

# PLANIFIEZ votre RECHERCHE et la GESTION de vos DONNÉES avec le PLAN de GESTION de DONNÉES

Jeudi 19 octobre 2023 – 11h00  
Session #1 : APPORTS THÉORIQUES

Maélys LISERON-MONFILS  0000-0002-1040-7133

et Hélène TISSERAND  0000-0002-1150-4541

Documentalistes de l'Atelier de la donnée dat@UBFC

En partenariat avec la MSHE Claude-Nicolas Ledoux





# Qui sommes-nous ?

dat@UBFC, c'est :

L'atelier de la donnée en Bourgogne-Franche-Comté, pour l'accompagnement à la gestion et la valorisation des données de la recherche

Qui s'appuie sur l'expérience de l'équipe « données de la recherche de l'OSU THETA »

Et s'adresse aux **7 établissements** et **51 laboratoires** répartis sur toute la **région Bourgogne-Franche-Comté**

**Sylvie Damy**, laboratoire Chrono-environnement

**Léna Bouillard, Maélys Liseron-Monfils, Rachel Prost et Hélène Tisserand**, documentalistes

**Raphaël Melior et Nadir Tabbou**, informaticiens

Un réseau de **référent.e.s** et de membres du **CoPIL** répartis sur tout le territoire de la région BFC

# L'Atelier de la donnée dat@UBFC

## Les SERVICES ACTIFS :



### Service 1 : Portail d'exposition des données

- Description et exposition  
de jeux de données :  
<https://search-data.ubfc.fr/>

- Interopérabilité avec des  
portails nationaux et  
internationaux

Interopérabilité avec RDG,  
BibCNRS, Datacite



### Service 2 : Actions de formation

- 6 formations proposées  
à l'École doctorale et à la  
formation permanente des  
personnels et en  
partenariat avec la MSHE  
Claude-Nicolas Ledoux

- Formations ponctuelles  
sur demande (ex : Agro  
Dijon, Bibliest)



### Service 3 : Actions de sensibilisation

- Événements thématiques  
(Printemps de la  
donnée, Open Access  
Week)

- Interventions dans les  
laboratoires sur demande



### Service 4 : Accompagnement des chercheur.se.s

- Description des données sur  
le portail

- Rédaction / relecture des  
PGD

- Réflexion sur la gestion  
des données au sein des  
laboratoires

- Accompagnement des  
porteur.se.s de projet ANR /  
Europe



### Service 5 : Diffusion de l'information sur l'ouverture des données de la recherche

- Site Web

<https://data.ubfc.fr/>

- X/Twitter [@dataubfc](https://twitter.com/dataubfc)

- Partage au sein du  
réseau des Ateliers de la  
Donnée (Recherche Data  
Gouv)

## Les SERVICES à VENIR :



Service 6 : Développement d'un réseau d'expert.e.s pour la  
gestion des données spécifiques



Service 7 : Service de stockage en lien avec  
Recherche Data Gouv



# OBJECTIFS de la FORMATION

---



Comprendre les enjeux et les bénéfices des plans de gestion de données



Comprendre ce qu'est un DMP et ses principales composantes



Créer un DMP avec l'outil DMP OPIDoR



[Passeport pour la science ouverte : Guide pratique à l'usage des doctorant.e.s](#)

→ « Prévoir la gestion des données » (p.8)



# PLAN de la FORMATION

---

- **Session #1 : Apports théoriques**
  - > Le PLAN de GESTION de DONNÉES dans le CONTEXTE de la SCIENCE OUVERTE
  - > POURQUOI rédiger un PGD ? Les BÉNÉFICES à attendre
  - > RÉDIGER son PGD avec l'outil DMP OPIDoR
  - > En PRÉVISION de l'ATELIER PRATIQUE
- **Session #2 : Atelier pratique**
  - > PRISE en MAIN de DMP OPIDoR
  - > CRÉATION / RÉDACTION de votre 1er PGD



