



Ministère de l'Enseignement Supérieur et de la Recherche Scientifique
Universitaire de Mila
Institut des Sciences de la Nature et la Vie
Département des science biologiques et agricoles
Master académique
Spécialité : Biochimie Appliquée



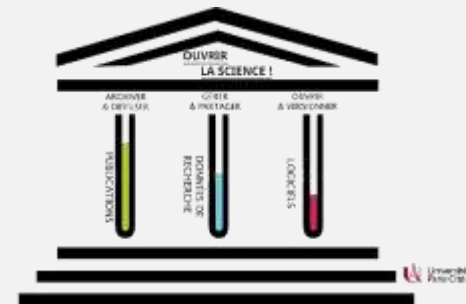
Intitulé de la matière : Logiciels libres et Open-Sources 2

1

CHAPITRE 01:

Open science et gestion avancée des données

Préparé par M. Kamel Belmehboul



Année univ. 2025-2026

1. Définition et enjeux de l'open science

2

La science ouverte (Open Science) est un mouvement mondial qui vise à rendre la recherche scientifique plus transparente, accessible et collaborative.

Elle repose sur l'idée que les connaissances scientifiques sont un bien commun de l'humanité, et qu'elles doivent être partagées librement pour favoriser l'innovation, l'efficacité et la reproductibilité des résultats.

Dans un contexte où la recherche devient de plus en plus numérique et interconnectée, l'Open Science transforme la manière dont les chercheurs produisent, partagent et valorisent leurs travaux.

1. 2 Défis et perspectives d'avenir

3

Malgré ses avantages, la mise en œuvre de la science ouverte rencontre plusieurs défis :

- **Défis techniques** : nécessité de plateformes fiables, sécurisées et interopérables pour stocker et partager les données.
- **Défis juridiques et éthiques** : gestion des droits d'auteur, respect de la vie privée, protection des données sensibles (notamment en biologie et en santé).
- **Défis culturels** : certains chercheurs hésitent encore à partager leurs résultats, craignant la concurrence ou la perte de reconnaissance individuelle.
- **Défis financiers** : la mise en place d'infrastructures ouvertes exige des investissements publics importants.

1. 3 Exemples d'application

4

Dans le domaine de la biologie, la science ouverte a révolutionné la recherche :

- Les bases de données comme **GenBank** ou **Protein Data Bank** permettent aux chercheurs du monde entier d'accéder librement à des millions de séquences d'ADN et de protéines.
- Les plateformes collaboratives facilitent la comparaison et la validation croisée des résultats expérimentaux.
- Le partage des protocoles de laboratoire (via **Protocols.io**, par exemple) renforce la reproductibilité des expériences.

1. 4 Principes fondamentaux de la Science Ouverte

5

La **science ouverte** repose sur :

