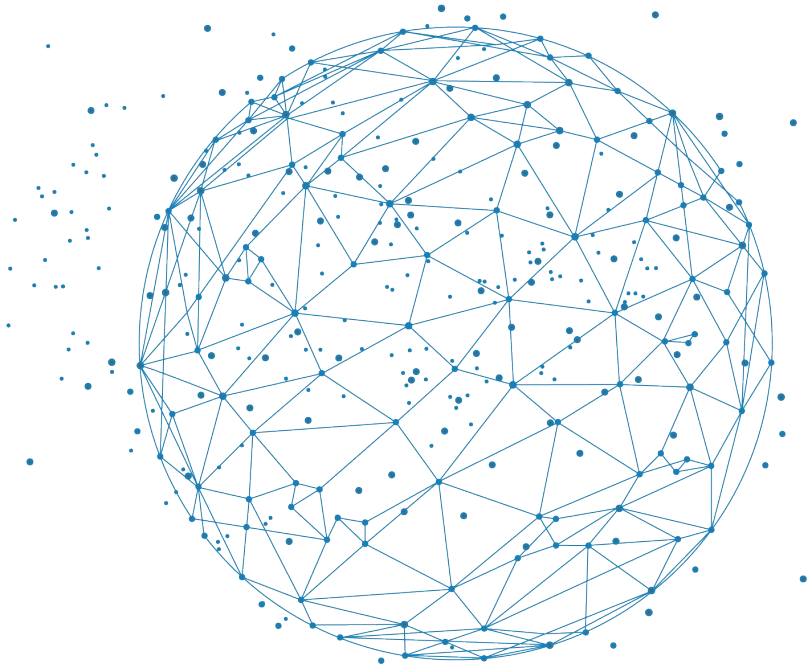
FAIR DATA IDF - 22 & 23 mars 2022



Module 2

Pratiques d'hygiène numérique pour la gestion des données au cours du projet de recherche

Valentin Loux & Cédric Midoux

ARTbio IBPS

migale

INSTITUT PASTEUR

PLANT BIOINFORMATICS FACILITY

IFB INSTITUT FRANÇAIS DE BIOINFORMATIQUE

elixir FRANCE

cnrs

INRAE

Inserm

cea

INVESTIR L'AVENIR

Les intervenants du module 2

- Valentin Loux
  - @vloux
  - https://orcid.org/0000-0002-8268-915X

- Cédric Midoux
  - @CedricMidoux
  - https://orcid.org/0000-0002-7964-0929

- Slides inspirées de
  - Fred de Lamotte - Montpellier
  - Julien Seiler - Strasbourg

Programme

- Intro
- Sécurité
- Traçabilité / Réutilisation
- Outils et solutions
- Un dernier exercice pour la route …