# Le **PLAN** de **GESTION** de **DONNÉES** dans le **CONTEXTE** de la **SCIENCE OUVERTE**

# SCIENCE OUVERTE, DONNEES OUVERTES

## Science ouverte

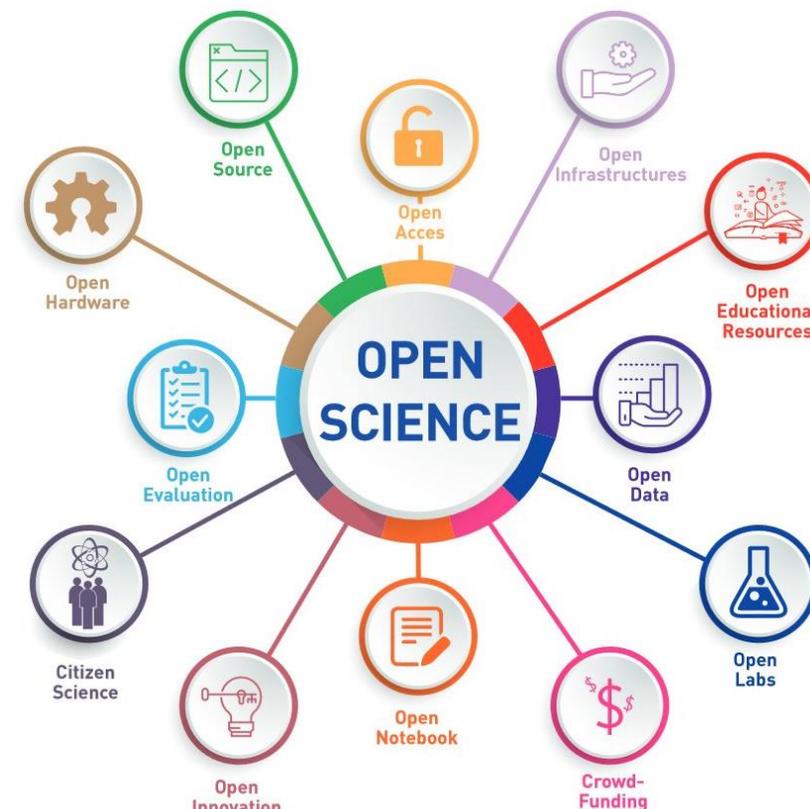
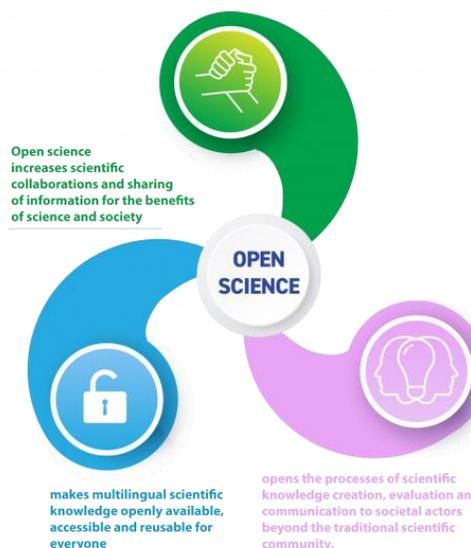
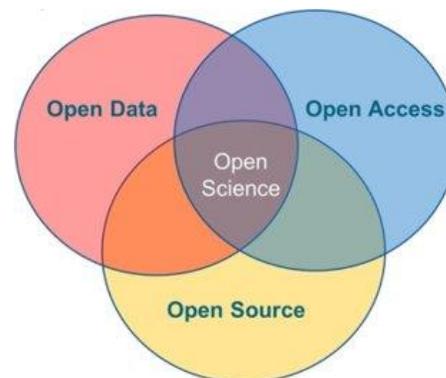
(*open science* ou *open research*) : mouvement visant à rendre la recherche scientifique et les données qu'elle produit accessibles à tou.te.s

## Données ouvertes (*open data*) :

données numériques dont l'accès et l'usage sont laissés libres aux usager.e.s

« Public money, public data »

« Aussi ouvert que possible, aussi fermé que nécessaire »



[UNESCO – Open Science](#)

# Les PLANS de GESTION de DONNÉES... c'est quoi ?

1

« Un Data Management Plan (DMP) ou plan de gestion de données est un document formalisé explicitant la manière dont seront **obtenues, documentées, analysées, disséminées et utilisées** les données produites au cours et à l'issue d'un processus ou d'un projet de recherche. »

A. Cartier, M. Moysan, N. Reymonet  
[Réaliser un plan de gestion de données : guide de rédaction](#)  
(V1, 09/01/2015). Universités Paris Diderot, Sorbonne Paris Cité, Paris Descartes

2

« Le plan de gestion des données est un **outil de gestion**.

Il se présente sous la forme d'un **document structuré en rubriques**.

Il a pour objectif de **synthétiser la description et l'évolution des jeux de données de votre projet de recherche**.

Il prépare le **partage, la réutilisation et la pérennisation des données**. »

D'après DoRANum, [Parcours interactif sur la gestion des données de la recherche](#)

## Au coeur du DMP, les DONNÉES de la RECHERCHE

1

Définition de l'OCDE :

« **Enregistrements factuels** (valeurs numériques, textes, images et sons) [...] utilisés comme **sources principales pour la recherche scientifique** et généralement reconnus par la communauté scientifique comme **nécessaires pour valider des résultats de recherche**. »

2

Jeu de données (ou dataset) :

« **Agrégation [...]** de **données brutes ou dérivées** présentant une certaine "unité", **rassemblées pour former un ensemble cohérent**. »

[Rémi Gaillard](#)

C'est donc un lot **techniquement homogène**, ou **intellectuellement cohérent** (bien que composé de lots techniquement hétérogènes).

# RÉPONDRE aux exigences des FINANCEURS et TUTELLES

## Attendus pour Horizon Europe

[https://ec.europa.eu/info/funding-tenders/opportunities/docs/2021-2027/horizon/guidance/programme-guide\\_horizon\\_en.pdf](https://ec.europa.eu/info/funding-tenders/opportunities/docs/2021-2027/horizon/guidance/programme-guide_horizon_en.pdf)

Un Research Data Management responsable, basé sur les principes FAIR :

- Obligatoire : DMP
- Obligatoire : informations pour pouvoir utiliser les données et valider les résultats
- Recommandé : ouverture code, software, workflows

Open Access pour les publications :

- Dépôt immédiat dans un entrepôt de confiance
- Licences CC-BY et CC-BY NC/ND

→ Développement de [ORE \(Open Research Europe\)](#) pour publier en Open Access

## Attendus pour l'ANR

<https://anr.fr/fileadmin/aap/2022/aapg-2022-v1.01.pdf>

« Les bénéficiaires de l'ANR s'engagent à **garantir le libre accès immédiat aux publications scientifiques évaluées par les pairs et à adopter, pour les données de la recherche, une démarche dite FAIR (Facile à trouver, Accessible, Interopérable, Réutilisable) conforme au principe "aussi ouvert que possible, aussi fermé que nécessaire" »**

