

## Diagramme de Flux des Données de Recherche

### Mode d'emploi

[À propos du Diagramme de Flux des Données de Recherche](#)

[Qu'est-ce qu'un Diagramme de Flux des Données de Recherche ?](#)

[Pourquoi devrais-je concevoir mon propre Diagramme de Flux des Données de Recherche ?](#)

[Que dois-je prendre en compte pour concevoir mon Diagramme de Flux des Données de Recherche ?](#)

[Comment préparer mon Diagramme de Flux des Données de Recherche ?](#)

[Comment utiliser le modèle et l'outil du Diagramme de Flux des Données de Recherche ?](#)

[Exemples de Diagramme de Flux des Données de Recherche](#)

### À propos du Diagramme de Flux des Données de Recherche

Le Diagramme de Flux des Données de Recherche (DFDR) est proposé par la Faculté de médecine. Le DFDR a été élaboré par les membres du groupe de travail **Stockage des données** du **Projet Data**. Cet outil est à destination principale des responsables de laboratoire ou de plateformes facultaires.

Le DFDR s'appuie sur le [cycle de vie des données](#) proposé par le RDMkit. RDMkit est la boîte à outils ELIXIR pour la gestion des données de recherche en sciences de la vie (<https://rdmkit.elixir-europe.org>). Le DFDR a été conçu avec le logiciel [Draw.io](#). Draw.io est une marque déposée de JGraph Ltd (Angleterre) et draw.io AG (Suisse).

### Qu'est-ce qu'un Diagramme de Flux des Données de Recherche ?

Le DFDR s'adresse aux groupes de recherche et aux plateformes de recherche. Le but du DFDR est triple :

- Instrument de réflexion pour évaluer le cycle de vie des données de recherche dans votre laboratoire
- Document d'aide à la décision pour mettre en œuvre de meilleures stratégies de gestion des données
- Outil pour identifier les besoins d'amélioration en termes de pratiques et d'infrastructure

Le DFDR est une représentation simple et efficace du cycle de vie de vos données de recherche.

### Pourquoi devrais-je concevoir mon Diagramme de Flux des Données de Recherche ?

- **Vue d'ensemble des pratiques des laboratoires**

Votre groupe génère et traite d'importantes quantités et une large variété de types de données. Les outils, technologies et votre programme de recherche évoluent. Vous avez besoin d'une vue d'ensemble des pratiques de gestion des données de recherche de votre groupe pour garantir leur alignement avec vos objectifs, ainsi qu'avec les recommandations institutionnelles et les obligations légales.

- **Homogénéité des pratiques**

Vous souhaitez que tous les membres de votre groupe de recherche ou de votre plateforme aient le même niveau d'information sur la gestion des données. Vous voulez que l'ensemble du groupe suive les mêmes bonnes pratiques, et faciliter l'intégration des nouveaux membres. Vous avez besoin d'un cadre commun pour retrouver facilement les données.

- **Identification des axes d'amélioration**

En réfléchissant à votre DFDR, vous pouvez constater que vous avez trop de copies des mêmes jeux de données, ou mettre en évidence des transferts de données lents. Le DFDR est votre outil d'entrée pour améliorer vos pratiques de gestion des données de recherche et pour interagir avec les Services de support concernés.

- **Préparation de DMP et communication avec les éditeurs/trices de revues**

Votre DFDR est un atout précieux pour préparer le Data Management Plan (DMP) demandé par le FNS ou exigé par la CUREG. Toutes les informations sont rassemblées dans un diagramme simple. Votre DFDR peut également être utile lors de communications avec les éditeurs/trices de revues scientifiques.

### Que dois-je prendre en compte pour concevoir mon Diagramme de Flux des Données de Recherche ?

La conception de votre DFDR est une réflexion sur les pratiques de votre laboratoire ou de votre plateforme. La première chose à considérer est de consacrer suffisamment de temps à cette réflexion.

Vous pouvez aussi envisager d'inclure plusieurs personnes dans la phase de réflexion : votre vision des pratiques ne correspond peut-être pas à la réalité des faits.

