



HAL
open science

Sobriété numérique et données de recherche : place à la pratique !

Alicia León y Barella

► **To cite this version:**

Alicia León y Barella. Sobriété numérique et données de recherche : place à la pratique!. Printemps de la Donnée 2024, INRAE; Université Haute-Alsace; Université de Strasbourg; INSA; PNDB; AgroParisTech; Université de Lille; Sorbonne Université; Data Terra, Jun 2024, Lille, France. hal-04708350

HAL Id: hal-04708350

<https://hal.inrae.fr/hal-04708350v1>

Submitted on 24 Sep 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.



Distributed under a Creative Commons Attribution 4.0 International License



la Fabrique
de la science
ouverte

Sobriété numérique et données de la recherche : place à la pratique !

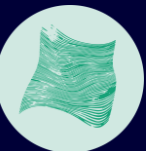
24 juin 2024

<https://fabso.univ-lille.fr>

28/08/2024

Sobriété numérique, Science ouverte et données de recherche

Alicia León y Barella
Cheffe du service Science ouverte –
SCD Université de Lille



Sobriété numérique, Science ouverte et données de recherche

À l'université de Lille

La gestion des données à l'université de Lille

- Enjeu intégré à la *Feuille de route pour la Science ouverte* (juillet 2021)



Selon le principe « aussi ouvert que possible, aussi fermé que nécessaire », rendre dans la mesure du possible accessibles les données scientifiques produites par la communauté scientifique lilloise, dans le respect des principes FAIR (Faciles à trouver, Accessibles, Interopérables, Réutilisables) et développer une culture de la gestion et du partage des données.

ACTION 7 Former et sensibiliser les doctorants, chercheurs et personnels d'appui, et promouvoir les pratiques vertueuses en matière de données (ouvertes),

Si la connaissance en matière d'accès ouvert aux publications progresse aujourd'hui au sein des communautés de recherche, il n'en est pas encore de même pour ce qui est des données. Il y a nécessité à développer, dans les années qui viennent, une politique forte de sensibilisation et de formation, notamment au regard du coût particulièrement élevé, tant sur le plan financier que sur le plan scientifique, d'une mauvaise gestion des données produites. L'Université de Lille amplifiera son action formative, notamment à travers le développement et le renforcement de son dispositif de sensibilisation « Fabrique de la Science Ouverte ». Elle interviendra tant au niveau individuel qu'au niveau collectif – équipes / unités de recherche – pour soutenir le développement des compétences nécessaires en matière de données, y compris dans leurs dimensions juridiques, éthiques ou encore d'intégrité scientifique.

ACTION 8 Accompagner les porteurs de projets dans l'élaboration du plan de gestion de données et plus généralement pour ce qui concerne les attendus en matière de Science Ouverte.

Les appels à projets, qu'ils soient internes, nationaux ou européens, imposent de plus en plus systématiquement la rédaction puis la mise en œuvre d'un plan de gestion de données. Au-delà de l'obligation réglementaire, la formalisation d'un tel plan permet de mettre en place une démarche raisonnée dans l'usage, l'utilisation, la conservation et l'ouverture éventuelle des données produites. L'Université de Lille intensifiera ses efforts pour former les porteurs de projets aux enjeux liés aux plans de gestion de données et les accompagner dans l'écriture de ces derniers.

ACTION 9 Accompagner les lauréats d'appels à projets dans la mise en œuvre de leur plan de gestion de données, et dans la gestion du cycle de vie des données produites.

La gestion du cycle de vie des données – création, traitement, analyse, conservation, accès, préservation, réutilisation – est une dimension majeure du projet de recherche. L'Université de Lille développera un accompagnement des laboratoires dans la curation de leurs données, tant pour celles qui ont vocation à être ouvertes que pour celles qui resteront accessibles dans un périmètre restreint.

ACTION 10 Définir et mettre en œuvre une stratégie en matière d'entrepôts de données.