« Le coordinateur ou la coordinatrice du projet s'engage à **fournir dans les 6 mois qui suivent le démarrage du projet, une première version du Plan de Gestion des Données (PGD) selon les modalités communiquées dans les conditions particulières**»

## Décret sur l'intégrité scientifique

03/12/2021 – [article 6](#)

« Les personnels des établissements publics de la recherche **doivent mettre en œuvre des plans de gestion de données** (et les établissements **doivent définir une politique de conservation, de communication et de réutilisation des résultats bruts des travaux scientifiques menés en leur sein**). »

# La REPRODUCTIBILITE de la recherche avec les principes FAIR

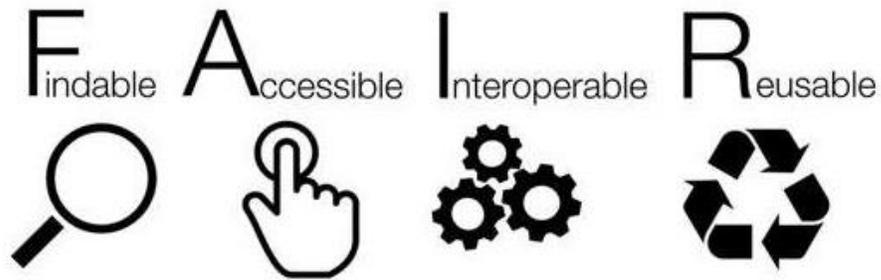


Image : SangyaPundir

2014 : naissance des principes FAIR

2016 : publication fondatrice des [principes FAIR](#) dans *Nature*

2022 : publication sur les [principes FAIR pour les logiciels de recherche](#)



# POURQUOI rédiger un PGD ? Les **BÉNÉFICES** à attendre

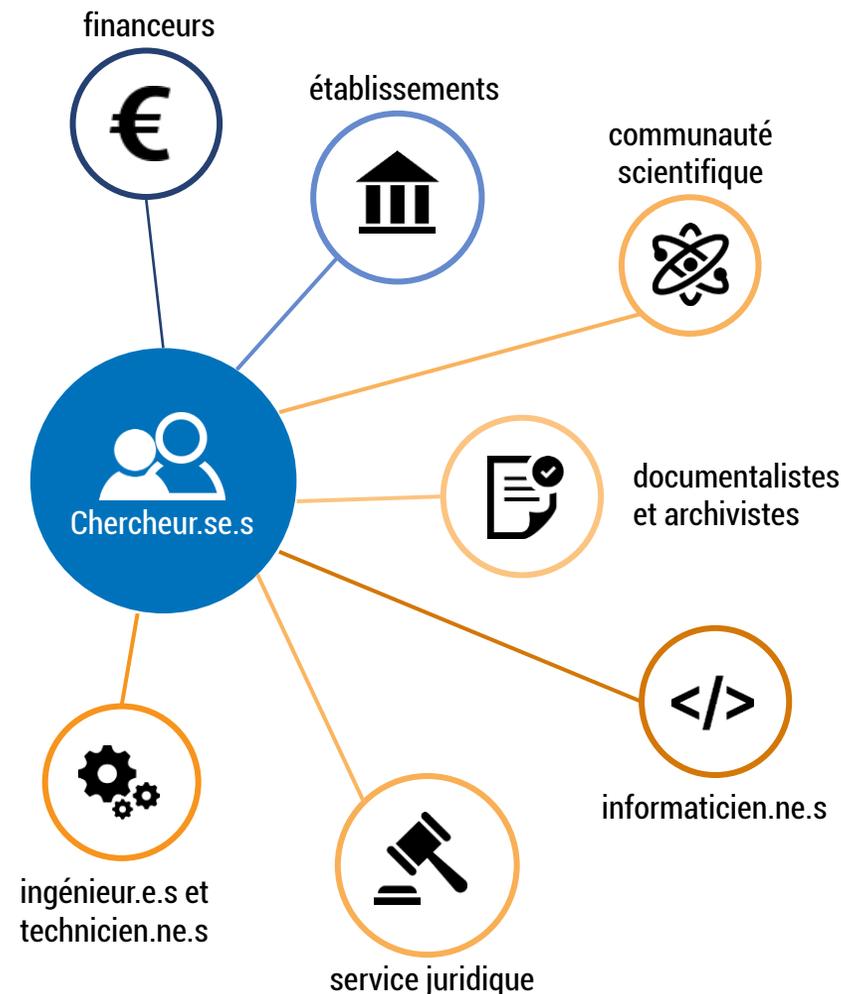
# Rédiger un DMP : Qui ? Quand ? Comment ?



## DMP et cycle de vie des données

Le DMP permet d'**organiser** et **anticiper** toutes les étapes du **cycle de vie** pour chaque jeu de données.

Il **accompagne** le projet tout au long du cycle de vie des données afin d'obtenir des **données FAIR**.



# Concrètement...

1 PROJET  
=  
1 DMP

Document **assez court**  
de quelques pages :  
rubriques avec des  
questions.

Un DMP est utile  
quelque soit la décision  
prise quant à **l'ouverture  
des données** (accès libre,  
restreint, fermé).  
Il mentionnera les  
raisons d'un non-  
partage

Pas de trame unique :  
**plusieurs modèles**  
(institutions, organismes,  
financeurs).  
Mais on y retrouve souvent  
les mêmes éléments.