IFB
INSTITUT FRANÇAIS DE BIOINFORMATIQUE

elixir FRANCE

Module 2

Pratiques d'hygiène numérique pour la gestion des données

Intro & rappels

ARTbio IBPS

migale

INSTITUT PASTEUR

PLANT BIOINFORMATICS FACILITY

cnrs

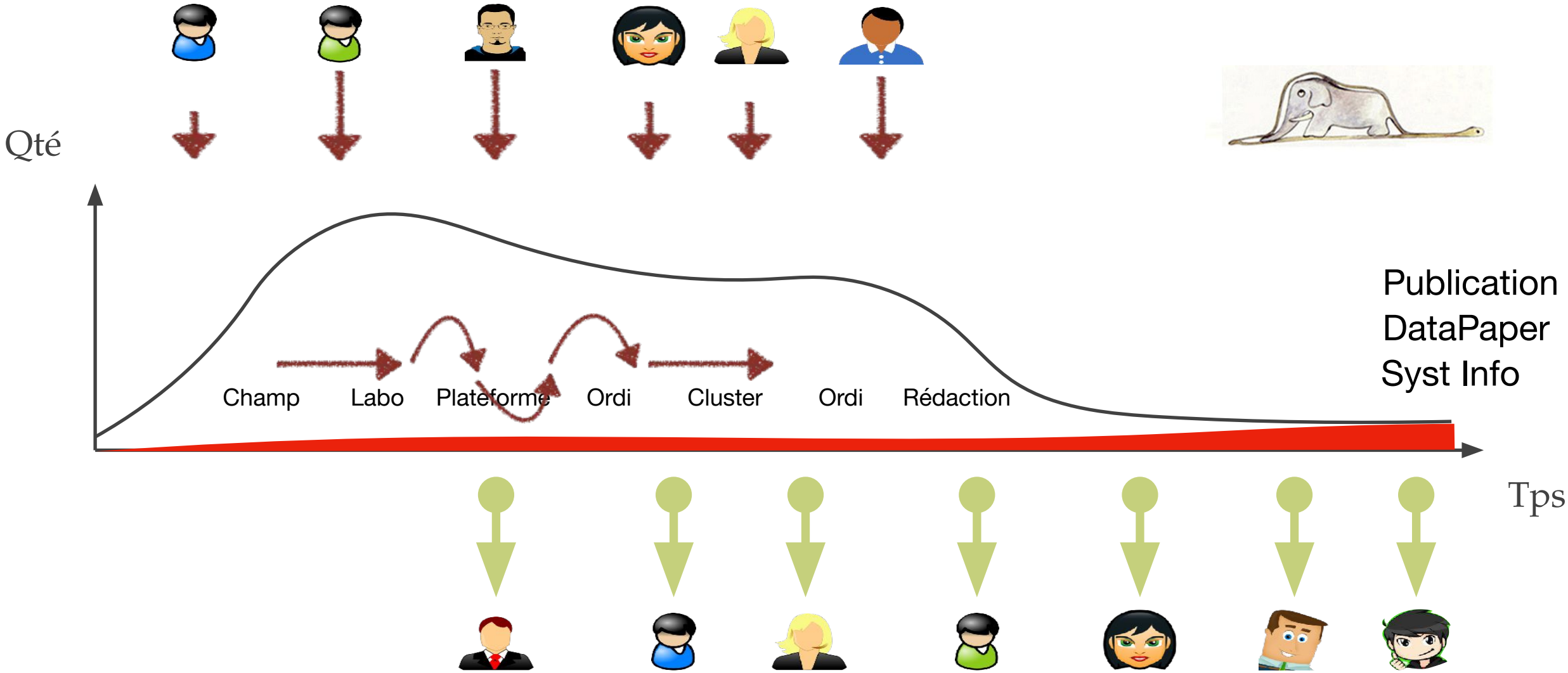
INRAE

Inserm

cea

INVESTIR L'AVENIR

Rappel : un projet sur la durée



Cycle de vie des données

Source

Plusieurs personnes

Plusieurs techniques

Plusieurs lieux

Plusieurs années

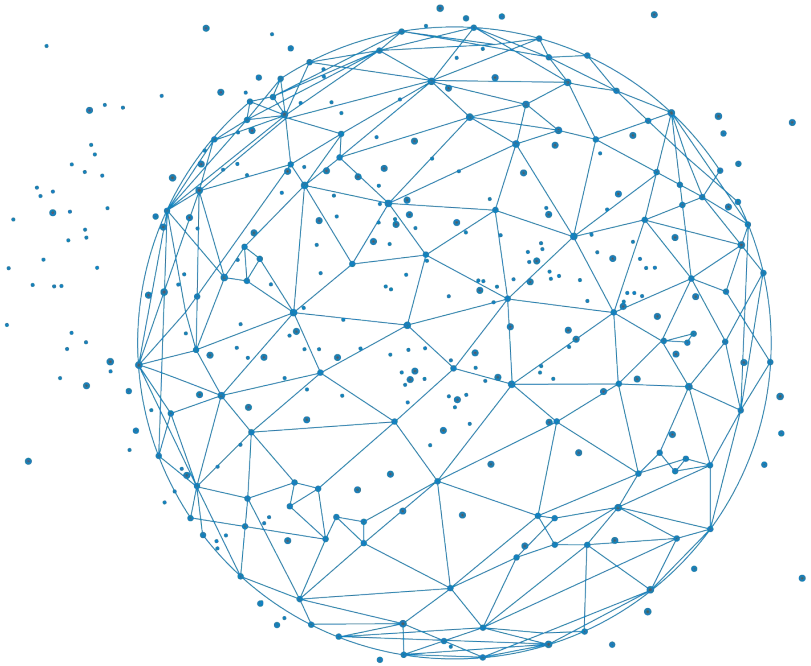
Ne rien perdre

Pouvoir retrouver

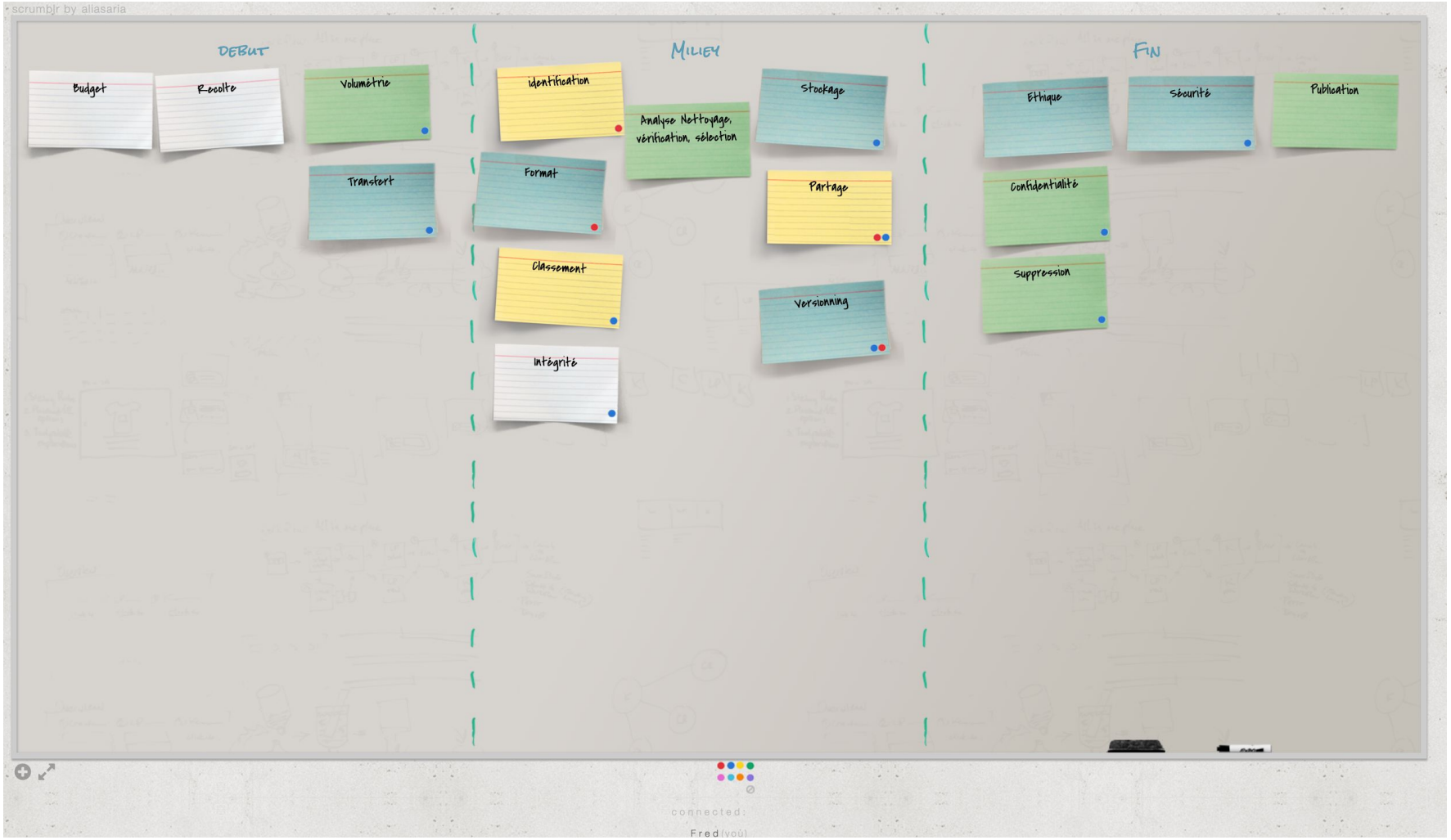
Pouvoir réanalyser

Pouvoir partager

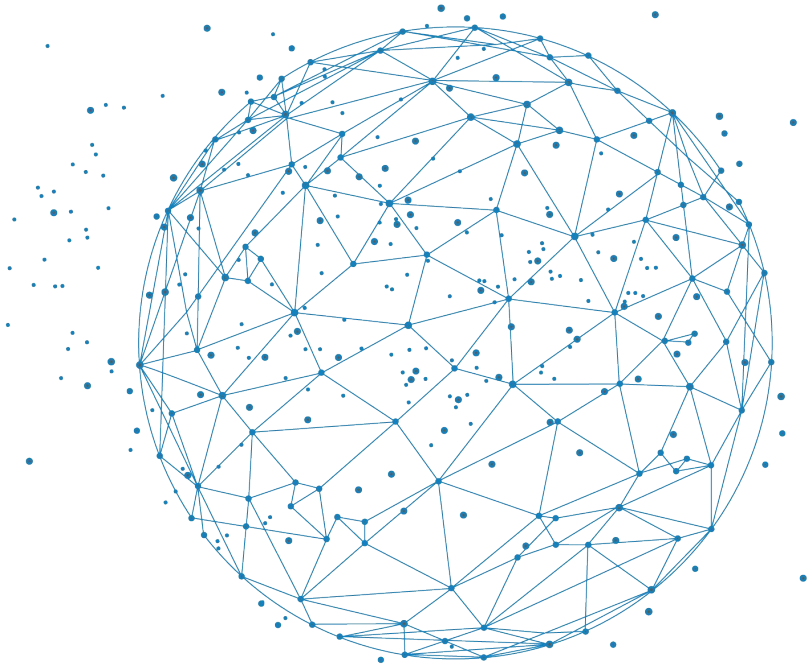
La vie des données



La vie des données le long du projet



Un cas d’usage



Un cas d’usage

Tout au long de cette session nous allons vous présenter des bonnes pratiques et des outils en nous mettant en situation au travers d'un cas d'usage.

Pour ce faire, imaginons que nous sommes une équipe de recherche et que nous souhaitons démarrer un nouveau projet de recherche.

Ce projet, nécessitera de mener de nombreuses expérimentations et acquisition de données diverses.

Nous espérons également qu'il nous permettra de proposer quelques bons papiers.

Nous nous efforcerons également tout au long du projet de garder en tête les principes FAIR que nous souhaiterons notamment mettre en oeuvre au travers de la publication de données

Attention, cette formation contient du placement de produits :-)

IFB
INSTITUT FRANÇAIS DE BIOINFORMATIQUE

elixir FRANCE

Module 2

Pratiques d'hygiène numérique pour la gestion des données

Sécurité des données

ARTbio IBPS

migale

INSTITUT PASTEUR

PLANT BIOINFORMATICS FACILITY

cnrs

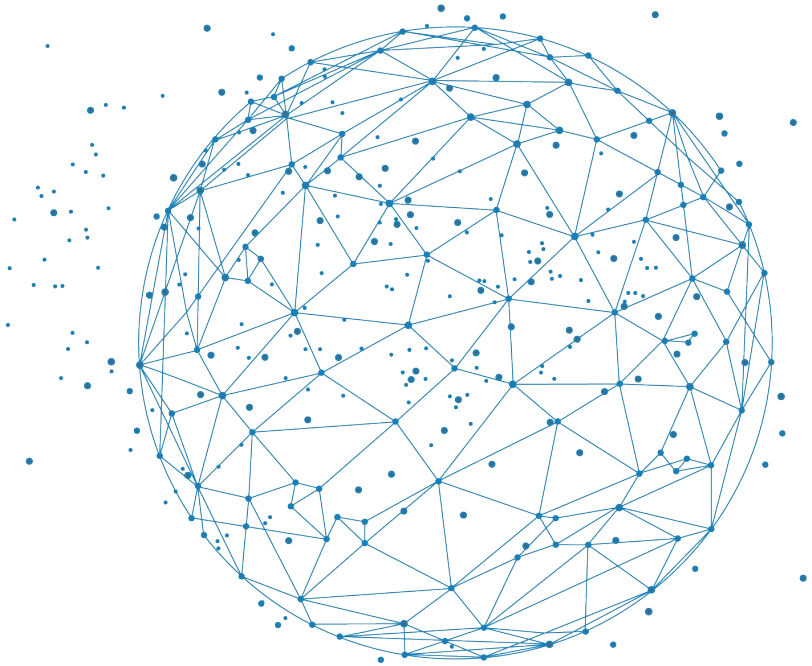
INRAE

Inserm

cea

INVESTIR L'AVENIR

Notre espace de stockage



Situation :

Je reçois un matin une clef USB par la poste, je m’aperçois que c’est les données critiques que Sam Lee me promet depuis 7 mois.

Que pensez vous de la façon de faire de Sam Lee ?

Fonction fondamentale : **la conservation des données**

Stockage :

- désigne des méthodes et des technologies permettant de conserver des données
- concerne tous les types de supports de stockage de masse (DD, Clé USB…) ou support de stockage dématérialisé (cloud)
- intègre des problématiques d'usage collaboratif : dépôt, partage.

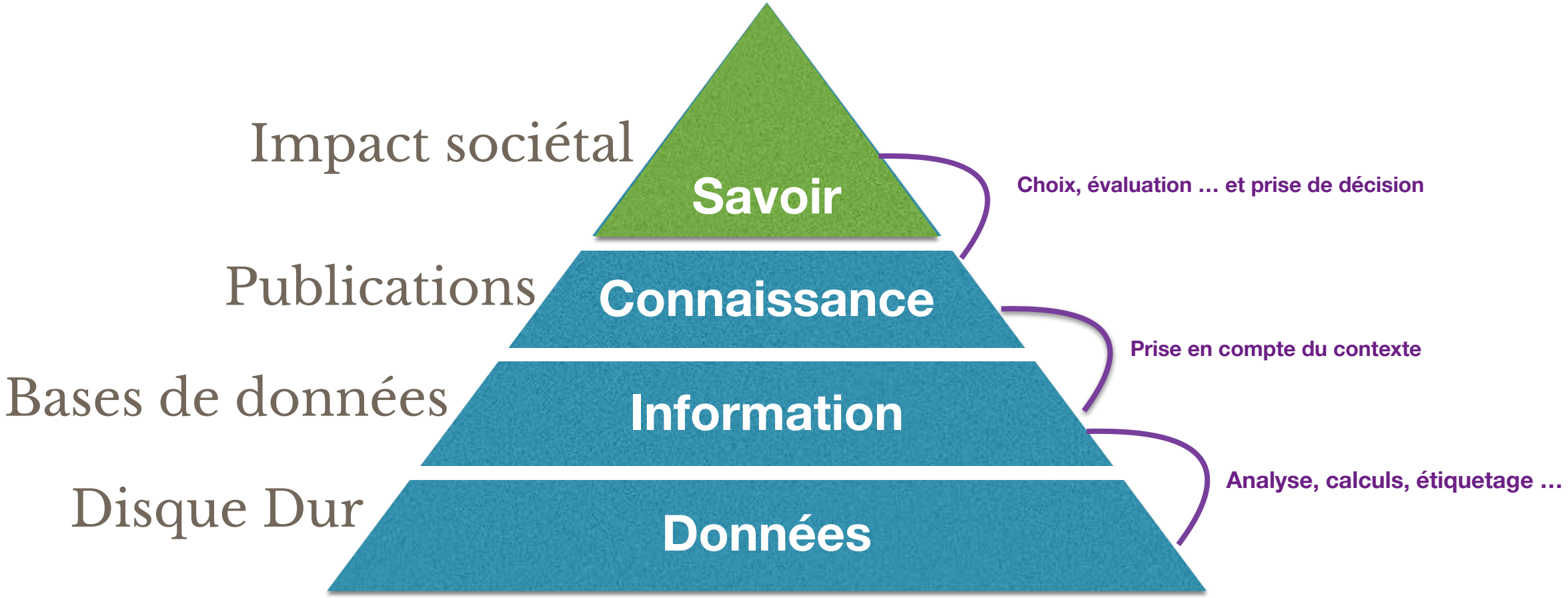
Critères de sélection pour choisir un support de stockage :