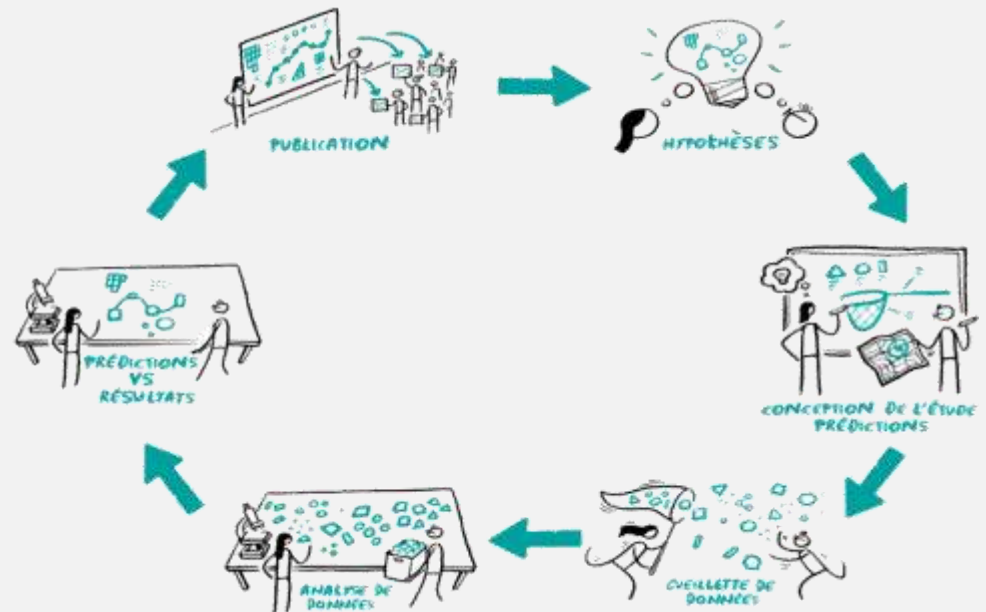
- **L'accès ouvert** : diffusion gratuite des articles via des plateformes comme **HAL** ou **Pub Med Central**.
- **Les données ouvertes** : partage des données de recherche (ex. : résultats biologiques sur **GenBank**) pour vérification et réutilisation.
- **Les logiciels libres** : outils open-source (**R**, **Python**, **Git Hub**) favorisant la transparence et la reproductibilité.
- **La reproductibilité scientifique** : documentation complète des méthodes et protocoles pour permettre la vérification des résultats.
- **La collaboration citoyenne** : implication du public et des institutions dans la recherche (science participative).

2. Principes de la reproductibilité scientifique

6

- La **reproductibilité scientifique** signifie qu'un autre chercheur doit pouvoir **obtenir les mêmes résultats** en suivant les **mêmes méthodes**.

C'est un principe essentiel pour garantir la **fiabilité** et la **crédibilité** des recherches.



2. 1 Principes clés

7

- **Documentation complète** : décrire clairement toutes les étapes de l'expérience (méthodes, outils, logiciels, paramètres).
- **Partage des données** : rendre accessibles les jeux de données utilisés pour permettre la vérification.
- **Partage du code** : publier les scripts ou programmes (ex. : sur GitHub) pour reproduire les analyses.
- **Utilisation d'outils ouverts** : privilégier les logiciels open-source (R, Python, Jupyter).
- **Workflow reproductible** : organiser les étapes du travail de façon traçable et transparente.

2. 2 Exemple simplifié

8

En biologie :

- Des chercheurs ont étudié l'effet d'une nouvelle protéine sur la croissance des cellules cancéreuses. Ils ont publié leurs résultats avec tous les détails nécessaires : type de cellules, concentration de la protéine, durée de l'expérience, température et milieu nutritif.

La reproductibilité consiste à permettre à d'autres chercheurs de répéter exactement la même expérience avec les mêmes conditions. Si les résultats sont similaires, cela confirme la validité scientifique de l'étude. Sinon, des doutes peuvent apparaître sur la fiabilité des conclusions.

Objectif : garantir que les découvertes biologiques sont réelles et non dues au hasard.



2. 3 Exemple simplifié

9

En intelligence artificielle :

- Un groupe de chercheurs développe un modèle d'intelligence artificielle pour classer des images d'arbres en 'sains' ou 'malades'. Ils publient leur article avec le code source, les données d'entraînement et les paramètres utilisés.

D'autres chercheurs téléchargent le code et les données, puis reproduisent le même entraînement. S'ils obtiennent la même précision (par exemple 92%), les résultats sont considérés comme reproductibles. Sinon, cela indique un manque de transparence ou de documentation scientifique.



Objectif : vérifier que le modèle fonctionne réellement comme décrit et non par des réglages cachés.

3. Format ouvert et interopérabilité des données

10

- **Format ouvert :**

Un **format ouvert** est un type de fichier dont les spécifications sont publiques et accessibles à tous.