# POURQUOI rédiger un DMP ?

## Outil de **GESTION** de **PROJET**

 Permet d'avoir une **réflexion collective**

 **Anticipe** les **coûts** et **points de vigilance**

 Clarifie les **rôles**, **responsabilités** et **droits** de chaque **contributeur·rice·s**

 Facilite le **partage** des **données** au sein de l'**équipe** du projet

 **Gain de temps** pour **publier** et **valoriser données** et **résultats**

## Obtention de **DONNÉES FAIR**

 Contribue à la **transparence**, à l'**intégrité scientifique** et à la **reproductibilité**

 **Anticipe** les éventuels problèmes **juridiques**, **éthiques** ou **techniques**

 Assure la **sécurité** des **données personnelles**, **sensibles** ou **stratégiques**

 Permet d'assurer la **pérennité** des données **coûteuses à reproduire** et/ou **non reproductibles**

 Favorise la **mise à disposition** des **jeux de données**

# Les 6 THÉMATIQUES du DMP

- 1 **DESCRIPTION** des données, de la **COLLECTE** ou de la **RÉUTILISATION** de données existantes
- 2 **DOCUMENTATION** et **QUALITÉ** des données
- 3 **STOCKAGE** et **SAUVEGARDE** pendant le processus/projet de recherche
- 4 Exigences **JURIDIQUES**, **ÉTHIQUES** et **CODES** de **CONDUITE**
- 5 **PARTAGE** des données et **CONSERVATION** sur le long terme
- 6 **RESPONSABILITÉS** et **RESSOURCES** en matière de gestion des données



# RÉDIGER son PGD avec l'outil DMP OPIDoR

# Où rédiger son DMP ?

Il existe  
différents outils  
pour rédiger son  
DMP :

Des outils et modèles  
anglophones :



[DCC](#)

*Digital Curation Center -  
Edimbourg*



[UC3](#)

*University of California  
Curation Center*

Recommandé pour l'ESR :

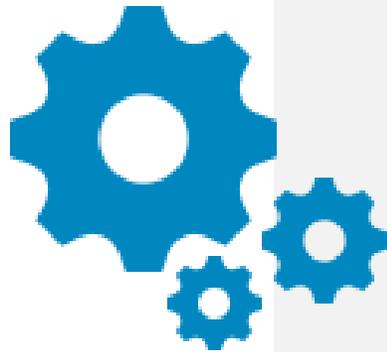


**DMP OPIDoR**

*Optimisation du Partage  
et de l'Interopérabilité  
des Données de la  
Recherche*

[INIST](#) CNRS

# DMP OPIDoR est un outil :



**Gratuit** : ouvert à toute la communauté de l'ESR

**Ouvert** : développé à partir du logiciel libre DMP Roadmap et code source disponible sur Github

**Évolutif** : évolue en fonction des remontées de utilisateur.rice.s

**Collaboratif** : partage d'un DMP aux collaborateur.rice.s d'un même projet

**Sécurisé** : hébergement sur des serveurs de l'INIST-CNRS (France)

# DMP OPIDoR : principales fonctionnalités



Création de DMP à partir de **modèles existants**



**Modèles personnalisables** par les organismes de recherche



Rédaction facilitée par affichage de **recommandations** liées au modèle choisi



**Partage** de DMP avec les collaborateur.rice.s



Ajout de **commentaires** par les collaborateur.rice.s



Définition du **niveau de visibilité** : 4 niveaux possibles

## Dernière version de DMP OPIDoR (11/2021) :



**Import automatique** de certaines informations, via le n° de projet ANR par exemple



**Nouvel onglet « contributeurs »**, qui prend en compte les **identifiants pérennes**



**Nouveau modèle structuré** selon le **cycle de vie des données** et conforme aux **principes FAIR** :

**Personnalisation des formulaires de saisie**

**Nouvel onglet budget**

**Téléchargement** du DMP au format .json (DMP « machine actionable »)

D'après : DoRANum - Parcours interactif sur la gestion des données de la recherche Disponible : <https://doranum.fr/de-la-recherche/enjeux-benefices/parcours-interactif-sur-la-gestion-des-donnees->



# En PRÉVISION de l'ATELIER PRATIQUE

# Vos « DEVOIRS » pour la SESSION #2 :

1

Se créer un compte  
sur DMP OPIDoR  
<https://dmp.opidor.fr/>

Si possible, lier son  
compte avec la  
fédération d'identité  
RENATER



## Modifier le profil

Renseignements | Mot de passe | Préférences des notifications

À noter que votre adresse électronique est utilisée comme nom d'utilisateur. Si vous modifiez cette information, n'oubliez pas d'utiliser votre nouvelle adresse pour vous connecter.

Vous pouvez modifier les renseignements ci-dessous.

\* **Courriel**

\* **Prénom**

\* **Nom**

\* **Organisme**

 ✕  
Mon organisme n'est pas listé.

