

Gestion des données de recherche

Noemi Trino

Conseillère en gestion des données de recherche et éthique – Département de soutien à la recherche et à l'innovation



ULB PhD Day 2025



Research and innovation support Department



Rachel Leproult



Noemi Trino



Dragana Petrovic

Department of Libraries and Scientific Information



Monica Minneci



Anthony Leroy



L' ÉQUIPE ULB GDR



Dataedo /cartoon

Pro@Dataedo

Prévenir la perte de données et assurer leur pérennité

- 80 % des jeux de données de plus de 20 ans ont été perdus.
- Vos données peuvent-elles être facilement réutilisées par d'autres ?
- Pourrez-vous accéder à vos propres données (et les comprendre) dans 5 à 10 ans ?

Augmenter l'impact et les citations de votre recherche

- Des données bien gérées améliorent la **visibilité**, la **réutilisation** et les **citations**, renforçant ainsi la reconnaissance des chercheurs.
- La **disponibilité des données** est désormais exigée par la majorité des revues scientifiques.

Répondre aux exigences des bailleurs de fonds

- Les principaux bailleurs de fonds exigent un **plan de gestion des données** pour l'octroi des financements.
- Une documentation adéquate garantit le respect des **exigences légales et éthiques**.



De bonnes raisons de se soucier de ses données de recherche

Renforcer l'intégrité scientifique et la transparence

- Une bonne gestion des données permet la **vérification des résultats** et facilite leur **réutilisation** dans de futurs travaux de recherche.

Éviter la fraude scientifique et les recherches invérifiables

- Les *paper mills* sont responsables d'un nombre croissant de publications frauduleuses, jusqu'à **2 % des articles** dans certains domaines (*Nature*, 2021).
- **54 % des données** des études publiées ne sont pas vérifiables (*Vines et al.*, 2014).



De bonnes raisons de se soucier de ses données de recherche

Une croissance massive des données

90 % des données mondiales ont été produites au cours des deux dernières années (Forum Économique Mondial).

- La production de données scientifiques augmente de **30 % par an** (International Data Corporation).

Une vulnérabilité croissante des données

- Des changements **politiques ou institutionnels** peuvent restreindre l'accès aux données ou entraîner leur suppression.

La gestion des données de recherche comme garantie dans un environnement en évolution rapide

- La GDR permet d'assurer un **accès à long terme aux données**, indépendamment des décisions ou contraintes externes.

Comment gérer concrètement les données de recherche de manière durable et réaliste pendant un doctorat ?



La gestion des données de recherche et la résilience scientifique

📁 **La gestion des données de recherche (GDR)** comprend des activités telles que :

- ✓ **Planification**
- ✓ **Collecte**
- ✓ **Organisation**
- ✓ **Documentation**
- ✓ **Stockage**
- ✓ **Partage**
- ✓ **Préservation**

💡 Chaque chercheur est responsable de la gestion de ses données conformément à la [politique de l'ULB en matière de gestion des données de recherche](#).



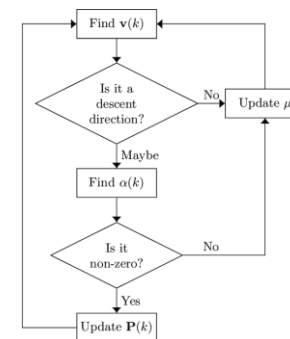


Les données de recherche sont toute information utilisée pour soutenir des résultats de recherche. Elles peuvent être :

- Quantitatives ou qualitatives
- Factuelles ou non factuelles
- Numériques, textuelles ou audiovisuelles

Exemples :

- Feuilles de calcul, documents, mesures de test
- Photographies, enregistrements audio et vidéo
- Diapositives, artefacts, spécimens, échantillons
- Modèles (numériques, analytiques, ingénierie)
- Logiciels de simulation, algorithmes, scripts
- Questionnaires, enquêtes, transcriptions
- Méthodologies



may mean that the Russians are about to announce the discovery of value added by manufacture. A section entitled "Examples of the application of composite indices in the Soviet Union and in People's Poland" is not very informative because there is no reference to anything that happened after 1952. A longer section, "Composite indices in the statistics of capitalist countries," makes the general point that such indexes may be completely unrealistic; if they are not it is usually a matter of coincidence, and even in the case of "correct indices published in capitalist countries a critical evaluation of their applicability from the point

Questionnaire – Métro de Montréal


Pour chacune des questions suivantes, répondez par Vrai ou Faux.