Cela veut dire que **tout le monde peut l'utiliser, le lire ou le modifier**, sans payer de licence ou utiliser un logiciel propriétaire

3. 1 Exemples de Formas ouverts

11

- CSV (au lieu d'Excel .xlsx) pour les tableaux de données,
- .TXT ou .Odt .Doc pour les textes.
- .PNG ou .SVG pour les images.
- .JSON ou .XML pour données structures.

Les formats ouverts assurent que les données peuvent être **lues dans le futur**, même si un logiciel commercial disparaît



3. 2 Interopérabilité des données

12

- **L'interopérabilité** signifie que **les données peuvent être échangées, comprises et utilisées facilement** entre différents systèmes, laboratoires ou logiciels.
Pour cela, les données doivent suivre **des standards communs** (format, métadonnées, vocabulaire, unités, etc.)
- **Exemple :**
Dans un projet de biologie, si plusieurs laboratoires partagent leurs données sur les gènes au format **CSV ou JSON** avec les mêmes noms de colonnes et unités, ils peuvent **fusionner leurs résultats et les comparer facilement**.

4. Workflow collaboratif

13

Le workflow collaboratif:

- Le workflow collaboratif désigne un ensemble organisé de processus et d'outils permettant à plusieurs personnes de travailler ensemble sur un même projet, de manière coordonnée et transparente.

Il vise à structurer les étapes du travail collectif, à attribuer les responsabilités et à assurer la traçabilité des contributions.

4. 1 Exemples d'application

14

En biologie :

- Un groupe de chercheurs en biologie peut utiliser **un workflow** collaboratif pour gérer un projet de **séquençage génétique**.

L'un collecte les échantillons, un autre effectue les analyses, et un troisième compile les résultats sur une plateforme partagée comme

Git Hub ou **OSF (Open Science Framework)**.

4. 2 Workflow collaboratif avec Git et Git Hub

15

Pourquoi Git/GitHub pour des biologistes ?

- **Versionner vos données et scripts** : revenir aux versions précédentes.
- **Collaborer** sans écraser le travail des autres (chacun travaille sur sa copie puis fusionne).
- **Traçabilité** : qui a changé quoi (messages commit).
- **Reproductibilité** : stocker code + données + README.



git

GitHub



4. 3 Git

16

- Git est un **outil de gestion de versions**.

Il permet d'enregistrer toutes les modifications faites sur un projet (fichiers, code, données).

Grâce à **Git**, on peut :

- Sauvegarder l'historique du travail.
- Revenir à une version précédente.
- Travailler à plusieurs sans écraser le travail des autres.

Commandes de base :

- **git init** : créer un nouveau dépôt Git (dossier versionné).
- **git add fichier** : ajouter un fichier à la zone de sauvegarde.
- **git commit -m "message"** : enregistrer les changements avec un message.
- **git log** : voir l'historique des versions.