**Langue**

 ▾

### Identifiants institutionnels

 [Lier vos identifiants institutionnels](#)

Enregistrer

# Vos « DEVOIRS » pour la SESSION #2 :

2

Réunir les informations nécessaires à la création de votre 1er PGD

## Renseignements sur le projet :

Possibilité de **récupération automatique des informations administratives des projets ANR** grâce notamment à l'identifiant ANR

- Titre du projet
- Acronyme du projet
- Résumé du projet
- Sources de financement du projet
- Partenaires associés au projet
- Coordinateur du projet
- Date de début du projet
- Date de fin du projet

## Vos « DEVOIRS » pour la SESSION #2 :

3

Réfléchir à la structuration de votre PGD :

- par « produit de recherche »
- par **work packages**
- tout ce qui peut sembler cohérent par rapport à votre projet

**Produit de recherche** : jeu de données, logiciel, protocole, échantillon...  
Nécessite une **gestion spécifique** à sa nature ou sa discipline.

Il s'agit des **différents types de données** que vous allez produire dans le cadre de votre projet.



**Typologie des produits de recherche** dans DMP OPIDoR :

- Audiovisuel
- Collection
- Jeu de données
- Image
- Ressource interactive
- Modèle
- Objet physique
- Service
- Logiciel
- Son
- Texte
- Workflow
- Autre

Voir : [https://dmp.opidor.fr/static/research\\_output\\_types](https://dmp.opidor.fr/static/research_output_types)

# ONGLET *Rédiger* → Les « TIPS »



DESCRIPTION des données, de la COLLECTE ou de la RÉUTILISATION de données existantes



## FOCUS :

Recenser l'**identité** des **contributeurs** et **contributeurices**  
→ utiliser des **identifiants pérennes** type ORCID

Lister les différents **types de données** produites et leur **nature**

Privilégier des **formats** de fichier **ouverts** et **non-propriétaires**  
Si impossible d'utiliser des formats ouverts, expliquer pourquoi

Citer / créditer les **données pré existantes** qui vont être utilisées



## RESSOURCES :

- [Fiche synthétique](#) sur les **identifiants pérennes** (source : DoRANum)
- Site d'[ORCID](#)

**Exemples** : Analyses génomiques, Images satellites, Relevés de températures, Transcription d'entretiens, etc...

Liste indicative de [formats ouvertes et fermés sur DoRANum](#)

- Utiliser la **fonction de citation** des portails d'exposition ”
- Pour les jeux de données pourvus d'un DOI, il existe l'outil [DOI Citation Formatter](#)
- **Logigramme** de l'**Institut Pasteur** (Émilie Cotte et Fanny Sébire, 2022) qui expose les [questions juridiques liées à la réutilisation de données](#)

# ONGLET *Rédiger* → Les « TIPS »



DESCRIPTION des données, de la COLLECTE ou de la RÉUTILISATION de données existantes



## FOCUS :

Décrire les **protocoles, unités de mesures, conditions expérimentales, protocoles de collecte, logiciels etc...** utilisés pour générer les données du projet

Identifier les **objets d'observation** et les **caractéristiques mesurées** sur ces objets

**Données personnelles** : données qui permettent **d'identifier une personne, directement** (ex : nom, prénom) **ou indirectement** (n° de compte en banque, adresse IP par ex.)

**Données sensibles** : données qui nécessitent des mesures de protection renforcées en raison des **risques importants d'atteintes à la vie privée** qu'elles représentent

**exemples** : données relatives à l'**origine ethnique** ; aux **opinions politiques**, aux **convictions religieuses** ou **philosophiques** ; à l'**appartenance syndicale** ; à l'**orientation sexuelle** ; données **génétiques** et **biométriques** ; données de **santé** ; données d'**infraction**



## RESSOURCES :

- Liste d'équipements, **plateaux techniques** pour la production ou l'acquisition de données sur [CatOPIDoR](#)
- **Plateforme locales** existantes : [GéoBFC](#) (MSH Dijon/MSHE), [Mimento](#) (FEMTO), [plateforme ADN](#) (MSH Dijon), [SHERPA](#) (MSHE), [PUDD](#) (MSH Dijon), [PEA<sup>2</sup>T](#) (Chrono environnement), [GISMO](#) (Biogéosciences), ...

Précautions particulières pour la **collecte et la gestion des données personnelles** → voir [slide 38](#)

# ONGLET *Rédiger* → Les « TIPS »



DESCRIPTION des données, de la COLLECTE ou de la RÉUTILISATION de données existantes



## FOCUS :

Mentionner les **centres de références** thématiques (Recherche Data Gouv) ou autres **infrastructures de recherche disciplinaires** sollicitées dans le cadre du projet de recherche

Utiliser des **vocabulaires standardisés** pour **décrire** les données



## RESSOURCES :

[Centres de références thématiques](#) de l'écosystème Recherche Data Gouv :

- **DATA TERRA** et **PNUB** pour le système Terre et environnement
- **CDS** pour l'astronomie et l'astrophysique
- **Huma-Num** et **PROGEDO** pour les SHS
- **IFB** pour la biologie-santé

- Des exemples de **thésaurus** et **vocabulaires** reconnus par les communautés disciplinaires → <https://search-data.ubfc.fr/help/referentials.php>

- Pour trouver des **vocabulaires standardisés** : [Ontobee](#)

# ONGLET *Rédiger* → Les « TIPS »



## DOCUMENTATION des données



### FOCUS :

Les **métadonnées** sont un « **Ensemble d'informations structurées** qui décrit, explicite, localise une ressource informationnelle, dans le but d'en **faciliter la recherche, l'usage, et la gestion.** »

Elles peuvent être soit « **embarquée** » dans le fichier (c'est le cas des photos par ex.) ou venir **en supplément** des données (dans un catalogue de métadonnées par ex.)

