



HAL
open science

DONNER Une nécessaire transformation des pratiques

Isabelle Charpentier

► **To cite this version:**

| Isabelle Charpentier. DONNER Une nécessaire transformation des pratiques. 2024. hal-04838254

HAL Id: hal-04838254

<https://hal.science/hal-04838254v1>

Preprint submitted on 14 Dec 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

DONNEES ~~ES~~ ER

Une nécessaire transformation des pratiques

I. Charpentier

1 CONTEXTE

- Innovation

2 INFRASTRUCTURE DE DONNEES

- Outils (Entrepôt, Catalogue)
- Cheminement des données et des métadonnées
- Bilan

3 PROPOSITIONS

- 1. Recommandations pour le dépôt des données
- 2. Aller à la pêche grâce à Mondrian
- 3. Comprendre les freins à la science ouverte
- 4. Améliorer le partage de données ouvertes
- 5. Partager avec les acteurs territoriaux de la donnée
- 6. Objets frontières digitaux, Data viz et IA
- 7. Retours d'expérience
- 8. Les données dans Transform
- 9. Gouvernance
- 10. Mise en œuvre à toutes les échelles

Novateurs : À l'affût du changement et partagent volontiers leur expérience avec d'autres.

Premiers adoptants : également friands de nouveauté, se laissent très vite convaincre par les novateurs. N'hésite pas à donner leur avis sur la nouveauté.

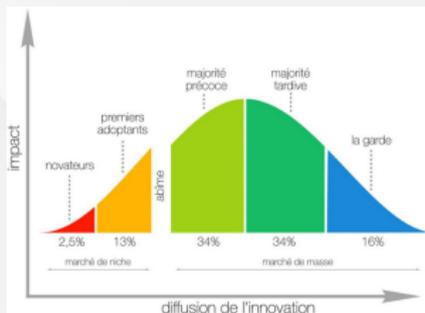


Schéma d'E. Rogers

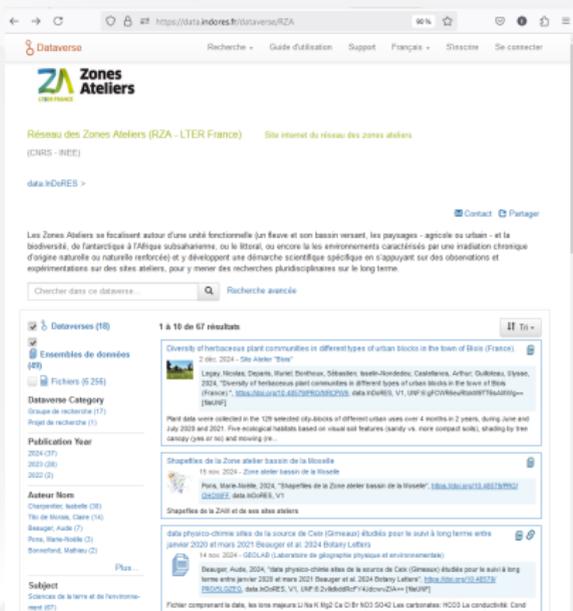
Majorité précoce : constituée de « clients » plus réfléchis qui attendent les retours d'expérience avant de franchir le pas.

Majorité tardive : attend qu'une large part de la population donne des preuves de l'intérêt du changement.

La garde : constituée par des personnes qui perçoivent avant tout le changement comme une menace. Avantages de la nouveauté doivent être très largement supérieurs.

Réflexions possibles

- acceptabilité du partage de la donnée
- appropriation des technologies...