Question	Vrai	Faux
1. L'inauguration du métro de Montréal a eu lieu en 1966.		
2. 5000 ouvriers pendant 4 années pour sa construction.		
3. Le métro de Montréal comptait à l'origine 20 stations.		
4. Il fut un jour le métro le plus rapide au monde.		
5. Ses voitures pneumatiques sont toujours utilisées aujourd'hui.		
6. Les locomotives actuelles sont en service depuis 10 ans.		
7. C'est dans la ville de Québec que sont construites les nouvelles voitures.		
8. C'est 468 nouvelles voitures qui seront mises en service graduellement à compter de 2014.		
9. Pour son transport, la voiture modèle a dû être démantelée en 10 parties.		
10. Les tests techniques se déroulent pendant la nuit au métro de Montréal.		
11. Certains passages laissent à peine 5 centimètres de		

Date	Start Time	Lunch Start	Lunch End	End Time	Overtime	Hou
4/1/2019	8:00am	1:00pm	1:30pm	5:00pm	0	
4/2/2019	8:00am	1:00pm	1:30pm	5:00pm	0	
4/3/2019	8:00am	1:00pm	1:30pm	5:00pm	2	
4/4/2019	8:00am	1:00pm	1:30pm	5:00pm	0	
4/5/2019	8:00am	1:00pm	1:30pm	5:00pm	0	
4/6/2019	900am	2:00pm	2:30pm	6:00pm	0	
4/7/2019	900am	2:00pm	2:30pm	6:00pm	0	
4/8/2019	800am	1:00pm	1:30pm	5:00pm	1	
4/9/2019	800am	1:00pm	1:30pm	5:00pm	1	
4/10/2019	800am	1:00pm	1:30pm	5:00pm	1	

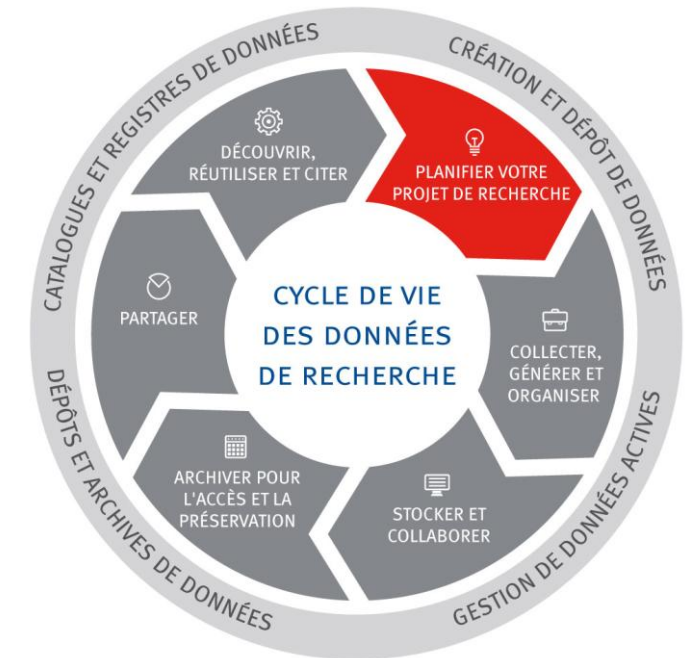


Que sont les données de recherche

 Un **Plan de Gestion des Données (Data Management Plan, DMP)** est un document **formel et évolutif** décrivant **comment** les données collectées seront gérées **tout au long du cycle de vie** d'un projet.

Il permet de :

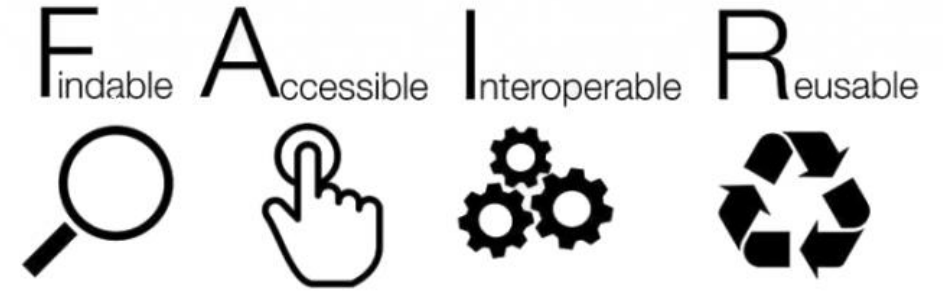
- ✓ Assurer la qualité, la sécurité et la durabilité des données
- ✓ Garantir l'accessibilité et la réutilisation des données
- ✓ Mettre à jour les informations à mesure que le projet évolue