- la fréquence d'utilisation des données,
- les besoins en capacité de stockage (taille),
- la sécurité des données,
- la vitesse d'accès à la donnée
- la fiabilité et le coût du support

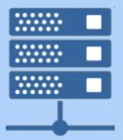
Les **besoins courants** pour la gestion de données lors d'un projet de recherche

- Des espaces de stockages adaptés à vos données (données scientifiques, documents bureautiques, bases de données, code source)
- Des outils adaptés à la gestion des droits des collaborateurs
- Des solutions de publication et d'archivage des données

Data Pyramid



Comparatif de systèmes de stockage des données

Support de stockage	Sécurité	Accès	Coût	Remarque d'utilisation
 Ordinateur professionnel	★★☆☆ Sujet au piratage informatique, aux détériorations et pannes	★☆☆☆ Pas adapté au partage, nécessite l'utilisation d'un support externe ou d'Internet (mail, cloud…)	★★★★ Pas de coût supplémentaire ou coût peu important	- Pour un stockage temporaire - Nécessité de crypter les données confidentielles et sensibles
 Support externe	★☆☆☆ - Sujet au vol, à la perte du support - Durée de vie limitée (dégradation du matériel)	★★★☆ Facilement transportable, il permet de transférer les données vers un autre ordinateur	★★★★ Pas de coût supplémentaire ou coût peu important	- Pour un stockage temporaire - Nécessité de crypter ou de sécuriser physiquement les données confidentielles et sensibles
 Serveur institutionnel	★★★★ Stockage fiable, durable et sécurisé (contre le vol, le piratage, les incendies…)	★★☆☆ La connexion au serveur institutionnel ne facilite pas le travail avec des personnes extérieures	★★☆☆ Coût assez important mais pas forcément répercuté sur l'usager	- Pour un stockage plus pérenne - Adapté pour le stockage de données sensibles et des versions « stables » de vos données - Toutes les institutions ne proposent pas ce service
 Serveur Cloud	★★☆☆ On ne sait pas vraiment où sont stockées les données, ni ce qu'elles deviennent	★★★★ Permet un travail synchronisé avec toutes les personnes ayant été autorisées au partage	★★☆☆ Payant à partir d'une certaine limite de stockage	- Pour un partage avec des personnes externes à l'institution - Ne pas y mettre de données sensibles ou confidentielles - Pas de contrôle sur la procédure de sauvegarde des données

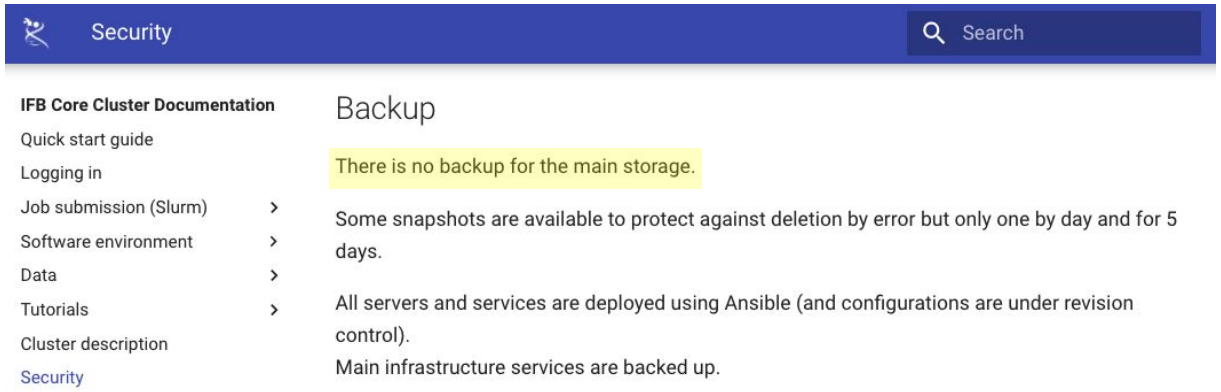
Source

Performance vs Sécurité :

- Une infrastructure de calcul nécessite une solution de stockage **performante** :
  - accès massivement parallèle aux données
  - disques rapides
- Pour gagner en performance, on désactive les mécanismes de **sécurité** :
  - Moins voire pas de snapshots
  - Pas de réplication
  - Pas de sauvegarde
- Pour gagner en sécurité, on réduit la performance

A capacité identique, le coût d'une infrastructure performante et d'une infrastructure sécurisé est le même

Infrastructure de calcul ne rime pas toujours avec infrastructure de stockage



Charte d'utilisation ROMEO

Conditions d'accès et règles de bon usage des ressources ROMEO

Version 2017/12

Créé en 2002, le Centre de Calcul Régional ROMEO accompagne les chercheurs de la région dans leurs activités numériques. La description complète des ressources et de leur utilisation est décrite sur http://romeo.univ-reims.fr

La présente demande, d'ouverture ou de maintien de compte sera étudiée et validée par le comité scientifique du centre de calcul et mis en œuvre par le personnel ROMEO.

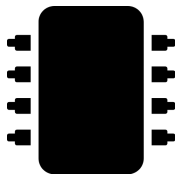
L'utilisateur s'engage, sous risque de fermeture de son compte sans préavis, à :

- consulter, corriger et améliorer les informations contenues sur le site pour toute question
- consulter les *notes de maintenance* sur le site web et sur les messages d'accueil des machines
- ne pas utiliser la machine comme espace de stockage ou de sauvegarde
- ne pas utiliser la machine comme passerelle depuis l'extérieur vers le réseau de l'URCA
- maintenir à jour ses coordonnées dans la rubrique *mon compte* du site web
- mettre à jour les projets dont il est responsable ou membre ainsi que la liste de ses publications dans la rubrique « mon compte » du site web
- mentionner l'utilisation de ROMEO sur vos communication :
  - *Ce travail a été réalisé avec le concours du Centre de Calcul Régional ROMEO*
  - *This work was partially supported by the French HPC Center ROMEO*
- prendre toute mesure afin d'empêcher l'utilisation de compte par des tiers (ne pas divulguer son mot de passe, choisir un mot de passe suffisamment complexe)
- participer aux événements organisés par le Centre de Calcul
- lire son mail régulièrement et répondre aux demandes venant du Centre de Calcul
- de manière générale, se conformer aux règles d'utilisations (batch, utilisation des scratchs, …) disponibles dans la rubrique *techno-centre* du site web
- libérer les espaces scratchs après leur utilisation
- communiquer avec l'équipe technique à l'adresse romeo@univ-reims.fr
- utiliser le site de support pour toute demande d'intervention https://romeo.univ-reims.fr/ticket
- participer à la diffusion des résultats scientifique (posters, vidéos, …)
- respecter les aspects légaux liés aux logiciels
- ne pas utiliser les ressources du centre a des fins criminelles, de violation ou tentative

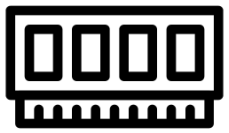
National Network of Computing Resources

Une offre de service
cloud et cluster couvrant
l'ensemble du territoire Français

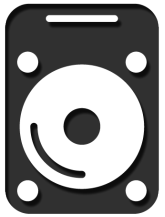
Cluster	Localisation du Data center	Coeurs	RAM (Go)	Stockage (To)
IFB Core	IDRIS - Orsay	5 042	26 542	2 000
Genotoul	Toulouse	6 128	34 304	3 000
ABiMS	Roscoff	2 608	10 600	2 500
GenOuest	Rennes	1 824	7 500	2 300
Migale	Jouy en Josas	2481	11 000	800
BiRD	Nantes	560	4 000	500



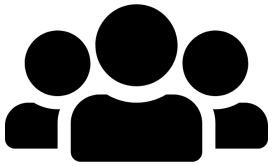
4300 coeurs



20 To RAM



2 Po



Une communauté d'entraide

Plus de 400 outils

SSH
Jupyter
RStudio
Galaxy

Allocation des espaces de stockage par projet :

- 250 Go par projet, extensible sur demande argumentée
- Un projet peut être accessible à plusieurs utilisateurs
- Un utilisateur peut demander plusieurs espaces projet
- Pas de sauvegarde
- Stockage pour le calcul, **pas archivage**

Bientôt disponible :

- Mise à disposition d'un espace scratch avec un quota plus important pour des besoins ponctuel (suppression automatique des fichiers les plus anciens)
- Sauvegarde des espaces projets

Comment transmettre vos données de recherche ?