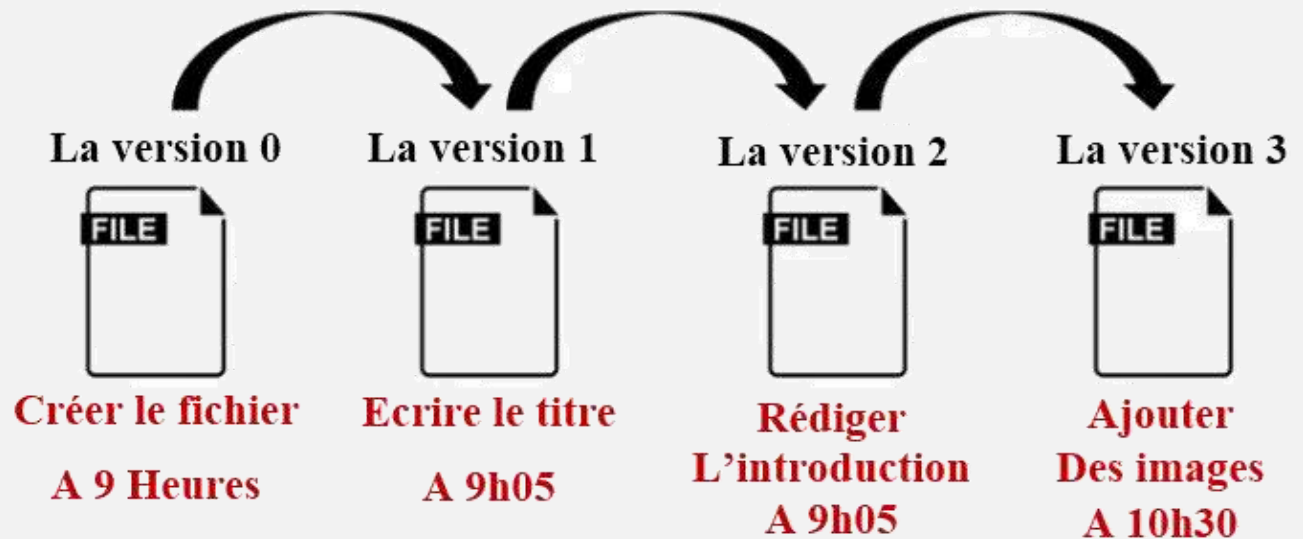


4. 3.1 Git

17

La fonction du **GIT**

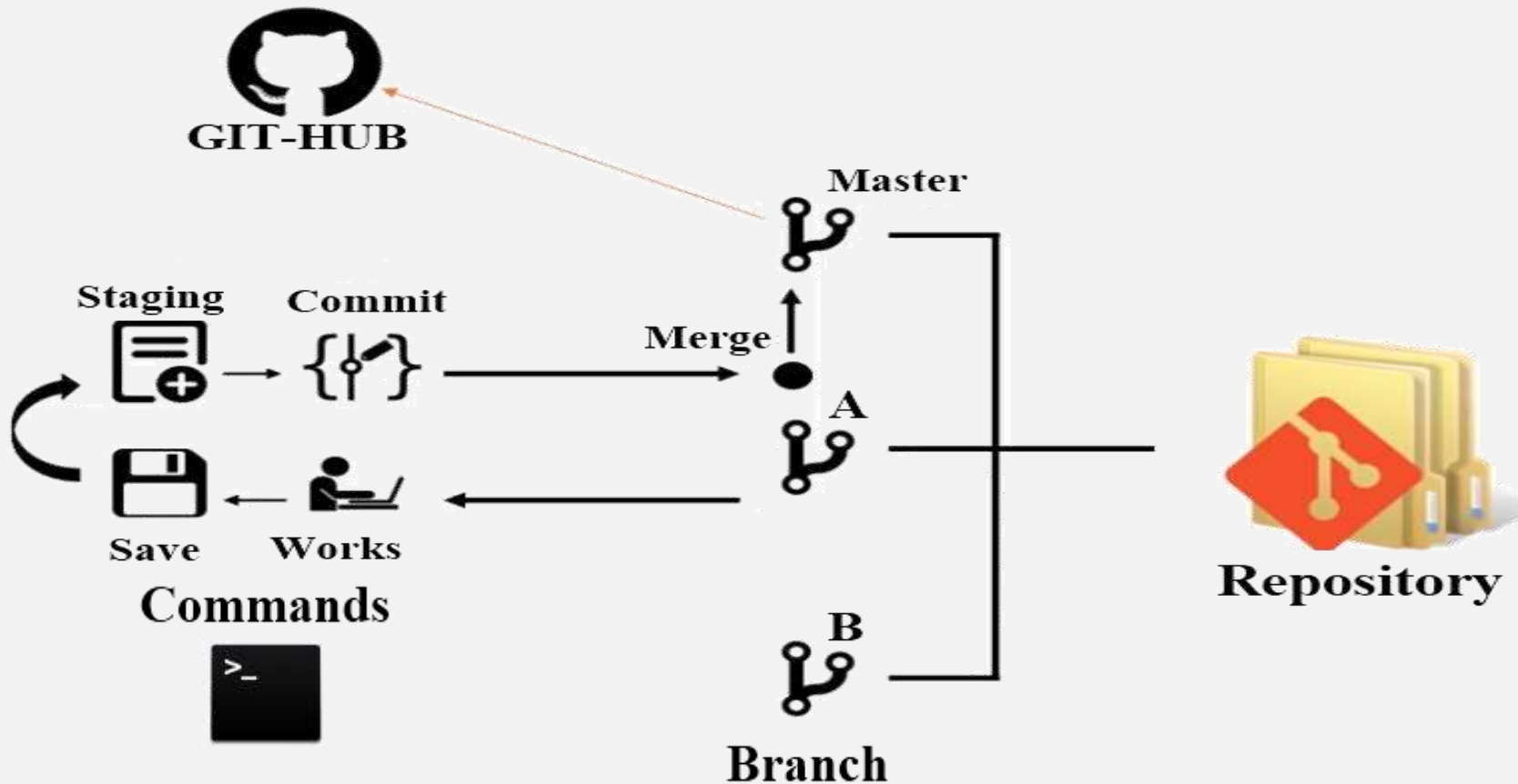
Suivre des fichiers spécifiques en prenant une capture de leur état lorsque l'utilisateur le demande