- 📌 **Collecte des données** – Quels types de données seront collectés, traités ou générés ?
- 📌 **Gestion des données** – Quels formats, méthodologies et meilleures pratiques seront utilisés ?
- 📌 **Stockage des données** – Où et comment les données seront-elles stockées pendant et après le projet ?
- 📌 **Considérations légales et éthiques** – Sécurité, confidentialité, aspects réglementaires et éthiques.
- 📌 **Partage et accessibilité** – Les données seront-elles ouvertes ? Quelles restrictions s'appliquent ?
- 📌 **Préservation à long terme** – Comment assurer leur pérennité après la fin du projet ?
- 📌 **Objectif : Garantir que les données restent accessibles et réutilisables sur le long terme.**



Le Plan de Gestion des Données



-  **Findable (Facile à trouver)** – Utiliser des **métadonnées riches**, des **mots-clés**, et des PID (DOI).
-  **Accessible** – Archiver les données dans des **référentiels de confiance**.
-  **Interoperable** – Utiliser des **formats et normes ouverts** plutôt que propriétaires.
-  **Reusable (Réutilisable)** – Ajouter **de la documentation** (README) et des **licences ouvertes** (CC0, CC-BY 4.0).

***Métadonnées:** Informations structurées qui décrivent, expliquent, localisent et contextualisent les données.

Exemple : nom(s) de l'auteur, ORCID de l'auteur(s), DOI, licence, langue, journal, titre, etc.

DOI: Un **Digital Object Identifier est un identifiant persistant ou un *handle* utilisé pour identifier de manière unique divers objets, standardisé par l'Organisation internationale de normalisation (ISO).

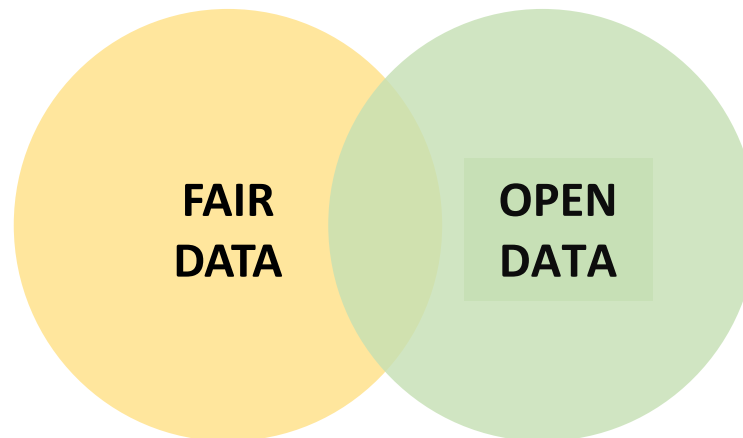
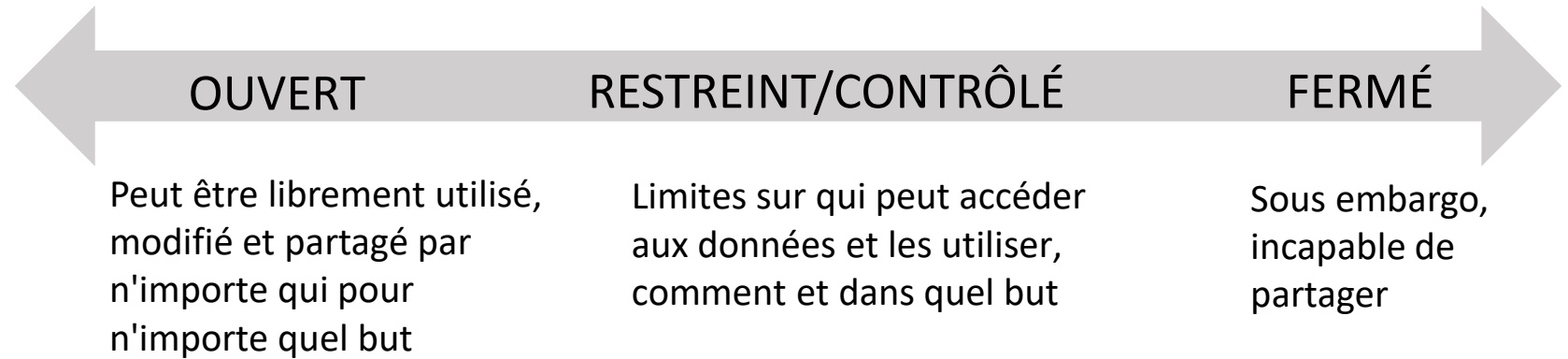