Pas adapté → Adapté

Messagerie instantanée

Email

- Pas conçu pour le transfert de données
- Les communications peuvent être interceptées
- Localisation du stockage et durée de rétention inconnues

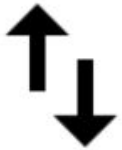
Envoi d'un disque

Dropbox, Drive, etc

- Risque de perte
- Risque d'accès non autorisés
- Acceptable si les données sont chiffrées

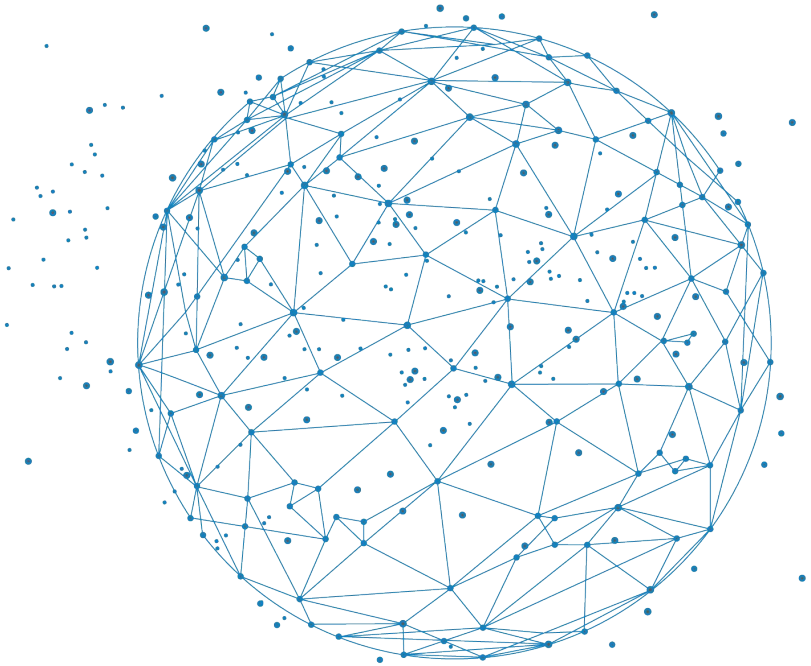
Cloud privé

Service d'un consortium



- Optimisé pour le transfert de données scientifiques
- Sécurisé
- Support gratuit

Un environnement de travail sûr



Un cas d'usage

Contexte :

Les chercheurs/ingénieurs effectuent leur travail sur leur PC, mais avec obligation de sauvegarder sur un serveur.

Situation :

Ce matin, je m'aperçois que mon PC est inaccessible : visiblement le disque dur est mort.

À part acquérir un nouveau poste de travail, comment vais-je récupérer mon environment logiciel & données ? Combien de temps est nécessaire pour retrouver un environnement de travail adapté ? Quelles pertes sont tolérables ?

Un environnement de travail sûr

Comprendre l'environnement de travail que vous utilisez avant de démarrer votre projet :

Votre poste de travail :

- Y'a-t-il des sauvegardes (stratégie 3-2-1) ?
  - **3** copies sur au moins **2** systèmes différents dont au moins **1** est distant
    = 0 inquiétude
  - *Par exemple : stockage en RAID (copie locale) + sauvegarde sur un disque externe qui reste au labo*
- Votre environnement est-il mis à jour régulièrement ?
- Disposez-vous d'un antivirus (à jour) ?
- Vos données sont-elles chiffrés (en cas de vol) ?

Vos solutions de stockage :

- Y'a-t-il des sauvegardes (stratégie 3-2-1) ?
- Est-ce que la pérennité est en phase avec vos besoins ?
- L'environnement est-il mis à jour régulièrement ?

source

Vos ~~mots~~ phrases de passes (au pluriel)

Password Length	Numerical 0-9	Upper & Lower case a-Z	Numerical Upper & Lower case 0-9 a-Z	Numerical Upper & Lower case Special characters 0-9 a-Z %$
1	instantly	instantly	instantly	instantly
2	instantly	instantly	instantly	instantly
3	instantly	instantly	instantly	instantly
4	instantly	instantly	instantly	instantly
5	instantly	instantly	instantly	instantly
6	instantly	instantly	instantly	20 sec
7	instantly	2 sec	6 sec	49 min
8	instantly	1 min	6 min	5 days
9	instantly	1 hr	6 hr	2 years
10	instantly	3 days	15 days	330 years
11	instantly	138 days	3 years	50k years
12	2 sec	20 years	162 years	8m years
13	16 sec	1k years	10k years	1bn years
14	3 min	53k years	622k years	176bn years
15	26 min	3m years	39m years	27tn years
16	4 hr	143m years	2bn years	4qdn years
17	2 days	7bn years	148bn years	619qdn years
18	18 days	388bn years	9tn years	94qtn years
19	183 days	20tn years	570tn years	14sxn years
20	5 years	1qdn years	35qdn years	2sptn years

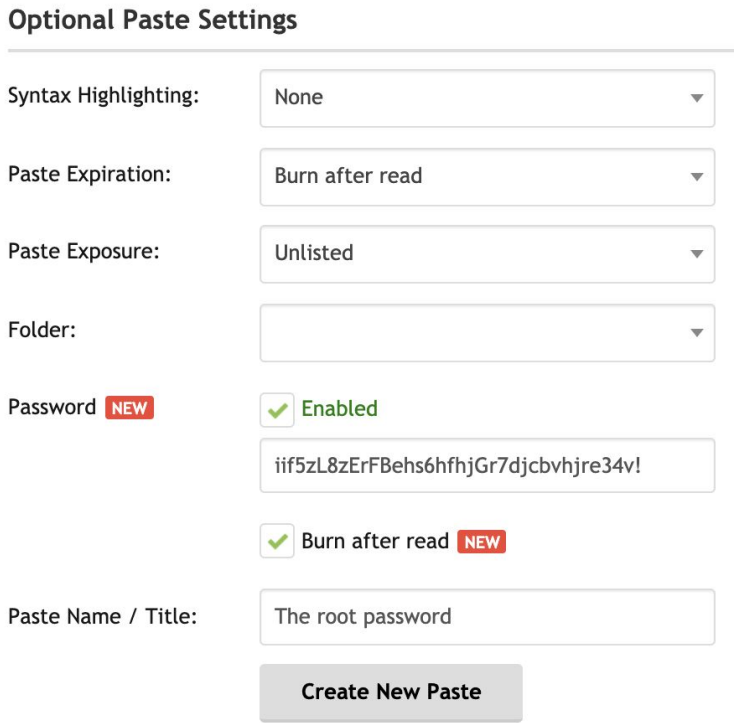
https://www.thesecurityfactory.be/password-cracking-speed/

At a current rate of 25$ per hour, an AWS p3.16xlarge [...] we're capable of trying a whopping 632.000.000.000 different password combinations per second!

Un environnement de travail sûr

- Utilisez-vous un mot de passe différent pour chaque fournisseur de service ?
- Utilisez-vous un gestionnaire de mot de passe ?
  - bitwarden (Open Source)
  - Dashlane, LastPass, 1password…
- Renouvelez-vous vos mots de passe régulièrement ?
- Utilisez-vous une procédure sécurisé pour communiquer un mot de passe à vos collègues ? (par exemple pastebin.com)

bitwarden



Un gestionnaire de mots de passe : Bitwarden

Bitwarden est un service en ligne qui vous permet de créer un coffre fort dans lequel vous allez pouvoir enregistrer tous vos mots de passe.

OpenSource	Gratuit	Accessible
et donc "pérenne" ***mots de passe exportables***	*mais n'hésitez pas à payer la souscription Premium pour soutenir le projet*	*Application Mac, Windows, Linux, Web, iPhone et Androïd*

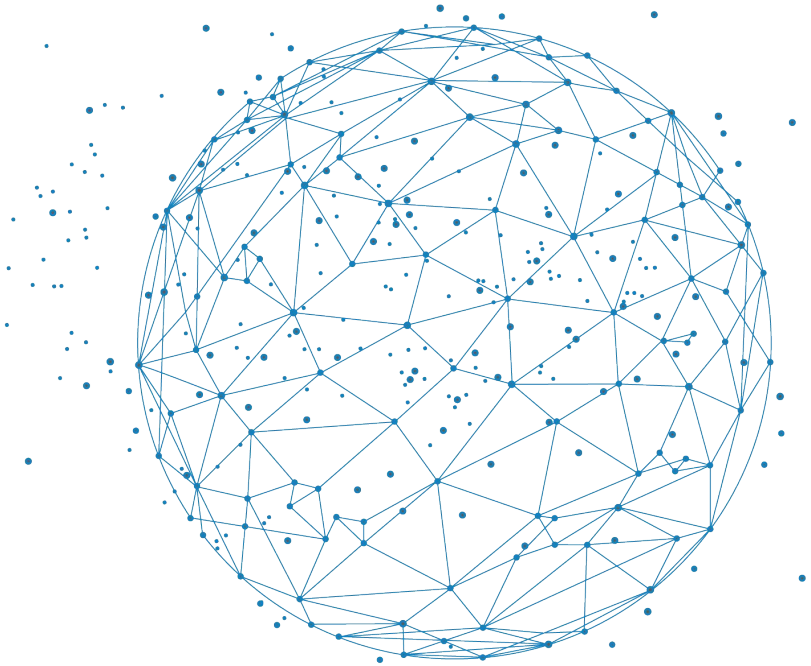
1. Créer votre compte sur https://bitwarden.com/
2. Choisissez votre mot de passe maître (size matters)
3. Installer les applications sur vos appareils et les extensions de vos navigateurs
4. Enregistrer vos mots de passes dans votre coffre fort Bitwarden
5. **Activez l'authentification à 2 facteurs**

En plus :

- Générateur de mot de passe robuste intégré
- Analyse de vos mots de passes et reporting
- Partage de mots de passe entre collègue

Quel gestionnaire de mots de passe choisir ? (Les numériques)

Protéger ses données



Un cas d’usage

Situation :

Une partie de vos données est considérée comme données sensibles.