On peut **choisir** un **standard de métadonnées généraliste** (DublinCore par ex.) ou **spécifique à un domaine** (EML en écologie par ex.)

D'autres éléments de **documentation des données** :

- **Dictionnaire** des données
- Liste des **codes utilisés**
- Fichier « **Lisez-moi** »

Et encore, pensez à :

- **Convention de nommage** des fichiers
- **Définition** d'une **arborescence commune**



### RESSOURCES :

- Le [catalogue dat@UBFC](mailto:catalogue.dat@UBFC) vous permet de **décrire vos données** avec des **métadonnées fines** qui **respectent des standards internationaux** et font l'objet d'un travail de **curation** de la part des **documentalistes**.
- **Trouver des standards et vocabulaires** avec l'outil [FAIRsharing](#)
- Des [standards de métadonnées disciplinaires](#) sur le site du **DCC** (Digital Curation Center)

Bonnes pratiques de **nommage des fichiers** :

- **25** caractères maxi.
- Pas de caractères **spéciaux**
- Pas d'**espaces**
- **Dates** au format **ISO** (AAAAMMJJ)
- **N° de version** si besoin

# ONGLET *Rédiger* → Les « TIPS »



## QUALITÉ des données



### FOCUS :

Pistes possibles pour évoquer la **qualité des données** :

- Décrive la procédure de **contrôle qualité**
- **Vérification** par les pairs ?
- Utilisation d'un outil particulier ?
- **Étalonnage** des outils de mesure ?
- **Répétition** des mesures ?



### RESSOURCES :

Utiliser un carnet de laboratoire électronique (CLE)

→ **Une infographie** ludique pour explorer les cahiers de laboratoire (**DoRANum – DataCC – URFIST Méditerranée**)

<https://doranum.fr/environnement/cahiers-laboratoire/>

→ **Guide** du Comité Science Ouverte pour aider **au choix d'un CLE** :

[https://www.ouvrirelascience.fr/wp-content/uploads/2022/09/GT\\_ELN\\_%E2%80%93\\_Outil\\_Grille\\_de\\_criteres.pdf](https://www.ouvrirelascience.fr/wp-content/uploads/2022/09/GT_ELN_%E2%80%93_Outil_Grille_de_criteres.pdf)

+ **Grille de critères à télécharger** :

[https://www.ouvrirelascience.fr/wp-content/uploads/2022/09/GT\\_ELN\\_%E2%80%93\\_Outil\\_Grille\\_de\\_criteres.ods](https://www.ouvrirelascience.fr/wp-content/uploads/2022/09/GT_ELN_%E2%80%93_Outil_Grille_de_criteres.ods)

# ONGLET *Rédiger* → Les « TIPS »



Exigences JURIDIQUES, ÉTHIQUES, CODE de CONDUITE



## FOCUS :

Rappel sur la **définition** des **données à caractère personnel**  
[slide 34](#)

Si le projet de recherche amène à **collecter/traiter** des **données à caractère personnel**, penser à :

- Recueil du **consentement éclairé**
- **Anonymisation / pseudonymisation**
- **Contrôle** spécifique sur **accès/stockage** de ces données
- Techniques de **chiffrement/cryptage**

**Données de santé** : données **sensibles** particulièrement exposées aux **risques de piratage informatique**, avec des **conséquences importantes** pour les personnes concernées  
→ **Besoin accru de confidentialité et de protection**

Mentionner l'éventuelle prise de contact avec le·la **DPO**



## RESSOURCES :

**Plus d'infos et de liens** sur la page dédiée sur notre site :  
<https://data.ubfc.fr/pour-aller-plus-loin/aspects-juridiques> (rubrique « Données à caractère personnel »)

**Pour les SHS** : guide [InSHS sur la protection des données à caractère personnel dans le contexte de la Science ouverte](#)

**PSEUDO** : outil de **pseudonymisation** en ligne d'Etalab

**AMNESIA** : Service d'**OpenAIRE** pour l'**anonymisation** des **données de recherche sensibles**

Fiche pratique réalisée par l'**université de Lorraine**

→ [Les données de recherche dans le domaine de la santé](#)

+ [Synthèse de la CNIL sur le traitement des données de santé à des fins de recherche](#)

**Contacts** dans la « boîte à outils » [slide 49](#)

# ONGLET *Rédiger* → Les « TIPS »



Exigences JURIDIQUES, ÉTHIQUES, CODE de CONDUITE



## FOCUS :

Préciser les **conditions d'accès aux données** et les éventuelles **restrictions**, par exemple : **données personnelles/sensibles**, **soumission de brevet/innovation technologique**, **embargo** etc...

Y a-t-il des **accords de consortiums** ? Des droits de **propriété intellectuelle** ?