Le partage des données forme un usage répandu mais contrasté dans le monde académique : si certaines disciplines le pratiquent depuis des décennies, d'autres ne disposent pas encore d'infrastructures partagées de stockage, de conservation et d'exposition. Or la question des entrepôts de données est encore immature en France, tant au niveau local qu'au niveau national : rares sont les universités qui ont entamé une réflexion structurée sur le sujet, et l'État lui-même ne s'est saisi du sujet que récemment. Le paysage évolue néanmoins rapidement, avec la mise en œuvre d'une stratégie nationale, en cours de définition, annoncée pour la fin 2021. L'Université de Lille initiera sa stratégie, articulée bien entendu avec la politique nationale, afin de répondre aux besoins nécessairement différents de chaque communauté, notamment celles ne disposant pas d'entrepôts disciplinaires adaptés.

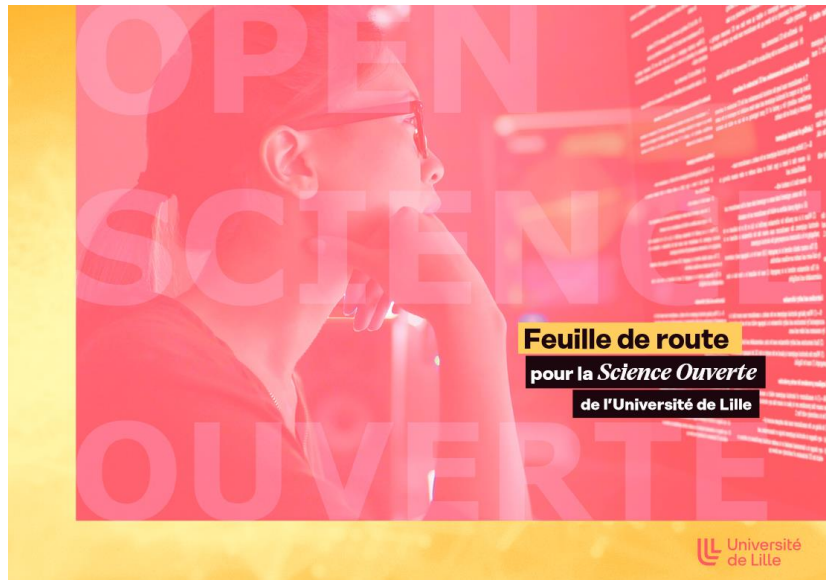
ACTION 11 Accompagner, soutenir et conseiller les chercheurs dans les pratiques émergentes.

Le champ des données voit apparaître depuis quelques années des pratiques nouvelles, qui ont vocation à être soutenues : diffusion des données qui sous-tendent les articles publiés, ou encore *data papers*, ces publications revues par les pairs qui écrivent des jeux de données accessibles, en vue de leur (ré)utilisation à des fins de recherche. Ces pratiques d'ouverture, prometteuses mais émergentes, seront accompagnées par l'université.

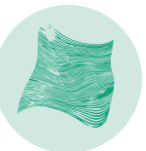
ACTION 12 Promouvoir les principes FAIR et accompagner leur mise en œuvre dans les bases de données produites dans le cadre de l'université.

Une donnée mal référencée, mal qualifiée et mal exposée est une donnée généralement perdue car inconnue et inaccessible à l'essentiel des communautés qui ont vocation à l'utiliser. Avec la massification des données produites tout au long du processus de recherche, il devient critique d'assurer une gestion rigoureuse de l'ensemble du cycle de vie de la donnée. Depuis 2019, l'Université de Lille expérimente une offre de services à destination des chercheurs qui gèrent des bases de données, notamment grâce à un partenariat avec l'Inist-CNRS pour ce qui concerne l'attribution de *Digital Object Identifiers* (DOI). L'université entend à l'avenir accompagner les chercheurs dans l'amélioration de la conformité de leurs bases de données aux principes FAIR (Faciles à trouver, Accessibles, Interopérables et Réutilisables) et ainsi renforcer la visibilité et la réutilisation des données produites, dans le respect des cultures et des pratiques disciplinaires.