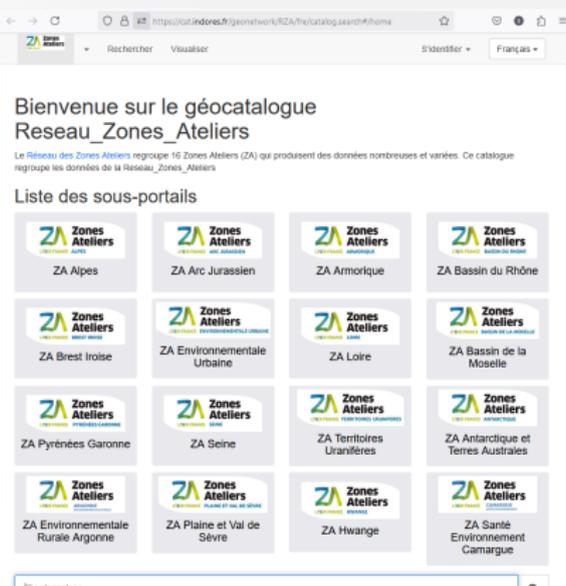


The screenshot shows the DataVerse website interface. At the top, there's a navigation bar with 'Recherche', 'Guide d'utilisation', 'Support', 'Français', 'S'inscrire', and 'Se connecter'. Below this is the 'Zones Ateliers' logo and the text 'Réseau des Zones Ateliers (RZA - LTER France)'. A search bar is present with the text 'Rechercher dans ce dataverse...'. The main content area displays search results for 'DataVerse (10)', listing various datasets with titles, dates, and brief descriptions. The results include:

- Diversity of herbaceous plant communities in different types of urban blocks in the town of Béziers (France)** (2 oct. 2024 - 06 Ateliers "Zéro")
- Plant data were collected in the 129 selected city-blocks of different urban uses over 4 months in 2 years, during June and July 2020 and 2021. Five ecological habitats based on visual soil features (sandy vs. more compact soils), shading by tree canopy (yes or no) and stoniness (y/n).**
- Shapettes de la Zone atelier bassin de la Micauda** (15 nov. 2024 - Zone atelier bassin de la Micauda)
- Pons, Marie-Noëlle, 2024, "Shapettes de la Zone atelier bassin de la Micauda".**
- Shapettes de la ZAH et de ses sites ateliers**
- data physico-chimie sites de la source de Cais (Girmeaux) étudiés pour le suivi à long terme entre janvier 2020 et mars 2021** (Swaguer et al. 2024) (Dutary Lefebvre)
- 14 nov. 2024 - GEGALAB: Laboratoire de géographie physique et environnementale**
- Swaguer, Aurélie, 2024, "data physico-chimie sites de la source de Cais (Girmeaux) étudiés pour le suivi à long terme entre janvier 2020 et mars 2021" (Swaguer et al. 2024) (Dutary Lefebvre)".**
- Fichier comprenant le data, les liens majeurs LiNAK (R2) De C2B: N03 S04C Les carbonates, H023 Les conductivité. Contient les données des sondages effectués dans le site de C2B: N03 S04C Les carbonates, H023 Les conductivité. Contient les données des sondages effectués dans le site de C2B: N03 S04C Les carbonates, H023 Les conductivité.**

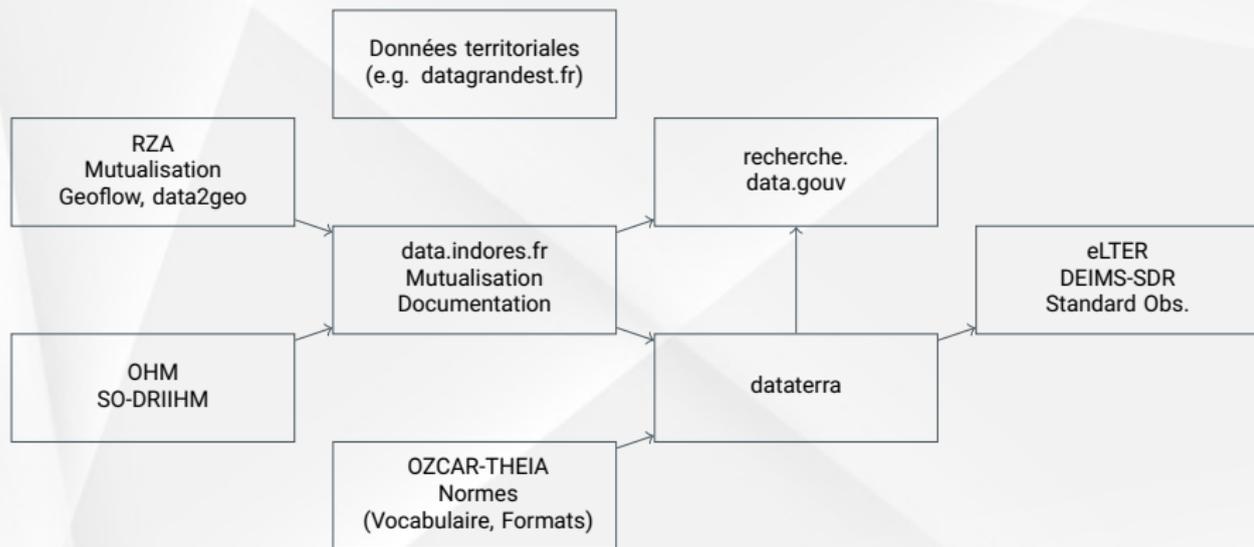
Data.indores.fr: Entrepôt thématique de type DataVerse ouvert pour les données socio-écologiques. Une ZA sans entrepôt universitaire local ou sans entrepôt thématique habituel (e.g. entrepôt d'OSU) peut y déposer ses **données (et métadonnées associées)**.

