

# **Introduction à la Gestion des données de recherche (GDR)**

**par Victoria Volkanova, bibliothécaire  
17 février 2022**



**UNIVERSITÉ DE MONCTON**  
EDMUNDSTON MONCTON SHIPPAGAN



# Éléments abordés

- ❑ Qu'est-ce que la GDR, son importance
- ❑ Les principes FAIR
- ❑ Au Canada : la Politique des trois grands conseils de recherche
- ❑ Trois piliers de la Politique
  - Stratégie institutionnelle
  - Plan de gestion des données (PGD)
  - Dépôts de données / Principes TRUST
- ❑ Bonnes pratiques en GDR
- ❑ La science ouverte / les données ouvertes
- ❑ Ressources nationales et locales

# Données de recherche : Définition

Les **données primaires** pour la recherche; nécessaires pour valider les conclusions de la recherche, ex. données textuelles, audio, iconographiques, numériques, etc.; données d'observation, factuelles, personnelles...

**Créées, stockées et gérées numériquement** => il est facile de les rendre accessibles, de les reproduire et de les combiner avec d'autres données.

Important : une **gestion active des données** tout au long de leur cycle de vie, et ce dès le début de votre projet de recherche.

# Importance de la GDR

## Pour les chercheur.e.s

- ✓ Meilleure organisation du travail
- ✓ Conservation sécuritaire des données
- ✓ Protection contre la perte ou le vol des données
- ✓ Visibilité et impact
- ✓ Nouvelles collaborations
- ✓ Réutilisation des données
- ✓ Intégrité, transparence

## Pour la science

- ✓ Avancement accéléré des connaissances
- ✓ Reproductibilité de la recherche
- ✓ Limite la répétition des travaux
- ✓ Validation des résultats de la recherche

# Principes directeurs FAIR

**F**indable = *Facilement repérables* :  
*faciliter la découverte des données*

**A**ccessible = *Accessibles* : *accéder et télécharger facilement les données*

**I**nteroperable = *Interopérables* : *exploiter les données quel que soit l'environnement informatique utilisé*

**R**eusable = *Réutilisables* : *réutiliser les données pour de futures recherches*

# Principes **FAIR**

## *Facilement repérables :*

- Attribuer des identifiants uniques et pérennes (**PID-Persistent IDentifiers**) aux données
- Décrire finement à l'aide des **métadonnées**
- Indiquer le **PID** dans les **métadonnées**
- Déposer ses données dans un **entrepôt de données**

# Principes FAIR

## *Accessibles :*

- Utiliser un protocole de communication
  - Standardisé (ex. HTTP, FTP)
  - Libre et ouvert vs protocoles propriétaires (Skype, Microsoft Exchange Server)
  - Sécurisé /permettant une authentification si nécessaire (HTTPS ou FTPS); spécifier les conditions exactes d'accès aux données privées, qui et comment
- Préserver l'accès aux métadonnées, même si les données ne sont plus disponibles

# Principes FAIR

## *Interopérables*

- Décrire les données avec un **vocabulaire contrôlé**, pour permettre aux machines d'avoir un langage commun
- Utiliser un vocabulaire qui **respecte les principes FAIR** (ex. Animal Diseases Ontology (ANDO), doi [10.15454/1.44525654526207E12](https://doi.org/10.15454/1.44525654526207E12))
- Contextualiser avec des **liens vers d'autres données** => réseau global d'informations scientifiques (web sémantique, format RDF)



# Principes FAIR

## *Réutilisables*

- Donner le plus d'information descriptive; à priori, toutes les **métadonnées sont pertinentes** (ex. protocole expérimental, paramètres, qui a traité les données, quand, où, v. du logiciel, etc.).
- Attribuer une **licence de réutilisation**



LICENCE OUVERTE  
OPEN LICENCE

**ODbL**

- Indiquer l'**historique / la provenance des données** pour prouver leurs crédibilité et fiabilité : dans les métadonnées ou Read Me.txt
- Utiliser des **standards de sa communauté** pour partager les données