Principes FAIR pour les données

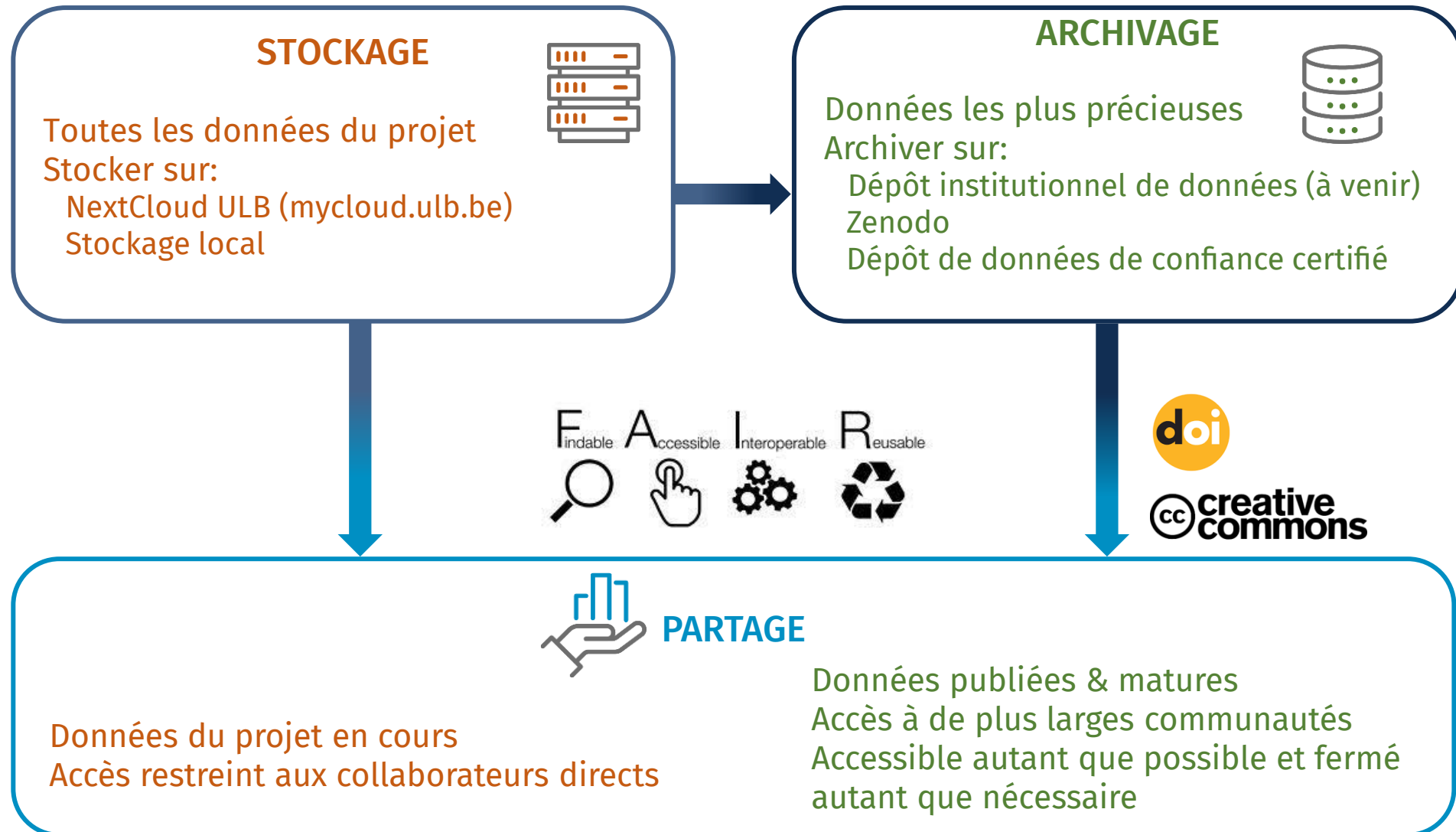
FAIR VS OPEN DATA



“Accessible autant que possible et fermé autant que nécessaire”