### Comment préparer mon Diagramme de Flux des Données de Recherche ?

La question très importante à laquelle vous devez répondre en premier est : « **est ce que je traite des données à caractère personnelle et/ou sensible ?** » Cette question est loin d'être évidente. Pour plus d'informations, vous pouvez consulter la définition des données personnelles ou sensibles et les recommandations éthiques et obligations légales à l'UNIGE, les commentaires de la CUREG2.0, et les lignes directrices de swissethics. Les données sensibles et non sensibles doivent être traitées séparément.

Vous pouvez commencer par élaborer une version préliminaire de DFDR sur une feuille de papier ou un tableau blanc. Il est conseillé de distinguer séparément le lieu de stockage des données (où les données sont physiquement) et la nature des données (par exemple : données brutes, données analysées), et de se concentrer dans un premier temps sur la nature des données. Votre flux de données suivra très probablement une partie du cycle de vie des données.



Source : Data Life Cycle du RDMkit ([https://rdmkit.elixir-europe.org/data\\_life\\_cycle](https://rdmkit.elixir-europe.org/data_life_cycle)). RDMkit : la boîte à outils ELIXIR pour la gestion des données de recherche en sciences de la vie (<https://rdmkit.elixir-europe.org>).

## Comment utiliser le modèle et l'outil du Diagramme de Flux des Données de Recherche ?

Le diagramme est réalisé à l'aide de Draw.io. Cette solution gratuite et open source a été choisie car elle est très intuitive et facile à manipuler. Vous retrouverez davantage d'information sur la [documentation](#) de l'outil.

- **Outil DRAW.io et accessoires**

Draw.io existe en deux versions : un [logiciel téléchargeable](#) pour une installation en local, ou une application web accessible à l'adresse <https://app.diagrams.net/?src=about#>

Une fois le programme installé ou l'application web ouverte, vous devez importer les icônes que vous utiliserez. Dans le panneau de gauche, cliquez sur le crayon de la section « scratchpad », puis importez le fichier [Drawio workflow icons](#). Enfin, vous pouvez télécharger un modèle de départ :

- [Sans données sensibles – vide](#)
- [Sans données sensibles – prérempli](#)
- [Avec données sensibles – vide](#)
- [Avec données sensibles – prérempli](#)

Draw.io est très intuitif. Par exemple :

- Un double clic permet de créer un nouvel objet (par exemple un rectangle). Vous pouvez écrire à l'intérieur de ce nouvel objet (par exemple Données brutes).
- Pointer le curseur sur le contour d'un objet fait apparaître un point vert ; faites glisser ce point vert vers un autre objet et il reliera les deux objets par une flèche.
- Faites glisser une icône depuis le scratchpad (par exemple « attention » ou « métadonnées ») vers votre diagramme pour documenter votre DFDR.

Maintenant que tout est configuré et que vous maîtrisez Draw.io, vous pouvez concevoir votre DFDR.

- **Conception du DFDR**

Horizontalement, le DFDR suit le flux des données dans le temps, de gauche à droite.

Verticalement, les données placées en haut du DFDR sont stockées pour une courte période tandis que les données placées en bas sont stockées pour une longue période.

Dans le cas d'un **DFDR avec données sensibles**, les données non sensibles (en haut) et sensibles (en bas) sont séparées. Il est possible que des données sensibles deviennent non sensibles au cours du flux (désidentification / anonymisation), et relient ainsi les parties non sensibles et sensibles du diagramme (exemple ci-dessous).

Il est conseillé de se concentrer d'abord sur la nature des données à chaque étape du cycle de vie des données et sur la durée de stockage correspondante, et de compléter ensuite le lieu de stockage des données.

Vous pouvez annoter et commenter le diagramme pour en faciliter la compréhension. Vous pouvez également utiliser des flèches rouges pour mettre en évidence des transferts de données problématiques.

- **Quelques définitions :**

**Des termes qui doivent être définis :**



- **Back-up** : données temporairement sauvegardées.
- **Preserve** : la préservation des données consiste en une série d'activités nécessaires pour garantir la sécurité, l'intégrité et l'accès des données aussi longtemps que nécessaire. Ne doit pas être utilisée à la place de Back-up.
- **Share** : partage des données avec des partenaires de collaboration ou avec la communauté et la société en général.