Toutes les ZA ont un sous-Dataverse dans data.indores.fr



Cat.indores.fr: Géocatalogue de type GéoNetwork pour:

- les métadonnées socio-écologiques accompagnées de données géographiques (principalement géomatiques),
- les métadonnées des jeux de données déposés ailleurs (lien vers les données du RZA, reporting...)



Données et/ou métadonnées circulent de gauche à droite (local, "régional/thématique", national, européen).

Elles ne sont déposées qu'une fois, en un seul endroit.

Les métadonnées sont moissonnées par l'infra de niveau supérieur.

Les entrepôts locaux ne figurent pas.

Certaines interactions sont encore à préciser, notamment avec acteurs territoriaux de la donnée.

	ZA A	ZA AJ	ZA AR	ZA BR	ZA BRI	ZA RG	ZA CAM	ZA EU	ZA HG	ZA L	ZA M	ZA PVS	ZA PYG.	ZA S	ZA TA	ZA TU
Dataverse	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Sous-DV										1+		5		1		
ZA-Shape								1		(1)		(1)				
Frise*	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
ModZA*	1		1		1			1	1		1	1	1			(1)
Autres								3	2	1	1			4		
Liés			3				1			7				1		
RDG	9															

data2geo pour les fiches de métadonnées / * Dans l'ouvrage / 3 jeux déposés à l'échelle du réseau

Questionnements

1. Peu de jeu de données sur data.indores, ça démarre dans certaines ZA
2. Sont ils sur d'autres entrepôts académiques, dans des bases de données, EasyData, RDG, Zenodo... ?
3. Sont ils "retenus" par manque de temps, par besoin de facilitation, pour d'autres raisons inavouables ?
4. Que faire pour améliorer l'infrastructure et la collecte ?
5. "Un Article = Un jeu de données ?"

Bilan catalogue: Moissonnage incorrect car catalogues locaux moissonnés entièrement, mauvaises affectations des fiches de l'ancien catalogue. **à reprendre**

Le savoir est la seule richesse que l'on puisse entièrement dépenser sans en rien la diminuer.

(Amadou Hampaté Ba)

Urbi et Orbi

L'infra est là, il est temps de diffuser et valoriser nos données à l'extérieur

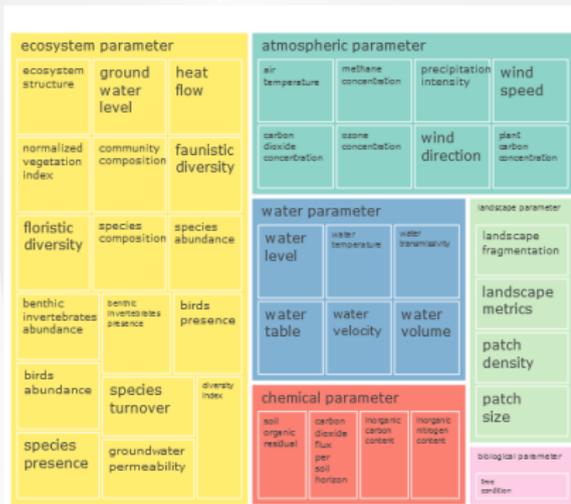
1. Recommandations pour le dépôt des données