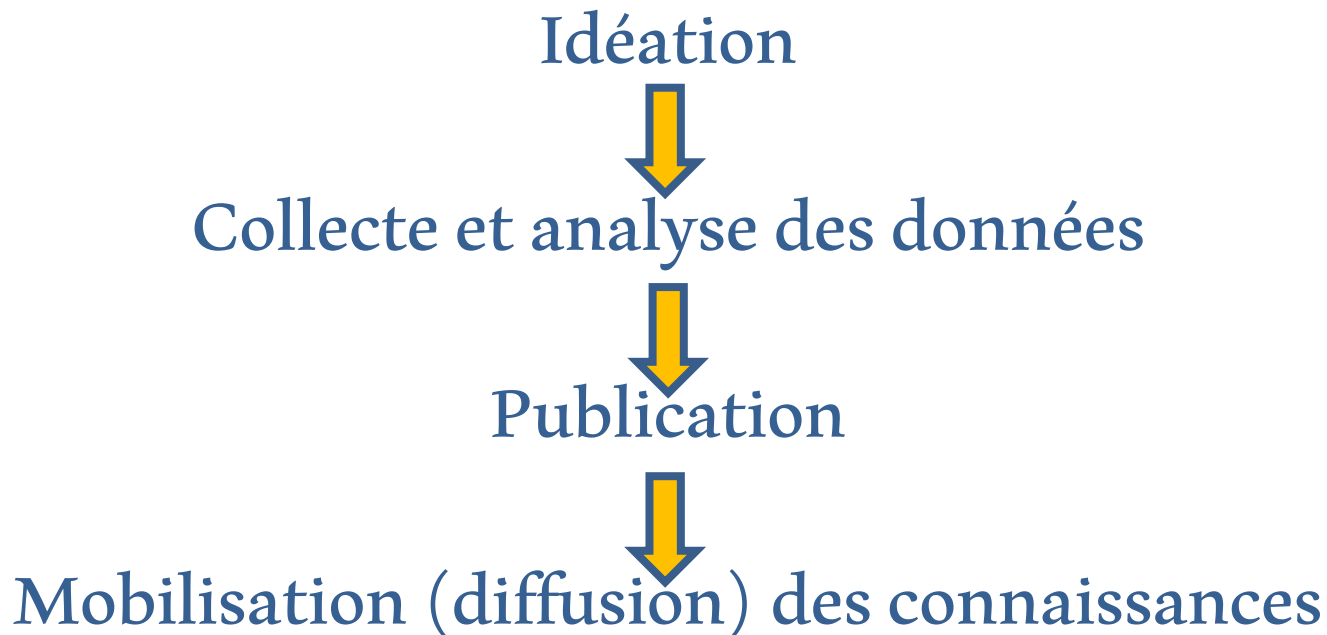
# Science ouverte

[...] la pratique qui consiste à rendre les intrants, les résultats et les processus scientifiques librement accessibles à tous avec un minimum de restrictions. Les résultats de la recherche scientifique comprennent i) des articles et des publications scientifiques examinés par les pairs, ii) les données scientifiques et de recherche, iii) la contribution du public à la science et le dialogue à cet égard.