La gestion des données à l'université de Lille

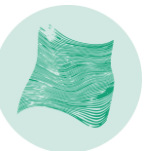


- Enjeu intégré à la *Feuille de route pour la Science ouverte* (juillet 2021)
- Un pilotage opérationnel
 - Service commun de documentation
 - DGD Recherche et valorisation
- Un comité de pilotage animé par le VP Recherche
 - DGD Recherche et valorisation ; SCD
 - Vice-présidents Recherche, Numérique, Valorisation, iSite, SHS, Evaluation



La gestion des données à l'université de Lille

- Un enjeu qui dépasse celui de la science ouverte :
 - Valorisation économique des données
 - Dimension éthique et juridique
 - Données générées dans les zones à régime restrictif
 - etc.
- Nomination à l'Université de Lille d'un **ADAC** en 2022, administrateur des données, des algorithmes et des codes de la recherche
 - Julien Roche, directeur du SCD
 - Définition de trois priorités :
 - La mise en place d'un schéma de gouvernance de la donnée de recherche,
 - La création d'une cellule opérationnelle de gouvernance,
 - La constitution d'un réseau d'acteurs opérationnels.



La gestion des données à l'université de Lille (en pratique)

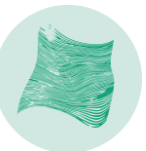
Lille
Open
Research
Data

- Accompagnement des équipes de recherche
 - dans la gestion de leur données
 - dans la mise en application des principes FAIR
 - dans la diffusion des données quand cela est possible
- Labellisation « Atelier de la donnée » en 2022 : Lille Open Research Data (LORD)

La gestion des données à l'université de Lille (en pratique)

bu.univ-lille.fr/chercheurs-doctorants/

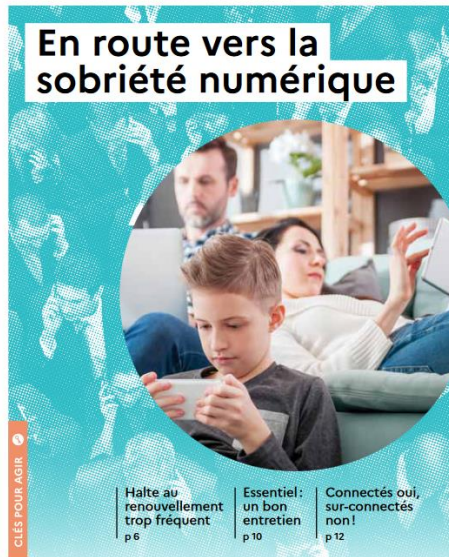
- **Webinaires Science ouverte**, et **accompagnement individualisé** (montage de projets, PGD, données FAIR...)
- **Formation doctorale** (Données, codes et logiciels)
- **Formation continue** des personnels
- **Administration de Lillodata** l'instance locale de l'entrepôt national Recherche Data Gouv
- Animation d'un réseau de **correspondants données**



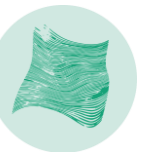
La sobriété numérique aux cœur des débats



INRO

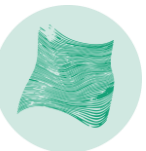


Fabrique de la Science ouverte
Sobriété numérique et données de recherche : place à la pratique !



La sobriété numérique aux cœur des débats, à Lille aussi !

- [Plan de transition écologique de l'université de Lille. Feuille de route 2023-2033, 2023.](#)
- Schéma directeur du numérique (2025-2029), à paraître
 - Un chargé de mission numérique responsable, Philippe Marquet
- Principes de gouvernance des données, algorithmes et codes de la recherche à l'Université de Lille, en cours de validation



Principes de gouvernance des données, algorithmes et codes de la recherche à l'Université de Lille



En cours de validation

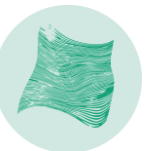
Principe 1 : Les données, les algorithmes et les codes doivent être partagés et réutilisés en respectant les valeurs de la science

Principe 2 : Les données doivent être Faciles à trouver, Accessibles, Interopérables, Réutilisables (FAIR)

Principe 3 : Les données doivent être aussi ouvertes que possible, aussi fermées que nécessaire, décrites dans tous les cas

Principe 4 : Les caractéristiques particulières des codes source et des logiciels doivent être reconnues