- Entrepôts à utiliser par ordre de préférence ?
 1. **Entrepôts locaux ou thématiques.** Faire remonter une fiche de métadonnées sur le catalogue du RZA (voir en-dessous)
 2. **data.indores:** pour les données socio-éco (RZA ou OHM, labos CNRS-EE). Import des métadonnées dans votre géocatalogue par le RZA. Un sous-entrepôt existe pour chaque ZA (idem OHM)
 3. **RDG: Plébisciter par de plus en plus d'universités.** Signaler le dépôt pour que le RZA importe les métadonnées dans votre géocatalogue
 - X. **Zenodo** Entrepôt non souverain, à éviter. Privilégier l'une des trois solutions précédentes
- Catalogues locaux
 4. Initialement conçus à destination des géomaticiens, pour le dépôt de données géographiques, utilisable pour la description de projets de recherche, éventuellement pour signaler des jeu de données que l'on ne veut pas vraiment ouvrir
 - Moissonnables par d'autres catalogues, utiles pour faire du reporting. Le catalogue du RZA moissonne les catalogues locaux (OSU par exemple). Un sous-catalogue existe pour chaque ZA (et OHM) dans cat.indores.fr. Le moissonnage n'est pas correct, trop de fiches des catalogues locaux, mauvaises affectations des fiches de l'ancien catalogue

Identifier les pratiques et recommandations pour nourrir le PGD scientifique

Code couleur: **Recommandé**, **Bonnes pratiques** ; **Attention, Effort supplémentaire** ; **Peut mieux faire**

2. Aller à la pêche grâce à Mondrian



"Mondrian" élaboré par DEIMS-SDR sur informations recueillies auprès des ZA.

Comment exploiter cela ?
Questionnements utiles ?

- Les données sont-elles déposées ? Si oui, où ?
- Les données sont-elles disponibles ? Si oui, quels sont les freins au dépôt ?
- (Avez vous fait la mise à jour sur DEIMS-SDR ?)

Les réponses sont autant d'éléments pour nourrir un PGD scientifique

Données analysables à l'échelle du réseau :

1. les infos ont elles été vérifiées sur DEIMS ?
2. les infos sont-elles confirmées par le dépôt de données (quelque part) ?
3. ce jeu de données est-il référencé dans un catalogue ?

3. Comprendre les freins à la science ouverte

Partage des données ? Acceptation du partage ? Acceptabilité du partage ?

Comprendre de manière pro-active

Rejoindre l'initiative d'enquêtes sur les données de la recherche

<https://gtso.couperin.org/gtdonnees/recensement-enquetes...>

Le GTSO Données recense les enquêtes sur les données de la recherche menées au sein d'établissements d'enseignement supérieur. Tout établissement peut contribuer et ajouter sa propre enquête en respectant les consignes suivantes :

- 1200 caractères espaces compris maximum ;
- Quelques phrases introductives ;
- Des tirets qui listent les principales conclusions,
- Éventuellement une phrase d'ouverture sur les perspectives ;
- Le lien vers l'enquête en ligne.

La contribution est à envoyer à gtso_donnees@listes.couperin.org

4. Améliorer le partage de données ouvertes

Analyser, e.g. Woods, H. B. and Pinfield, S. (2021). Incentivising research data sharing: a scoping review. Wellcome Open Research, 6.

À retenir : Plusieurs méthodes et pratiques peuvent améliorer le partage de données ouvertes :

- stratégie qui s'aligne sur les normes existantes : **INSPIRE pour les géomaticiens, usage des thésaurus**
- stratégie qui s'adapte aux besoins spécifiques de diverses communautés de recherche : **voir propositions 1 et 2, à définir dans PGD scientifique**
- communication efficace sur les politiques et services de partage de données **BBEES pour la doc, RZA, ZA...**
- présence de modèles illustrant de bonnes pratiques de partage de données **premiers adoptants, exemples, propositions 6 et 7**
- allocation de ressources suffisantes : **INFRA OK, RH dans le cadre OHM-RZA**

Certes, mais on devrait aller plus loin.