Adapté de '[Managing and sharing research data](#)', S. Jones, CC-BY



Stockage des données du projet (à court terme/pendant le projet) auxquelles vous et vos collaborateurs pouvez avoir accès

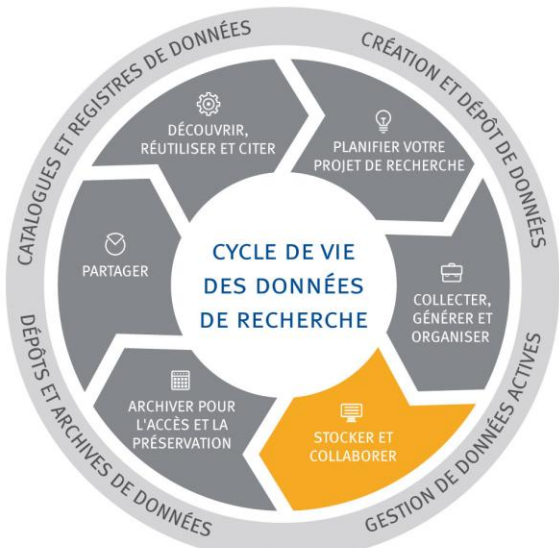
Règle 3-2-1

- 3 exemplaires
 - 2 formats de stockage
 - 1 « hors site »
- Ex : Votre ordinateur de bureau, un disque dur externe, Nextcloud (<https://mycloud.ulb.be>)

Dépôts de données (à long terme/après le projet):

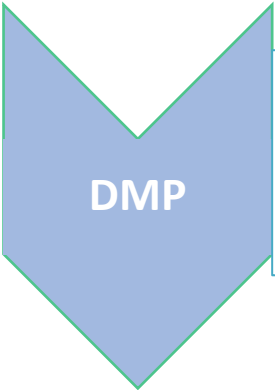
- **Dépôts généralistes** (Ex. Zenodo, Open Science Framework).
 - **Dépôts spécifiques à un domaine.**
 - **Dépôts institutionnels** (En développement à l'ULB).
- 📌 **Registres pour trouver des dépôts de données :**

🔍 re3data.org
🔍 [FAIRsharing.org](https://fairsharing.org)



STOCKAGE + ARCHIVAGE





Data Management Plan - tool

← → ↺

https://dmponline.be/plans/new

Create a new plan

Before you get started, we need some information about your research project to set you up with the b

* What research project are you planning?

☐ mock project for testing, practice, or educational purposes

* Select the primary research organisation

Université libre de Bruxelles

* Select the primary funding organisation

✓ --Select funder--

Belgian Federal Science Policy Office (BELSPO)

European Commission (Horizon)

European Research Council (ERC)

Fonds National de la Recherche Scientifique (FNRS)

Fonds voor Wetenschappelijk Onderzoek - Research Foundation Flanders (FWO)

Vlaams Agentschap Innoveren & Ondernemen (VLAIO)

Create plan

Cancel

2.1.1 Will data and other research outputs be identified by a persistent identifier?

☐ Yes: describe below

☐ No: describe below

Additional Information

B I [list icon] [list icon] [link icon] [table icon]

Save

Guidance

Comments

EC

Persistent Identifiers (PIDs) must be provided for the data, for all author(s) involved in the action and, if possible, for their organizations and the grant.

Examples of commonly used PIDs include:

- DOI or handle for data and other research outputs
- ORCID or ResearcherID for authors
- ROR ID for organizations
- grant DOI for grants
- but also accession numbers within specific disciplines, notably in the Life Sciences e.g. an accession number in a database such as the Protein Data Bank.

Persistent Identifiers for data or other research output may be provided by trusted repositories where the research output is deposited (see also question 2.2.3).