Que mettez vous en place pour protéger l’accès à ces données ?

Identifier et contrôler la corruption des données

- **Corruption** : introduction de modifications non intentionnelles des données

Les données peuvent être corrompues par :

- des modifications non souhaités
  - (ransomware, collègue ~~un peu c** ~~peu soigneux, une erreur dans un script, …)
  - un **transfert de données défectueux**
  - un plantage d'un disque dur
  - …

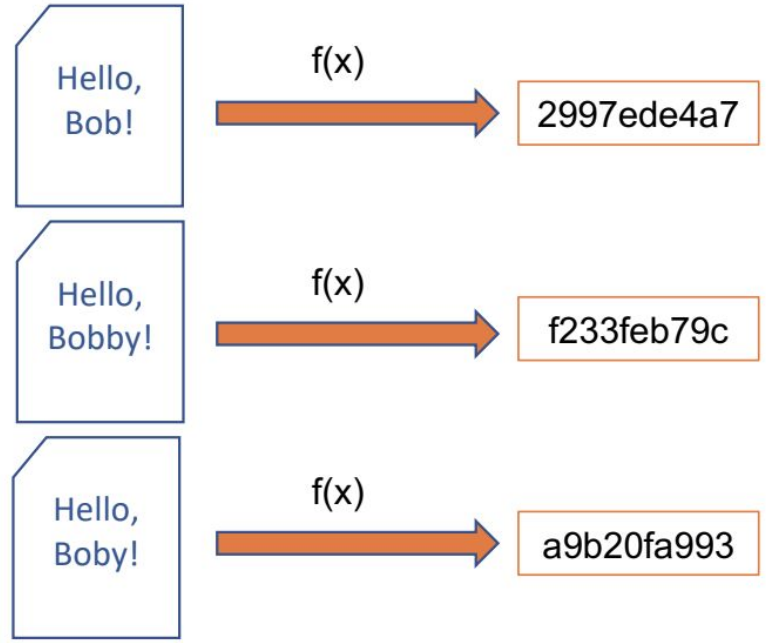
Recommandation 1 : générer des sommes de contrôles

Comment ?

- Linux / macOS : md5sum, sha256sum
- Windows : certutil

Quand ?

- **Avant un transfert de données**
  - Lorsqu'on réceptionne un nouveau jeu de données d'un collaborateur, d'une plateforme
  - Lorsqu'on transfert des données sur un stockage distant (exemple : EMBL ENA)
- **Stockage à long terme**
  - La version principale de chaque dataset
  - Les extraits de données utilisés dans les publications



Recommandation 2 : utilisez le contrôle d'accès

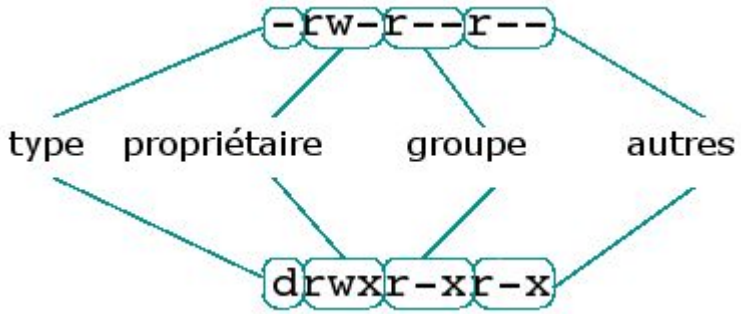
N'accordez que les permissions d'accès nécessaire :

- Limitez le nombre d'utilisateurs ayant accès à vos données
- Limitez la visibilité des données (réseau interne vs internet)
- N'utilisez jamais de partage public sans chiffrement des données !

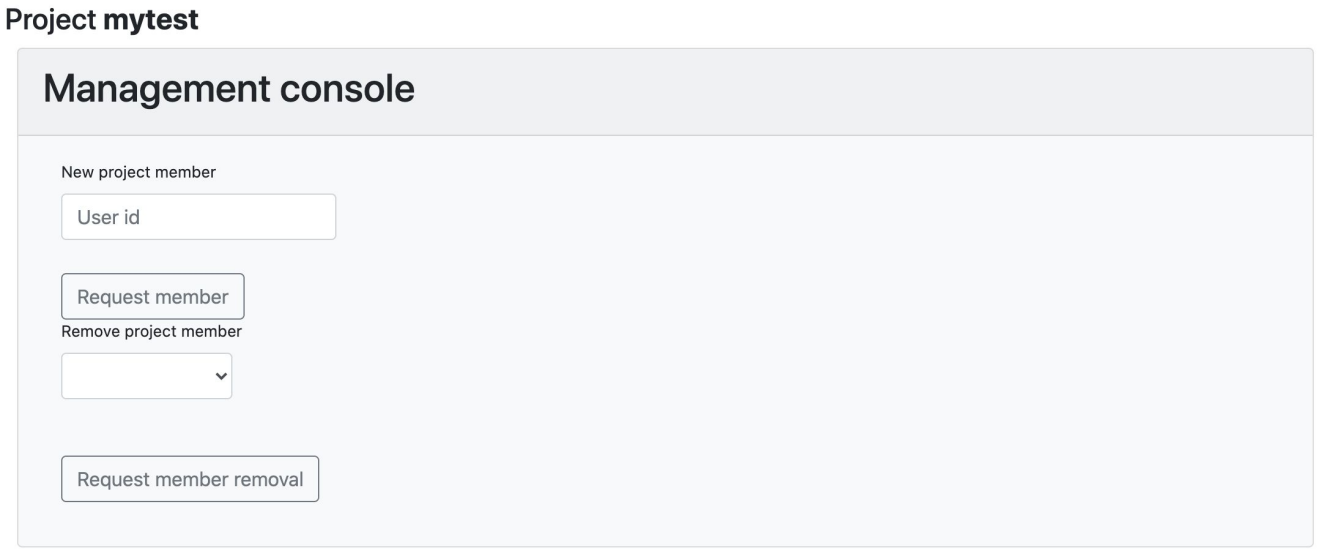
Mettez les données brutes en lecture seule !!

L'accès aux données sensibles doit être **documenté**

Gestion fine des droits au niveau UNIX



- DroitS POSIX
- ACL



exemple sur IFB

Limitez les copies au maximum !

- **Copie principale** (master)
  - Egalement appelé donnée “source” ou “brute”
  - En ***read-only***
  - Stratégie 3-2-1
- **Copie de travail**
  - A éviter au maximum
  - Utilisez des liens symboliques (`ln -s`) ou des alias vers la copie principale
- **Copie de sauvegarde**
  - Ne travaillez jamais sur votre copie de sauvegarde

IFB
INSTITUT FRANÇAIS DE BIOINFORMATIQUE

elixir FRANCE

Module 2

Pratiques d'hygiène numérique pour la gestion des données

Traçabilité et réutilisation des données

ARTbio IBPS

migale

INSTITUT PASTEUR

PLANT BIOINFORMATICS FACILITY

cnrs

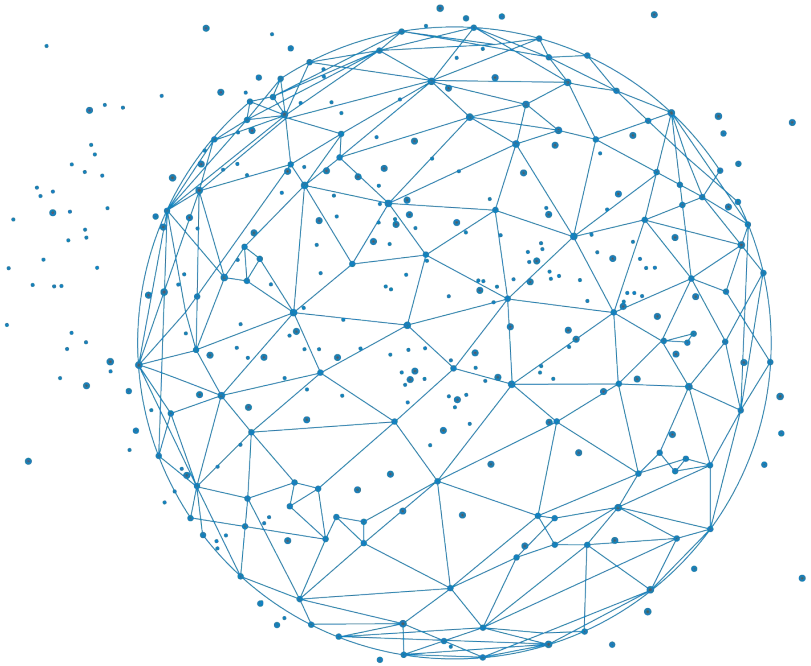
INRAE

Inserm

cea

INVESTIR L'AVENIR

Ranger ses données



Un cas d'usage

Un collègue vous envoi cette table de mesure.
Que lui proposeriez-vous pour améliorer cette table de données ?

species	habitat	weight	length	latitude/longitude	date
Alligator mississippiensis	swamp	431 lb	4 ft 2	29.531,-82.184	Sept 15, 2015
Puma concolor	forest	125 lb	2.2m	29.125,-81.682	08/10/2015
Ursus americanus	forest	88 kg	133 cm	N29°7'30"/W81°40'55.2"	07-13-2015

Bonus : Qu'est ce qui relève de la donnée et de la métadonnée ?