Envisager l'attribution de **licences** pour encadrer l'**accès** et la **réutilisation** des données



## RESSOURCES :

[Guide d'application de la Loi pour une République Numérique pour les données de la recherche](#)

[Ouverture des données de recherche : guide d'analyse du cadre juridique en France](#)

Logigrammes de l'Institut Pasteur (Émilie Cotte et Fanny Sébire, 2022) qui expose :

- les [questions juridiques](#) liées à la **diffusion de données**
- les [questions juridiques](#) liées à la **réutilisation de données**

Ressource interactive sur les [Licences de réutilisation dans le cadre de l'Open data et de la Loi pour une République Numérique](#) (DoRANum)

# ONGLET *Rédiger* → Les « TIPS »



Exigences JURIDIQUES, ÉTHIQUES, CODE de CONDUITE



## FOCUS :

Aspects **éthiques** à prendre en compte :

- Chartes de **déontologie**
- Protocoles ou codes de conduite liés à la **recherche sur le vivant**
- **Chartes locales** (chartes et feuilles de route Science ouverte)



## RESSOURCES :

Comité d'éthique sur la recherche à UBFC :

→ <https://www.ubfc.fr/recherche/cer-ubfc/>

Chartes et codes de déontologie de la recherche :

- [Charte française de déontologie des métiers de la recherche](#)
- [Code européen pour l'intégrité en recherche](#)
- [Protocole de Nagoya](#) sur l'accès aux ressources génétiques
- [Charte de déontologie et d'intégrité scientifique de l'ANR](#)

Chartes et feuilles de route Science ouverte des établissements UBFC :

# ONGLET *Rédiger* → Les « TIPS »



## TRAITEMENT et ANALYSE des DONNÉES



### FOCUS :

Décrire les **traitements** utilisés :

- Logiciels, paramètres définis, environnement, OS...
- Formats → privilégier les formats ouverts et non propriétaires
- Type de matériel



### RESSOURCES :

- **Liste de plateformes de calcul** (équipements, plateaux techniques) utilisées pour le traitement et l'analyse de données sur [CatOPIDoR](#)
- **Plateforme locales** existantes :
  - [GéoBFC](#) (MSH Dijon/MSHE)
  - [Mimento](#) (FEMTO)
  - [plateforme ADN](#) (MSH Dijon)
  - [SHERPA](#) (MSHE)
  - [PUDD](#) (MSH Dijon)
  - [PEA<sup>2</sup>T](#) (Chrono environnement)
  - [GISMO](#) (Biogéosciences), ...

# ONGLET *Rédiger* → Les « TIPS »



STOCKAGE et SAUVEGARDE des données PENDANT le projet



## FOCUS :

**STOCKAGE** : Enregistrement d'une information sur un **support physique** qui peut avoir des caractéristiques variées (clé USB, disque dur, serveur)

- Accès fréquent/rare aux données
- Partage avec les membres du projet
- Support externe (disque dur, clé USB) ou serveur
- Accès local (labo) ou distant (service type cloud)

**SAUVEGARDE** : Dupliquer des données pour les **mettre en sécurité** sur des **supports** de stockage différents, localisés dans des endroits différents (= **recopie à l'identique**)

## TEMPÉRATURE des données :

- Données **froides** : peu utilisées, on peut se permettre un temps d'accès un peu long (=disque dur)
- Données **chaudes** : données très actives, accès intensif (=espaces de stockage spécifique)



## RESSOURCES :

La **règle du 3-2-1** est recommandée par certaines universités anglaises ([Sheffield](#) et [Bath](#)) :

- 3 copies de vos données
- sur 2 supports différents
- avec au moins 1 copie à distance

Utilisation de **clouds privés** (AWS, Google Drive, DropBox...) :

- éventuellement possible si outil validé par l'ANSSI
- vérifier les prestataires qualifiés sur [SecNumCloud](#)
- **recours aux clouds privés interdit en IRR et ZRR**

Liste d'équipements et plateaux techniques utilisés pour le **stockage** et la **sauvegarde** sur [CatOPIDoR](#)

Plateformes locales proposées : **CCuB** (Dijon), **Mésocentre de Franche-Comté** (Besançon), **Plateforme ADN** (Dijon).

Autres plateforme locales : **Data-Chrono**, **clouds institutionnels**, ...

# ONGLET *Rédiger* → Les « TIPS »



## PARTAGE des DONNÉES



### FOCUS :

**ENTREPÔT de DONNÉES** : « Réservoir » de données de recherche qui peuvent être réutilisées grâce à la **documentation** de ces données (=métadonnées)

Justifier le choix de l'entrepôt :

- **Recommandations** de mon labo, établissement, financeur
- Entrepôt **disciplinaire** de **confiance**
- **Types, formats** et **volumétrie** de données acceptés
- Attribution d'**identifiants pérennes (DOI, Handle...)**
- Attribution de **licences**
- **Garanties** en termes de **sécurité** et d'**archivage**
- **Coût**
- **Souveraineté** sur les données (entrepôt situé en France ou en Europe)

Expliquer comment les données seront sélectionnées et nettoyées en vue du partage



### RESSOURCES :

Entrepôt et/ou catalogue pour déposer ou simplement décrire les données si elles ne peuvent pas être ouvertes → [dat@UBFC](mailto:dat@UBFC)

Liste d'entrepôts de données sur [CatOPIDoR](#)

Répertoire d'entrepôts de données [re3data](#) (registry of research data repositories)

Pour la **diffusion** et la **préservation** des **logiciels** et **codes sources** de la recherche, dépôt dans l'entrepôt **Software Heritage** (attribution d'un identifiant pérenne : le **SWHID**)

Si les données **ne peuvent pas être déposées dans un entrepôt**, **expliquer pourquoi** (données sensibles, qui font l'objet d'un brevet, embargo, etc...)

[OPEN REFINE](#) : logiciel libre de nettoyage et mise en forme de données

[Liste des éléments à vérifier](#) au moment de la **préparation des données pour le partage** (Doranum)

# ONGLET *Rédiger* → Les « TIPS »



## CONSERVATION des DONNÉES à LONG TERME



### FOCUS :

**ARCHIVAGE** : ensemble d'actions qui a pour but de **garantir sur le long terme l'accessibilité à des données** que l'on doit ou souhaite conserver pour des raisons **juridiques, historiques ou culturelles**.