Principe 5 : Les enjeux liés à la sobriété numérique doivent être pris en compte

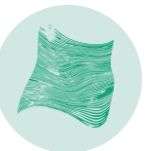




Principe 5 : Les enjeux liés à la sobriété numérique doivent être pris en compte

Le numérique, dont l'ouverture des données, est source d'une empreinte environnementale non négligeable. Pour limiter cet impact et aller vers une plus grande sobriété des données, il devient nécessaire de mettre en place de bonnes pratiques sur l'ensemble du cycle de vie des données :

- évaluer la pertinence à créer de nouvelles données quand d'autres peuvent être réutilisées,
- limiter le stockage de versions obsolètes, restreindre le volume des données quand c'est possible,
- s'interroger sur les modalités de diffusion selon l'usage et l'intérêt des jeux de données,
- choisir des logiciels de traitement des données open source, ne développer des logiciels que si nécessaire, chercher à réduire l'impact du code...

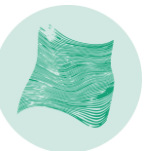


Sobriété numérique, Science ouverte et données de recherche

Pistes d'action et bonnes pratiques

Science ouverte et Sobriété numérique : des enjeux contradictaires ?

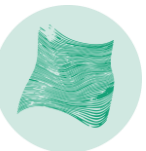
- **Sobriété numérique** : réduire l'impact environnemental des activités numériques
- **Science ouverte** : mouvement pour l'ouverture, la diffusion et le partage de ses recherches
 - Donner un accès pérenne aux productions scientifiques : publications, données, codes et logiciels, et tout le reste !
- **Données de recherche** : un enjeu particulier => une bonne gestion avant tout (une gestion sobre?)