Un cas d'usage

A

Untidy Data

species	habitat	weight	length	latitude/longitude	date
Alligator mississippiensis	swamp	431 lb	4 ft 2	29.531,-82.184	Sept 15, 2015
Puma concolor	forest	125 lb	2.2m	29.125,-81.682	08/10/2015
Ursus americanus	forest	88 kg	133 cm	N29°7'30"/W81°40'55.2"	07-13-2015

B

Tidy Data

meta-data

data

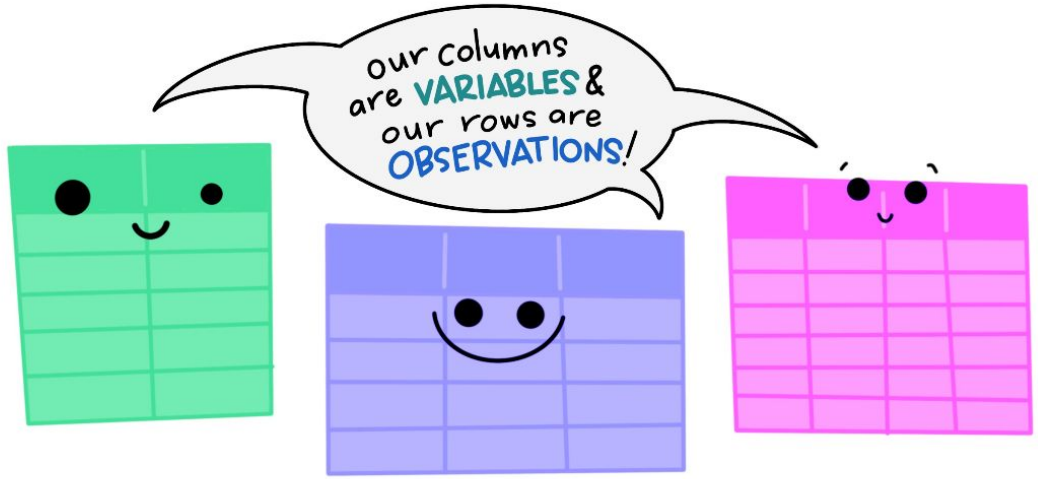
species_code	date	station_code	weight_kg	length_cm
TSN 551771	2015-09-15	1	196	127
TSN 55247	2015-08-10	2	57	220
TSN 180544	2015-07-13	2	88	133

station_code	habitat	latitude	longitude
1	swamp	29.531	-82.184
2	forest	29.125	-81.682

species_code	class	genus	species
TSN 551771	Reptilia	Alligator	mississippiensis
TSN 55247	Mammalia	Puma	concolor
TSN 180544	Mammalia	Ursus	americanus

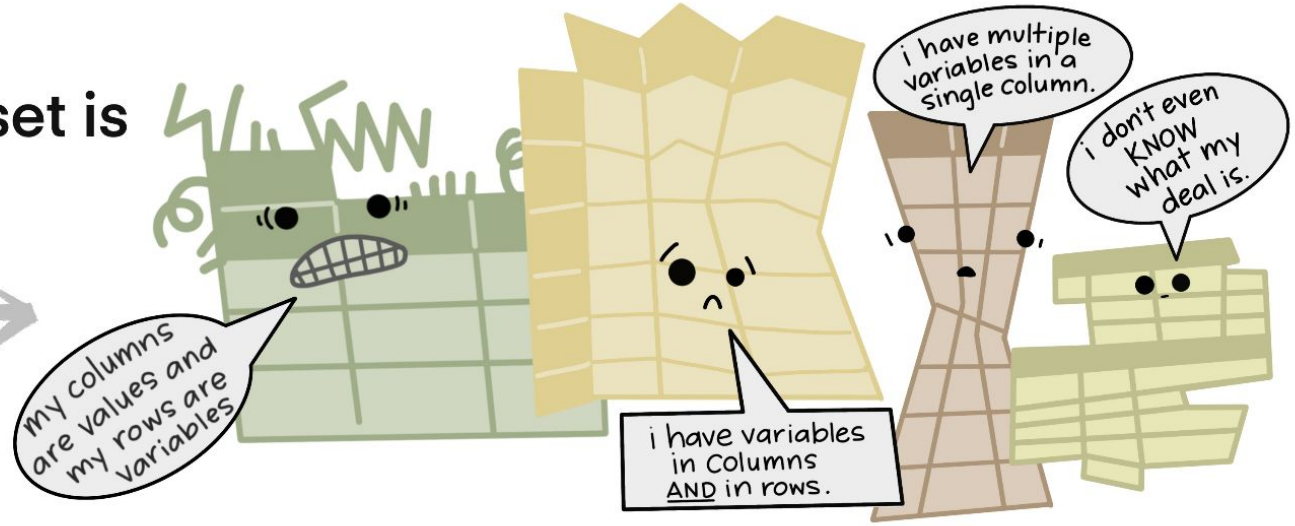
https://doi.org/10.1371/journal.pcbi.1005097.g001

The standard structure of tidy data means that "tidy datasets are all alike..."



"...but every messy dataset is messy in its own way."

–HADLEY WICKHAM



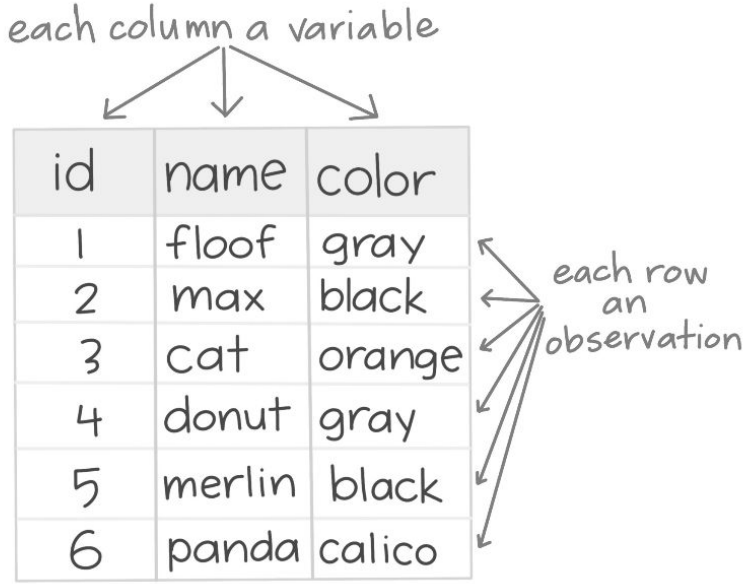
https://github.com/allisonhorst/stats-illustrations

"TIDY DATA is a standard way of mapping the meaning of a dataset to its structure."

–HADLEY WICKHAM

In tidy data:

- each variable forms a column
- each observation forms a row
- each cell is a single measurement

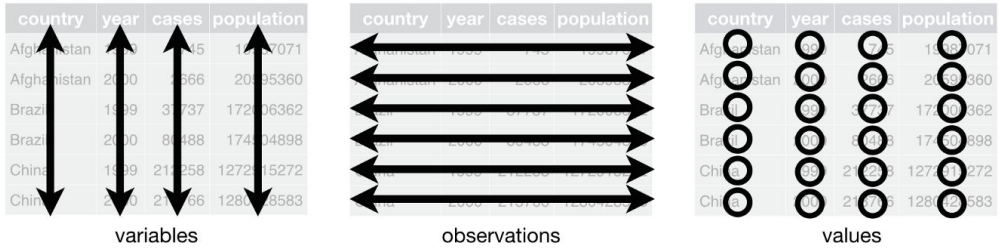


Wickham, H. (2014). Tidy Data. Journal of Statistical Software 59 (10). DOI: 10.18637/jss.v059.i10

https://github.com/allisonhorst/stats-illustrations

Utiliser une structure de données uniformes

- Chaque colonne est une variable.
- Chaque ligne est une observation.
- Chaque cellule est une valeur unique.



Exemples de chose a éviter

- Les en-têtes de colonne sont des valeurs, et non des noms de variables.
- Plusieurs variables sont stockées dans une seule colonne.
- Les variables sont stockées à la fois dans les lignes et les colonnes.
- Plusieurs types d'observation sont stockés dans la même table.
- Une unique d'observation est stockée dans plusieurs tables.

Permet d’uniformiser les outils d’analyses et de visualisation pour faciliter la réutilisation et l’interopérabilité des données

Gene name error

Ziemann *et al. Genome Biology* (2016) 17:177
DOI 10.1186/s13059-016-1044-7

Genome Biology

COMMENT Open Access

Gene name errors are widespread in the scientific literature

Mark Ziemann[1], Yotam Eren[1,2] and Assam El-Osta[1,3*]