**Archivage à long terme de contenus numériques :**

- Assurer la **pérennité** des **supports de stockage**
- Assurer l'**accès au contenu** même quand les **formats** de données deviennent **obsolètes**
- Assurer l'**intégrité** des données

**Conservation des données nécessaires la reproductibilité des résultats :**

- **Données de test** : à supprimer en fin de projet
- **Données de validation associées à une publication** : à garder 5 à 10 ans après la publication
- **données difficilement reproductibles** : à garder + de 10 ans après la publication
- **données non reproductibles ou à forte valeur** : à archiver

L'ANR demande la **conservation** des données de recherche **5 ans** après la fin du projet



### RESSOURCES :

**Archivage – opérateurs français :**

- [CINES](#) → prestataire pour l'ESR
- [Huma-Num](#)
- [Vitam](#) (programme interministériel d'archivage numérique)
- [Software Heritage](#) → pour les logiciels et codes sources

**Exemple de coûts** pour les services du **CINES** :

forfait accompagnement : **2 500€**

**1 043€** par an et par To archivé pour les projets < à 10 To

**221€** par an et par To archivé pour les projets > à 10 To

Le [NERC](#) a mis au point une « [checklist](#) » qui comprend **29 critères** répartis sur **8 thématiques** pour déterminer si une **conservation à long terme** des données se justifie

**Contrôler** si un **format** de fichier est éligible à la **conservation pérenne** sur la plateforme [FACILE du CINES](#)

# Vos SOURCES D'INFORMATION :

## ATELIER DE LA DONNÉE DAT@UBFC

DAT@UBFC PORTAIL ACCOMPAGNEMENT RESSOURCES ACTUALITÉS

Accueil » Pour aller plus loin

### Ressources

 Valorisation des données de recherche	 Principes FAIR	 PGD / DMP
 Entrepôts / Stockage	 Trouver des jeux de données	 Logiciels et codes sources
 Aspects juridiques	 Éthique et intégrité scientifique	 Identifiants pérennes

Site web du service dat@UBFC :  
<https://data.ubfc.fr>

# GESTION de vos DONNÉES de RECHERCHE : quelles FORMATIONS ?

En partenariat avec la MSHE :

**PLANIFIEZ** votre RECHERCHE et la GESTION de vos DONNÉES avec le PLAN de GESTION de DONNÉES

Jeu. 19/10 et ven. 10/11/23

MSHE C-N. Ledoux

Étudiant.e.s, doctorant.e.s et personnels

Partie #1 théorique : 1h00  
Partie #2 pratique : 2h00

**DOCUMENTEZ & VALORISEZ** vos DONNÉES de RECHERCHE : vers plus de VISIBILITÉ et de PÉRENNITÉ

Jeu. 15/02 et 07/03/2024

MSHE C-N. Ledoux

Étudiant.e.s, doctorant.e.s et personnels

Partie #1 théorique : 1h00  
Partie #2 pratique : 2h00

École doctorale, formation permanente et continue des personnels :

**PLANIFIEZ** votre RECHERCHE et la GESTION de vos DONNÉES avec le PLAN de GESTION de DONNÉES

Jeu. 25/01/24

DIJON & visio.

Doctorant.e.s & personnels

1h00

**INTRODUCTION** aux BASES de DONNÉES RELATIONNELLES

07 et 21/02/2024  
06, 13 et 20/03/24

BESANÇON & visio.

Doctorant.e.s et personnels

5 x 3h00

**OUVRIR** vos DONNÉES de RECHERCHE : dans quel CONTEXTE travaillez-vous ?

Jeu. 08/02/24

BESANÇON & visio.

Doctorant.e.s et personnels

1h30

**Soyez FAIR :** adoptez de BONNES PRATIQUES pour la GESTION de vos DONNÉES de RECHERCHE

Jeu. 14/03/24

BESANÇON & visio.

Doctorant.e.s et personnels

1h30

**DOCUMENTEZ & VALORISEZ** vos DONNÉES de RECHERCHE : vers plus de VISIBILITÉ et de PÉRENNITÉ

Jeu. 21/03/24

DIJON & visio.

Doctorant.e.s et personnels

1h00

**ARCHIVAGE à LONG TERME** de vos DONNÉES de RECHERCHE

Jeu. 16/05/24

DIJON & visio.

Doctorant.e.s et personnels

1h30



**MERCI**  
pour votre écoute !



Nous **CONTACTER**

[dataubfc-doc@ubfc.fr](mailto:dataubfc-doc@ubfc.fr)



Nous **SUIVRE**

[@dataUBFC](https://twitter.com/dataUBFC)



Point d'entrée **UNIQUE** sur les **DONNÉES** de la recherche à UBFC

<https://data.ubfc.fr/>



**CATALOGUE** des **DONNÉES** de la **RECHERCHE** en Bourgogne Franche-Comté

<https://search-data.ubfc.fr/>



Vendredi **10 novembre 2023**  
10h00

**PGD session #2**



Jeudis **15 février & 07 mars 2024**

**Documentez et valorisez vos données de recherche : vers plus de visibilité et de pérennité**



Source des **PICTOGRAMMES** utilisés :

<https://iconmonstr.com/>