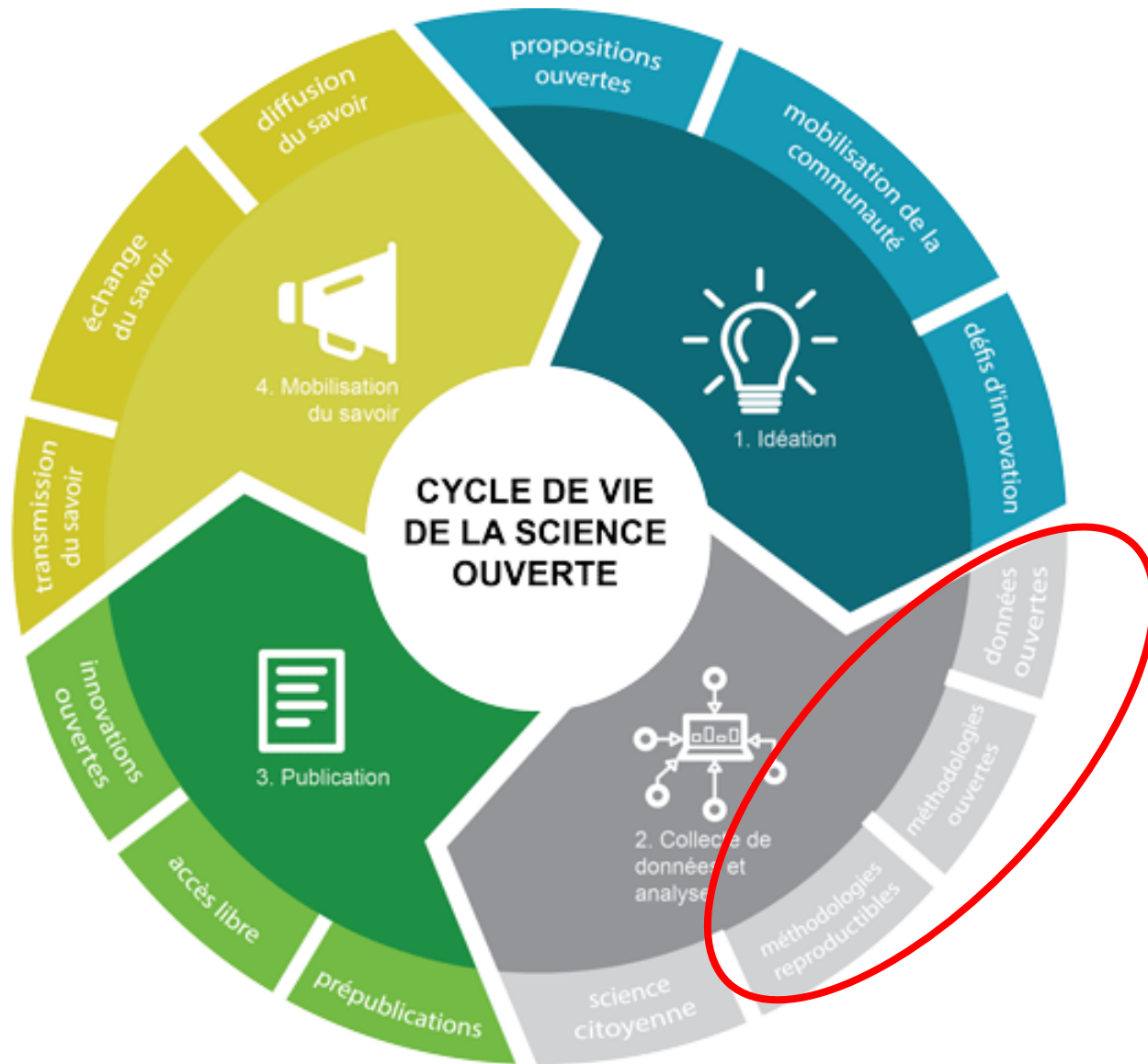
La science ouverte est rendue possible par les personnes, la technologie et l'infrastructure. Elle se pratique dans le respect de la vie privée, de la sécurité, de l'éthique et de la protection appropriée de la propriété intellectuelle.

# Science ouverte : cycle de vie

- La science ouverte peut intervenir à toutes les étapes du cycle de vie de la recherche :



# Science ouverte : diagramme



# Données ouvertes

*Est-ce que mes données sont partageables ?*

- Matrice de risque lié aux données de recherche avec les êtres humains (Portage)
- Arbre décisionnel pour le partage des données avec participants humains (Portage)
- Outil d'aide à la décision sur le partage des données (FRQ)

# Politique nationale

## Politique des trois conseils de recherche sur la GDR

- La GDR est un élément nécessaire de l'excellence en recherche
- Il ne s'agit pas des données ouvertes, le principe « aussi ouvertes que possible, aussi fermées que nécessaire »
- Reconnait les différences légitimes dans les normes de la GDR entre les disciplines, les domaines de recherche et les modes de recherche
- Vise à promouvoir l'excellence en matière de recherche au Canada en encourageant de saines pratiques de gestion et d'intendance des données de recherche
- Trois piliers : Stratégie institutionnelle; PGD; dépôts de données



# Pilier 1

## 1. **Stratégie institutionnelle** (échéance : mars 2023)

Objectif : comprendre les besoins, identifier les lacunes et commencer à les remédier; trouver des solutions institutionnelles.