**Sortir du cadre : Prendre le problème à l'envers en partant de la Réutilisation ?
Evaluer le potentiel ? Valoriser sur les territoires ?**

Thème 2024 : Les données au service de la territorialisation de la transition écologique



J'ai aimé :

- la COP territoriale
- les retours d'expériences dont le co-design de l'observatoire des forêts (28° Design et ONF)

Données, cartes, indicateurs sont clés pour les acteurs, friands de visualisation et de médiation... autant d'artefacts digitaux qualifiables d'objets-frontière !

Comment rendre visibles nos données sur les portails territoriaux ?

> à traiter dans un PGD scientifique

> le grand test et Grand Est : ZAEU, ZAM, ZARG, OHM Pays de Bitche, OHM Fesseheim & Data Grand Est

Quels projets de recherche autour de la donnée territoriale ?

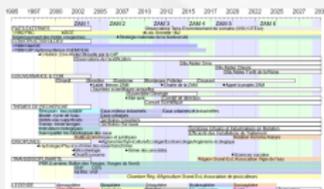
Quelle valorisation ?

Est-on déjà dans l'ingénierie avec champ libre aux bureaux d'études ?

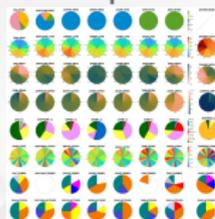
Regrouper sur data.indores un ensemble de données communes à toutes les ZA (utilisées dans le bouquin, ou bien déjà collectées par eLTER-DEIMS)



Shapefiles
ZAEU, ZAHG, ZAM



Frises ZA
ZAS



Indicateurs ZAA, ZAAJ
ZAAR, ZABRI, ZAEU, ZAHG
ZAM, ZAPYGAR, ZAPVS



Mondrian (DEIMS)
ZAA, ZACAM, ZARG

Construction une collection numérique d'ensembles de données et de logiciels liés aux objets limites socio-écologiques ?

Représentation digitale des systèmes socio-écologiques (schémas conceptuels, systèmes ARDI, ou autres)

- Possibilité des médiations de type Data Viz
- Digitalisation propice à des recherches avec IA

7. Retours d'expérience : Structurer son dataverse

ZA
Référénts
Sous-Collections

ZAL
Nicolas

- plateformes
- sites

ZAPVS
Sabrina

- plantes adventices / weeds
- oiseaux / birds
- pollinisateurs / pollinators
- rendement / yield
- général / general

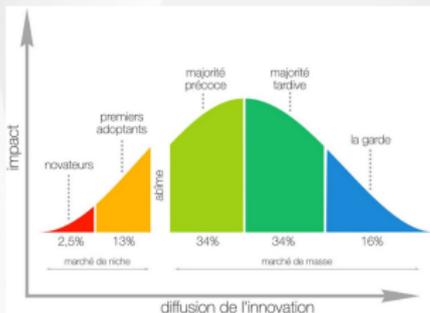
8. Les données dans Transform

Sur la base de ce qui a été proposé dans cette présentation :

- en impliquant la gestion des données très tôt dans le processus de recherche
- en utilisant les infrastructures de données existantes

Un vrai sujet pour évaluer la transformation des pratiques autour de la donnée dans la recherche socio-écologique

Franchir l'abîme en identifiant les personnes clés (dans ZA et/ou labos)



Novateurs : À l'affut du changement et partagent volontiers leur expérience avec d'autres.

Premiers adoptants :

Binôme "Chercheur-Référent" : Le référent accompagne, le chercheur dépose et partage son expérience.

Majorité précoce : Organiser des retours d'expérience sur la donnée. Le référent accompagne d'autres chercheurs.

Réflexions possibles

- Faire une place aux données dans la gouvernance des ZA ?
- Penser la donnée (sa collecte, ses usages, sa publicisation...) en amont des projets de recherche ?

10. Mise en œuvre à toutes les échelles