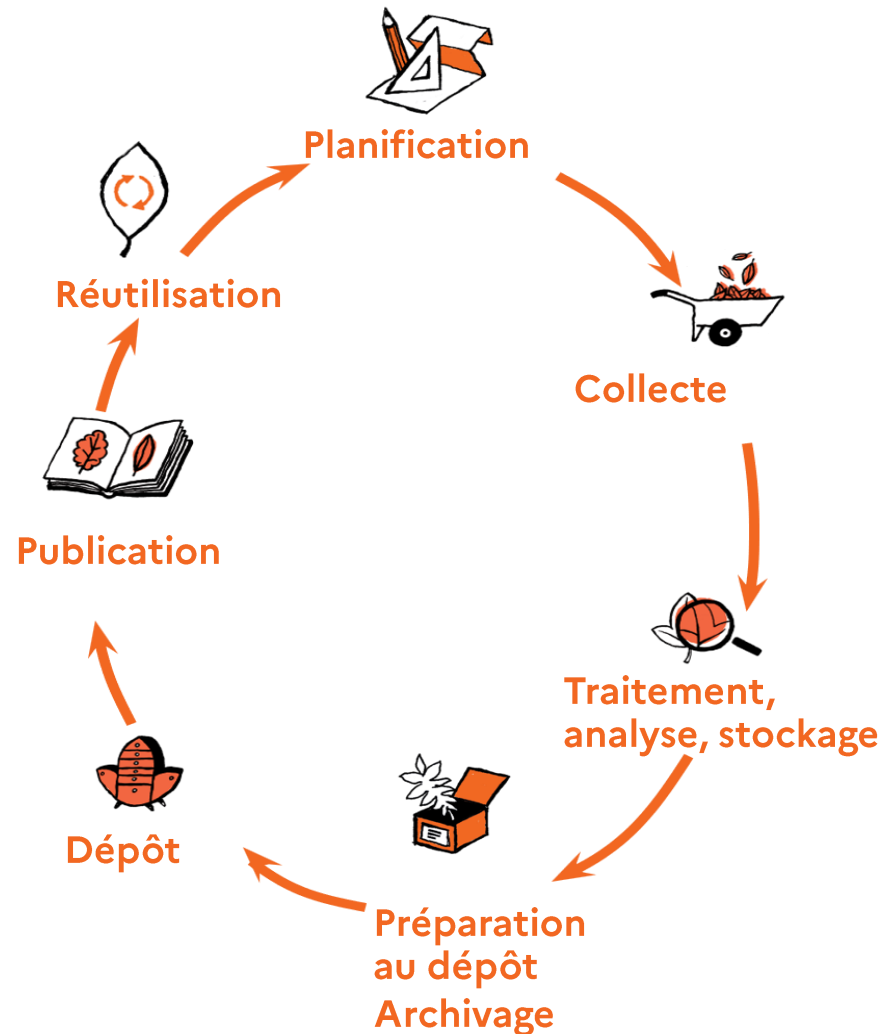
Les données de recherche en questions



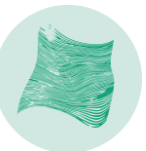
- Des données non documentées, non triées, mal nommées, ne permettant pas leur réutilisation
 - Multiplication des données produites et (re)produites (redondances)
 - Le stockage de versions intermédiaires (absence de tri, d'éliminations)
 - La réplication des données de recherche sur plusieurs lieux de stockage et entrepôts (sauvegardes, plusieurs versions, diffusion chez l'éditeur et dans un entrepôt...)
- (etc.)



Comment agir ?



- Agir sur les données pour en limiter la collecte, le stockage, la taille, sur tout le cycle de vie des données
 - Rechercher des données existantes avant d'en produire de nouvelles
 - S'interroger sur les durées de conservation, trier ses données au fur et à mesure, notamment quand volumétrie importante
 - S'interroger sur les modalités de diffusion selon l'usage et l'intérêt
 - Préférer un entrepôt national (Recherche Data Gouv) plutôt qu'une multitude d'entrepôts dans chaque établissement?

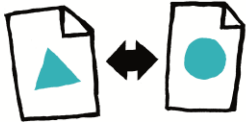


Comment agir ?

Faciles à trouver



Accessibles



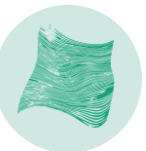
Intéropérables



Réutilisables

- Appliquer les principes FAIR
 - Diffuser des jeux de données structurés, pérennes, interopérables et réutilisables
- Notamment : construire des jeux de données...
 - Bien nommés
 - Bien organisés
 - Bien versionnés
 - Bien documentés
 - Bien protégés quand c'est nécessaire

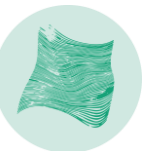
Pour pouvoir être diffusés et réutilisés



Comment agir ?

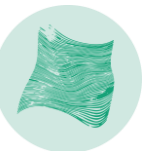


- Agir sur les logiciels de traitement des données et sur les logiciels que l'on développe pour ses recherches
 - Choisir des logiciels de traitement des données open source,
 - Ne développer des logiciels que si nécessaire,
 - Et quand on le fait...
 - Développer avec contrôle des versions lors du développement pour éviter la multiplication des versions (Git, subversion...)
 - Partager et archiver son code, pour qu'il puisse être réutilisé (forges logicielles, Software Heritage...)
 - Proposer/utiliser un accès aux données par API pour faciliter la récupération de données et réduire la transmission
 - Utiliser des serveurs de calcul sur des datacenters labellisés proposant des solutions écoresponsables (circuit fermé de refroidissement) et suivant la réglementation européenne



Comment agir ?

- **Le plan de gestion de données** : un outil pragmatique pour aller vers une science ouverte plus sobre ?
 - Définir une stratégie de gestion sobre des données
 - Adopter des pratiques rationnelles
 - Outil de concertation au sein d'une équipe: éviter les redondances
 - Sélection collective des jeux de données à conserver / à diffuser / à éliminer
 - L'occasion de documenter l'impact de ses données.



Merci de votre attention!

alicia.leonybarella@univ-lille.fr

donnees-recherche@univ-lille.fr

[Guichet unique - atelier de la donnée LORD](#)



Crédits



Illustrations : 4 minutes 34 - Guides et livrets de la collection *Passeport pour la science ouverte* - CC-BY-SA



En dehors des illustrations, les contenus de ce diaporama sont diffusés sous la licence CC-BY