La stratégie n'est pas prescriptive, mais doit adresser plusieurs éléments. Elle sera révisée, mais pas évaluée.

Son intention: l'esprit de collégialité, augmenter la capacité en matière de la GDR de l'institution et à travers le pays.

# Pilier 2

**2. Plan de gestion de données** (dès printemps 2022, les subventions exigeant un PGD seront identifiées)

« [Le PGD] est un document synthétique qui aide à organiser et anticiper toutes les étapes du cycle de vie de la donnée. Il explique pour chaque jeu de données comment seront gérées les données d'un projet, depuis leur création ou collecte jusqu'à leur partage et leur archivage. »\*

Un document évolutif, doit être mis à jour

\**Source* : DoRANum. (2018). [Fiche synthétique : Plan de gestion de données.](#)



# Pilier 2 (cont.)

ASSISTANT

Accueil PGD publics Modèles de PGD Aide À propos de

Langue Se connecter

Bienvenue

Mon tableau de bord Créer des plans Référence Aide

L'Assistance

## Créer un nouveau plan

Avant de commencer, nous avons besoin de renseignements sur votre projet de recherche pour vous offrir le modèle de PGD qui répond le mieux à vos besoins.

\* Quel projet de recherche planifiez-vous?

projet simulé aux fins d'essai, d'exercice ou d'enseignement

\* Indiquer l'organisme de recherche principal

- ou -  Aucun organisme de recherche n'est associé à ce plan, ou mon organisme de recherche n'est pas répertorié

Créer un plan

Annuler

## Pilier 2 (cont.)

Un PGD contient une vingtaine de questions portant sur les aspects suivants :

Collecte de données

Documentation et métadonnées

Stockage et sauvegarde

Préservation

Partage et réutilisation

Responsabilité et ressources

Conformité éthique et juridique

# Pilier 3

## 3. Dépôt des données de recherche

Un service / une archive en ligne qui assure la gestion des données de recherche à travers le cycle de vie des données. Inclure : les données, la documentation, les métadonnées descriptives

- DI / DI des données de recherche
- Dépôts de données nationaux : [Dataverse](#) ; [FRDR](#)
- Dépôts spécifiques aux disciplines
- Dépôts spécifiques aux domaines

*Pour trouver un dépôt : [re3data.org](http://re3data.org)*

# Principes directeurs TRUST

*Pour les dépôts de données fiables :*

Transparency = la transparence

Responsibility = la responsabilisation

User focus = l'orientation vers ou l'intérêt pour l'utilisateur

Sustainability = la pérennité ou la durabilité des données

Technology = la technologie utilisée

Source : [The TRUST Principles for Digital Repositories](#)  
(2020)



# Pilier 3

## *Choisir un dépôt : Critères d'évaluation*

- Réputation
- Pays de stockage des données
- Coûts et volume des données
- Langue pour la description
- Identifiant unique (doi)
- Administrateurs des jeux de données (niveau de contrôle)
- Types de licences et de formats
- Gestion des accès restreints

# Bonnes pratiques en GDR

## 1. Planifier

- Rédiger un PGD
- Obtenir le consentement des participants quant à la diffusion éventuelle des données avant leur collecte
- Identifier le bon processus pour protéger les données potentiellement sensibles (ex: anonymisation ou accès contrôlé)

# Bonnes pratiques en GDR

## 2. Stocker et sauvegarder

- Choisir la plateforme de stockage de données : sécurité, coût, lieu des serveurs, récupération de fichiers, versionnage, etc.
- Suivre la règle du 3-2-1 :
  - ✓ 3 copies sauvegardées
  - ✓ sur 2 supports de stockage distincts,
  - ✓ dont 1 situé hors campus
- Définir une fréquence de sauvegarde