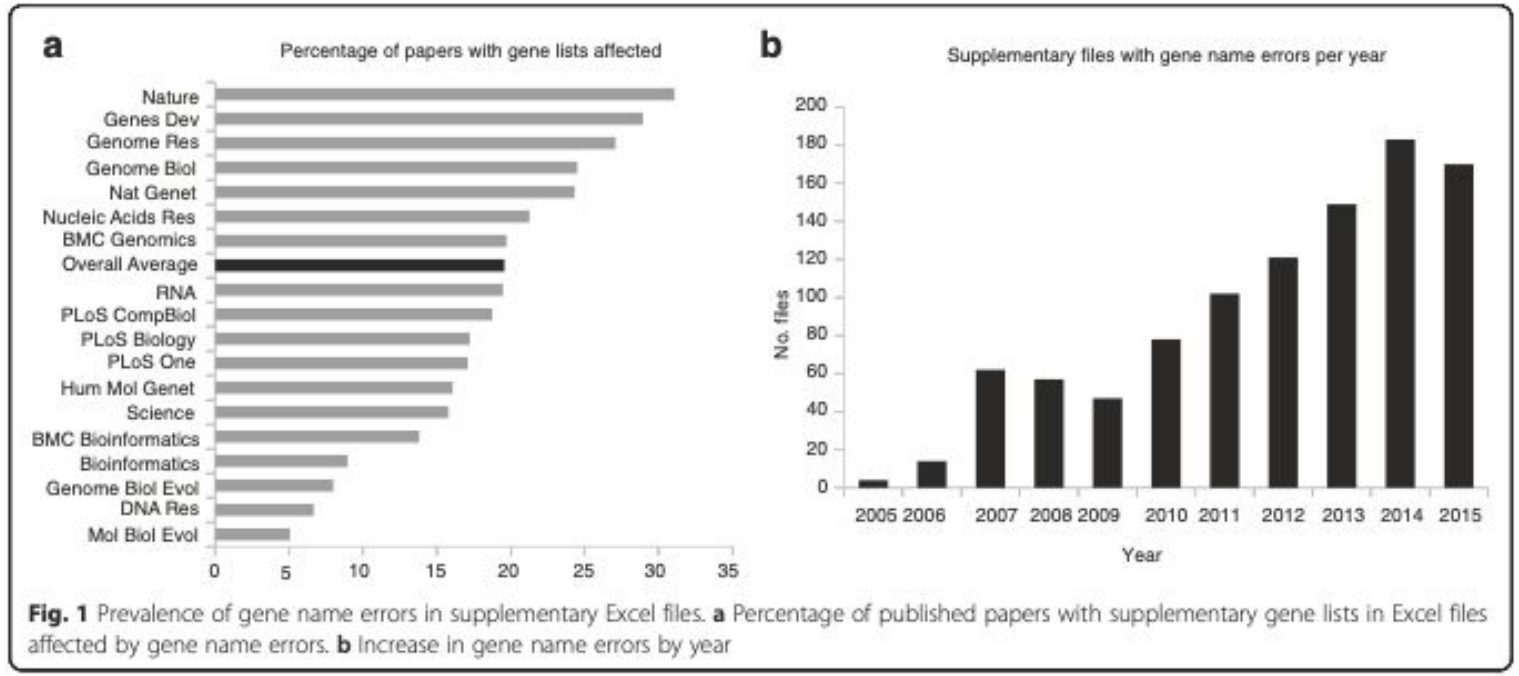
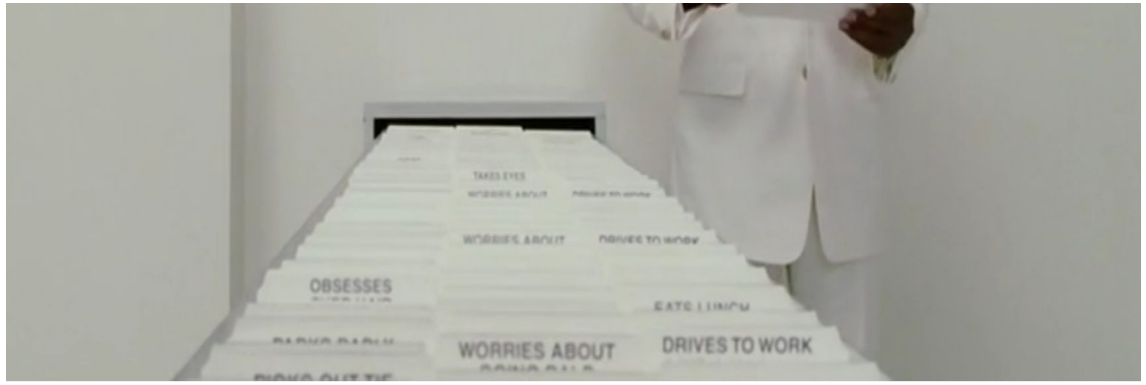
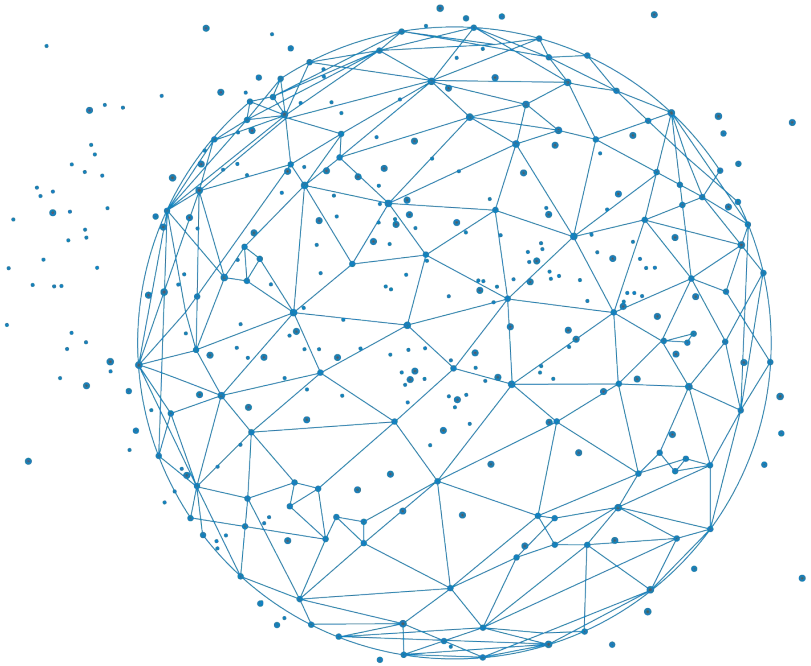


Fig. 1 Prevalence of gene name errors in supplementary Excel files. **a** Percentage of published papers with supplementary gene lists in Excel files affected by gene name errors. **b** Increase in gene name errors by year

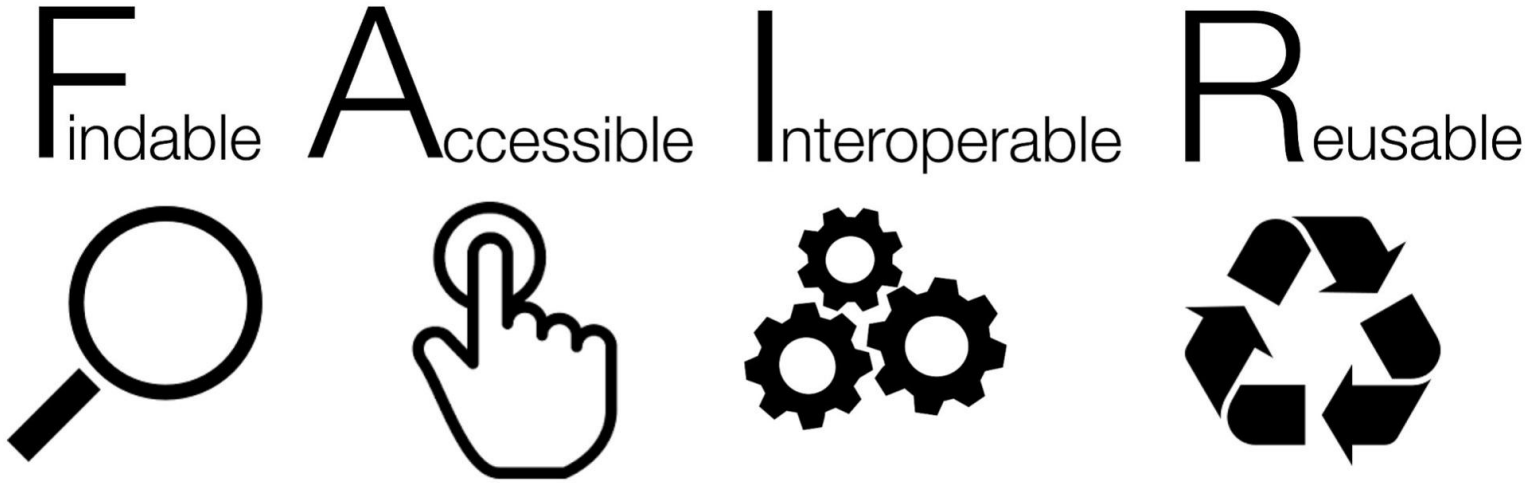
https://doi.org/10.1186/s13059-016-1044-7

Le nommage des fichiers





Nommer ses fichiers



Findable

It should be possible for others to discover your data. Rich metadata should be available online in a searchable resource, and the data should be assigned a persistent identifier.

- ☐ A persistent identifier is assigned to your data
- ☐ There are rich metadata, describing your data
- ☐ The metadata are online in a searchable resource e.g. a catalogue or data repository
- ☐ The metadata record specifies the persistent identifier

Situation :

C’est parti pour un projet de biologie intégrative sur 3 ans, au programme acquisitions de nombreux types de données (imagerie, séquençage, phénotypage) et analyses intensives.