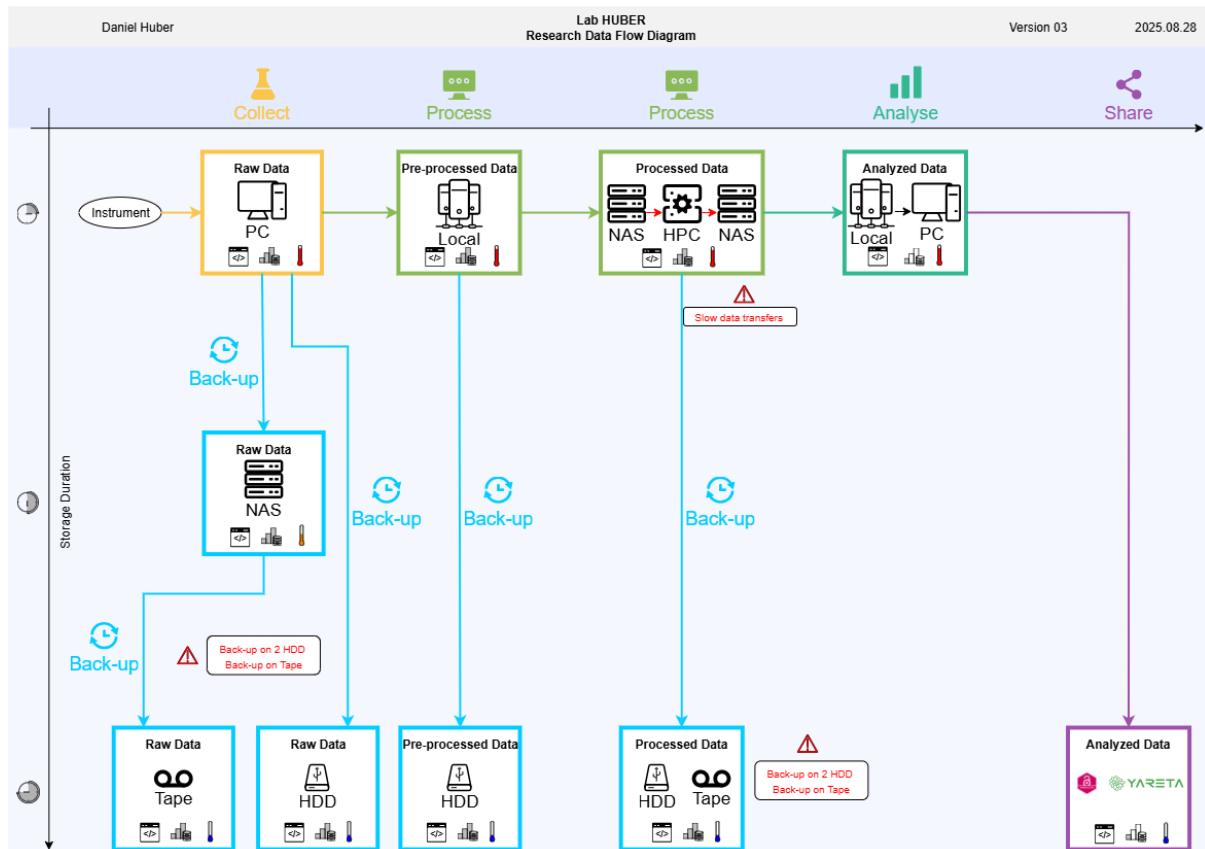
- Data temperature 🌡️: fréquence d'accès aux données (de froid = fréquence faible, jusqu'à chaud = fréquence haute)
- Challenges ⚠️: tout ce que vous jugez intéressant de mentionner
- Metadata 📁: « données sur les données », elles décrivent tout ce qu'un nouvel « utilisateur des données » aurait besoin de savoir pour trouver, comprendre, reproduire et réutiliser les données.

#### Vous devrez définir aussi :

- Storage duration ⏳: vous devrez définir ce que vous considérez comme stockage à court, moyen et long terme.
- Data volume 📈: vous devrez définir ce que vous considérez comme données de faible, moyen et fort volume.
- HPC HPC: vous devez spécifier quel cluster est utilisé : Baobab, Yggdrasil or Bamboo

#### Exemples de Diagramme de Flux des Données de Recherche

##### ▪ DFDR sans données sensibles



#### ▪ DFDR avec données sensibles