[Plateforme ACENET](#)

# Bonnes pratiques en GDR

## 3. Organiser

- Hiérarchie des dossiers
- Nomenclature des fichiers ([guide](#) de l'UQÀM)
- Documentation aidant à contextualiser les données :

✓ Fichier LISEZMOI.txt (modèle en [français](#) et en [anglais](#))

✓ [Dictionnaire de](#)

Nom	Définition	Type	Longueur	Règle	Exemple
CodeP	Code du participant	AN	6		635958
CivilP	État civil du participant	A	20	5 types d'état civil	uni civilement
Datenaiss	Date de naissance du participant	D	10	aaaa-mm-jj	1980-01-01
NbEnf	Nombre d'enfants du participant	N	20		2

✓ Métadonnées riches

⇒ Uniformité, constance et logique



# Bonnes pratiques en GDR

## 3. Organiser : exemple

### Exercice

#### Enquête sur la moutarde

questionnaire\_2021.pdf

enquetesurlamoutarde\_lesconsignes.txt

Enquête\_moutarde\_données\_2021\_1.xlsx

Enquête\_moutarde\_données\_2021\_2.xlsx

VideoATremblay.mp4

VerbatimATremblay.doc

VideoDSmith.mp4

VerbatimDSmith.doc

VideoJ.Jacques.mp4

VerbatimJ.Jacques.doc

VideoR.St-Pierre.mp4

VerbatimR.St-Pierre.doc

#### Enquete sur la moutarde

##### Documentation

Enquete\_moutarde\_questionnaire\_2021.pdf

Enquete\_moutarde\_consignes.txt

**Enquete\_moutarde\_Dictionnaire\_Donnees\_2021.pdf**

**Lisez-moi\_Enquete\_moutarde\_2021\_2021-09-27.txt**

##### Donnees\_sondage : données tabulaires

Enquete\_moutarde\_donnees\_2021\_v01.csv

Enquete\_moutarde\_donnees\_2021\_v02.csv

##### Verbatim : verbatim des entrevues

Enquete\_moutarde\_Verbatim\_P1\_2021.txt

Enquete\_moutarde\_Verbatim\_P2\_2021.txt

Enquete\_moutarde\_Verbatim\_P3\_2021.txt

Enquete\_moutarde\_Verbatim\_P4\_2021.txt

##### Video (si autorisé): vidéos des entrevues

Enquete\_moutarde\_Video\_P1\_2021.mp4

Enquete\_moutarde\_Video\_P2\_2021.mp4

Enquete\_moutarde\_Video\_P3\_2021.mp4

Enquete\_moutarde\_Video\_P4\_2021.mp4

# Bonnes pratiques en GDR

## 4. Formats de fichiers

- Privilégier des formats de fichiers ouverts. Cependant, dans certaines disciplines, les formats propriétaires sont acceptés.
- Consulter ce [tableau de UK Data Service](#) sur les formats recommandés

Ex: .txt; .csv; .xml; .tif; .mp4

- Objectifs :
  - Préservation à long terme
  - Éviter les erreurs de conversion
  - Permettre un meilleur partage des données

# Prochaines étapes à l'UdeM

- Comité du développement de la stratégie institutionnelle à l'Université de Moncton (en cours de composition)
- Gabarits de PGD pour l'U de Moncton
- Solutions de dépôts : ScriptoriUM? Dataverse? Plateformes existantes?
- Mise en application, sensibilisation

# Quelques ressources

- Série de webinaires sur la GDR
- Politique des grands conseils de recherche (2021)
- Matériel de formation du réseau Portage\*

*\*nouveau nom l'Alliance de recherche numérique du Canada (l'Alliance)*

- Page web Données de recherche (Bibl. Champlain)
- Feuille de route pour la science ouverte (Canada)
- DORA Num

**Questions ? Commentaires ?**

**Personnes-ressources :**

FESR : [Francis Leblanc](#)

Bibliothèque Champlain : [Victoria Volkanova](#)



**UNIVERSITÉ DE MONCTON**  
**CAMPUS DE MONCTON**

**Bibliothèque Champlain**