4. 3.2 Git

18

- Les éléments et le fonctionnement du GIT :



4. 4 Git Hub

19

- Git Hub est une **plateforme en ligne** qui héberge les projets Git.
Elle permet de :
- Sauvegarder le projet sur le cloud.
- Collaborer avec d'autres chercheurs.
- Partager le code ou les données de manière ouverte (Open Science).

Commandes utiles :

- `git remote add origin lien_du_dépôt` : connecter votre dépôt local à GitHub.
- `git push` : envoyer les modifications sur GitHub.
- `git pull` : récupérer les changements faits par les autres.

4. 5 Exemple concret pour les biologistes

20

Un groupe de chercheurs en biologie travaille sur des **données génétiques**.

- Chaque membre analyse une partie des données.
- Grâce à Git et GitHub, ils partagent leurs scripts et résultats.
- Si une erreur se produit, ils peuvent revenir à une version précédente sans perdre le travail collectif.

4. 6 Explication pratique étape par étape

21

Étape 1 : Installer Git

- Ouvrez votre navigateur et allez sur :

<https://git-scm.com>

Étape 2 : Créer un dossier de projet

Exemple : myproject

Ajoutez un fichier dedans (ex. : index.txt).

Étape 3 : Ouvrir le Terminal dans le dossier

Clic droit → Open in Terminal Ou Windows PowerShell

`cd path/to/myproject`



4. 7 Explication pratique étape par étape

22

Étape 4 : Initialiser Git dans le projet

- Tapez :

`git init`

Git commence maintenant à suivre votre projet.

Étape 5 : Ajouter les fichiers à Git

Tapez :

`git add .`

Le point signifie : ajouter tous les fichiers.

4. 8 Explication pratique étape par étape

23

Étape 6 : Enregistrer une version (commit)

- Tapez :
`git commit -m "first version"`

4. 9 Explication pratique étape par étape

24

- **Passons maintenant à Git Hub:**

Étape 7 : Créer un compte Git Hub

- Rendez-vous sur <https://github.com>
- Créez un compte.
- Cliquez sur New **Repository**
- Donnez un nom, par exemple : **myproject**
- Cliquez sur **Create Repository**

4. 10 Explication pratique étape par étape

25

Étape 8 : Connecter le projet local à Git Hub

- Copiez le lien affiché, par exemple :

<https://github.com/username/myproject.git>

Puis tapez :

git *remote add origin* <https://github.com/username/myproject.git>

Étape 9 : Envoyer le projet vers Git Hub

Tapez : **git** *push -u origin main*

Ou, si votre branche s'appelle "master" :

git *push -u origin master*

Fin cours