Tool : <https://dmponline.be/>

Support: rdm-support@ulb.be

- **Pages Web dédiées à GDR** contient:
 - la Guide Pratique et la Politique Institutionnelle ULB sur la gestion des données des pages centralisées
 - des ressources mises à jour:
 - <https://portail.ulb.be/en/research/research-data-management>
 - <https://portail.ulb.be/fr/recherche/gestion-des-donnees-de-recherche>
- **Formations**
 - <https://portail.ulb.be/en/research/research-data-management/training-and-support>
 - <https://portail.ulb.be/fr/recherche/gestion-des-donnees-de-recherche/formation-et-soutien>
- **Rencontres individuelles** (pour DMP ou toute autre question relative à GDR)
- **Relecture** de votre DMP et vous fournir un **feedback détaillé**
- **Modèles de DMP** avec conseils mis en oeuvre:
 - **DMPonline** - outil en ligne pour créer votre DMP (y compris divers modèles de bailleurs de fonds et conseils de rédaction) – gratuit pour la communauté ULB www.dmponline.be (connexion avec ULBID)
- **Contact** : rdm-support@ulb.be



SUPPORT GDR & DMP à l'ULB

- **ULB MyCloud** – Stockage institutionnel des données de recherche
<https://mycloud.ulb.be> support@ulb.be – assistance technique

DO I HAVE TO WRITE A DATA MANAGEMENT PLAN ?

Ressources externes

Do I have to write a DMP?

Outil pratique d'aide à la décision permettant d'identifier si un **plan de gestion des données** est requis et quelles obligations s'appliquent, selon le type de financement.



Compass to Research Data Management

Questionnaire interactif guidant les chercheurs à travers les réglementations, recommandations et obligations liées à la gestion des données de recherche en Fédération Wallonie-Bruxelles.

Navigating the data management ocean: specificities of my dataset

The following questions will help you narrow down the specificities of your dataset, and highlight the applicable rules or recommendations.

#1 Does your dataset include medical, health (physical, mental) or life sciences data ? *

SUPPORT GDR & DMP à l'ULB



Données personnelles ou sensibles, ou sujets humains ou animaux

- Approbation du comité d'éthique
- Anonymiser/pseudonymiser avant le partage
- Accord de partage des données inclus dans le consentement éclairé
- La Déléguée à la Protection des Données (DPO) de l'ULB peut être consultée pour des conseils méthodologiques et/ou pour toute question liée au RGPD: (email : rgpd@ulb.be).
- Une **analyse d'impact relative à la protection des données (DPIA)** peut être exigée en fonction des risques identifiés.

Recherche menée dans ou avec des pays hors Union européenne

- Les chercheurs doivent évaluer les **risques liés aux droits fondamentaux** avant toute collaboration.
- En cas de **partenariat formel**, une approbation du [Comité Respect du droit international et Dual use](#) est requise.



ASPECTS ÉTHIQUES ET JURIDIQUES

Recherche impliquant des outils et méthodes d'intelligence artificielle

Les projets utilisant l'IA doivent respecter plusieurs cadres juridiques :

- **RGPD** : base légale, minimisation des données, DPIA si nécessaire
- **Droits d'auteur et droits des bases de données**, y compris les exceptions de *text and data mining*
- **Conditions d'utilisation** des plateformes (scraping) et **sécurité informatique**
- **Règlement européen sur l'IA (EU AI Act)** : transparence et gestion des risques

Recherche à potentiel de dual use (civil–militaire)

- Certaines recherches peuvent avoir des usages civils légitimes mais aussi des **applications potentiellement nuisibles**.
- Dans ce cas, les chercheurs sont invités à contacter le [Comité Respect du droit international et Dual use](#) pour un accompagnement et une évaluation.



ASPECTS ÉTHIQUES ET JURIDIQUES

Gestion des données de recherche & DMP

rdm-support@ulb.be

Questions relatives au RGPD

rgpd@ulb.be

<https://portail.ulb.be/fr/documents-officiels/protection-des-donnees-a-caractere-personnel>

Comités d'éthique

Ethics committees – ULB

Respect du droit international et double usage

<https://www.ulb.be/fr/partenaires-et-reseaux/developper-un-partenariat>

Valorisation et propriété intellectuelle

ulbkto@ulb.be

<https://portail.ulb.be/en/research/service-technology-transfer-tto>

<https://portail.ulb.be/en/research/intellectual-property>

Science ouverte et publications en libre accès

<https://www.ulb.be/fr/transparence-et-integrite/science-ouverte-open-science>



Contacts et accompagnement à l'ULB



- Merci!
- Pour nous contacter: rdm-support@ulb.be

Your future self will thank you!