codes, logiciels, algorithmes

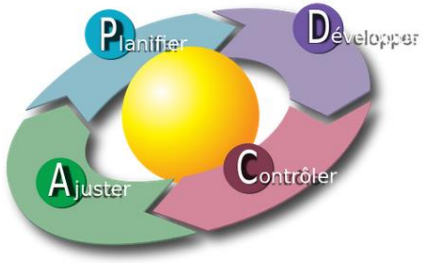


Plateforme nationale fédérée des données de la recherche



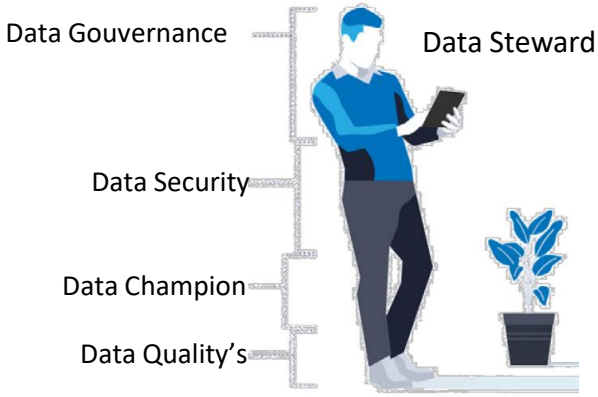
Venez les découvrir à notre atelier cet après midi !

Perspectives
Ce qu'il reste encore à améliorer



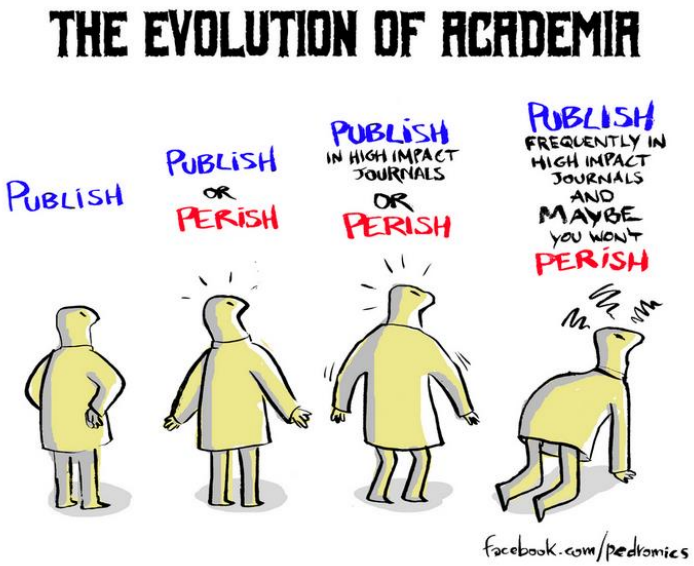
Donner confiance aux données réutilisables : mise en qualité FAIR et fiabilité des données

→ apport nécessaire des démarches qualité, de la métrologie pour les données mesurées



Accompagner les chercheurs

→ Vers un changement plus culturel que technologique



Un besoin de revoir les critères d'évaluation de la recherche

→ De nouveaux indicateurs et fin du « publish or perish »

Au final :

Accélérer le partage pour plus de connaissance

pour répondre aux enjeux globaux

Conclusion

- Les outils sont là !
- Besoin d'acculturation des chercheurs
- Besoin encore d'encadrer la mise à disposition des données
- Accélérer le partage pour plus de connaissance

pour répondre aux enjeux globaux

MQDS 2023
Measurement - Quality - Data Science

INRAE

Merci pour votre attention