Expliquez votre approche de nommage et d’organisation des fichiers
(le nom des fichiers doit obligatoirement comprendre au moins la date)

DONNER UN NOM BREF, EXPLICITE, UNIQUE et …

- Pas d'espace, pas de point, ni de caractères spéciaux (`&/+>:?%*` …)
  Utiliser des tirets (-) ou underscores (_) pour séparer les éléments

Règles dénomination fichiers ✗

ReglesDenominationFichiers ✓

- Dates au format `AAAAMMJJ`
  ou `AAAA-MM-JJ` (année, mois, jour)

20150405_CR 20160310_CR 20160515_CR

- Versionnez (cf GitLab)

Convention_V01 Convention_V02 Convention_VF

ISO 8601 : https://xkcd.com/1179/

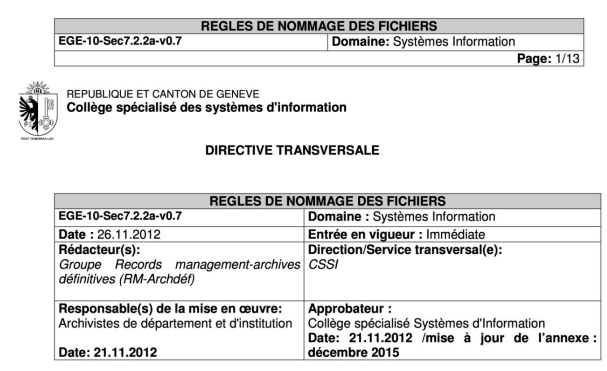
- Pas trop long

cenomestpeutetreexplicitemaisprobablementbeaucouptroplongetencorecapourraitetrepirecarilmanqueladateetlalistedetouslesauteurs_vf_2.txt

- Rangez

Reunion
20150407_CR
20150407_Minutes
20150407_OJ

- Et documentez vos règles



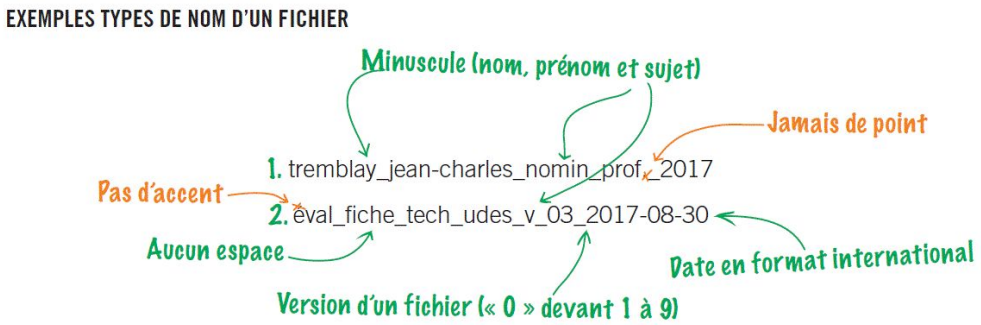
REGLES DE NOMMAGE DES FICHIERS
EGE-10-Sec7.2.2a-v0.7 | Domaine: Systèmes Information
Page: 1/13

REPUBLIQUE ET CANTON DE GENEVE
Collège spécialisé des systèmes d'information

DIRECTIVE TRANSVERSALE

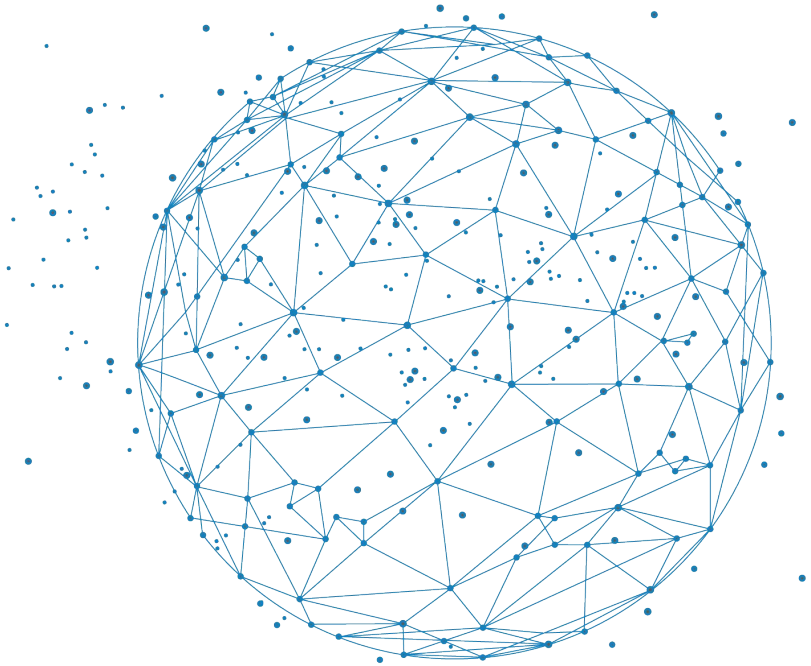
REGLES DE NOMMAGE DES FICHIERS	
EGE-10-Sec7.2.2a-v0.7	Domaine : Systèmes Information
Date : 26.11.2012	Entrée en vigueur : Immédiate
Rédacteur(s): *Groupe Records management-archives définitives (RM-Archdél)*	Direction/Service transversal(e): *CSSI*
Responsable(s) de la mise en œuvre: Archivistes de département et d'institution Date: 21.11.2012	Approbateur : Collège spécialisé Systèmes d'Information Date: 21.11.2012 /mise à jour de l'annexe : décembre 2015

Exemple de règles

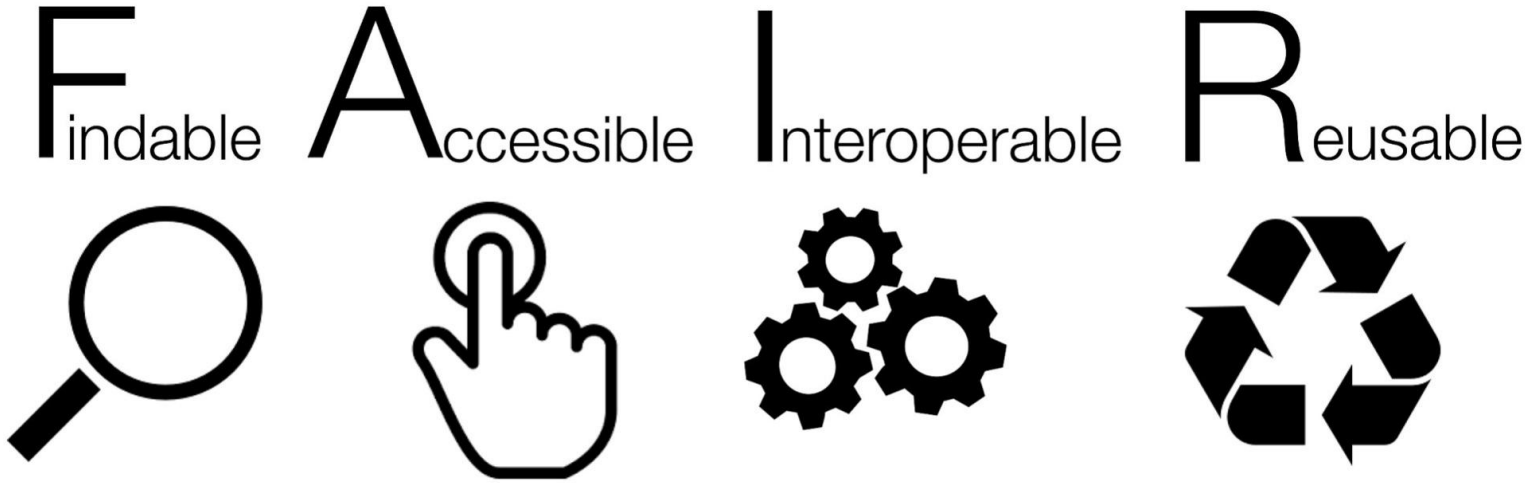


Eléments	Règle	Exemple
Sujet	**Obligatoire** Il s'agit du sujet principal traité au sein du document. Utiliser des noms communs, écrits en lettres minuscules non accentuées.	 projet formation évaluation
Séparateur	Les espaces sont interdits. Utiliser l'underscore (touche 8 du clavier) pour remplacer les espaces	 « _ »
Type de document	**Facultatif** Qualifie la nature du document. Toute abréviation sera en lettres majuscules.	 (CR) compte rendu (OJ) ordre du jour
Date	**Obligatoire** Date de création du document, date de l'événement. Format à l'américaine : AAAAMMJJ. Nommage d'une période : utilisation d'un séparateur «_» ou « -».	 20180122 201608 2010 201501_07 ou 201501-07
Version du document	**Obligatoire** Distingue les différentes versions d'un document, signalées par un « V » majuscule suivi de deux chiffres ; version provisoire (VP) et la version finale (VF), version validée (VV). Un nouveau document créé à partir d'une version finale doit être sauvegardé sous un nouveau nom de manière à ne pas écraser la version précédente.	 CR_CFVU_V0.0 CR_CFVU_V0.1 CR_CFVU_VP, VF ou VV
Extension	**Obligatoire** L'extension est ajoutée automatiquement par le système et n'apparaît peut-être pas sur vos écrans.	 .txt (fichier texte) .doc (fichier Word) .xls (fichier Excel)

Formats des fichiers



Formats des fichiers



Interoperable

Data and metadata should conform to recognised formats and standards to allow them to be combined and exchanged.

- ☐ Data is provided in commonly understood and preferably open formats
- ☐ The metadata provided follows relevant standards
- ☐ Controlled vocabularies, keywords, thesauri or ontologies are used where possible
- ☐ Qualified references and links are provided to other related data

Situation :

Vous devez traiter un fichier avec un format ‘propriétaire’, c’est à dire qui nécessite un logiciel dont le code source n’est pas disponible (et dans ce cas, non gratuit) pour lire le fichier. Votre institution n’a aucune licence pour ce logiciel, et ne projette pas d’en acquérir.

Quelles sont les solutions possibles ?

Formats et logiciel ?

Allez à : http://scrumblr.ca/fair_data_format

et listez les formats que vous connaissez dans les colonnes appropriées

Texte brute
Texte formaté
Données numériques
Tableurs & Stats
Image (matriciel, vectoriel, 3D)
Son & Vidéo
Cartographie
Bioinfo
Code & Script
Archive & compression

Un cas d'usage

Un cas d'usage

Des logiciels spécifiques sont ils nécessaire pour traiter les formats cités ?

Fonctionnent-ils en ligne ou après installation sur un ordinateur ?

Fonctionnent-ils avec un système d'exploitation particulier (Windows, Mac, Linux) ?

Sont-ils lies à un type d'ordinateur ou à un instrument particulier (ex : microscope) ?

Sont-ils gratuits ou payants ? Qui paye ?

S'ils n'existaient plus ou si vous n'y avez plus accès, pourriez-vous continuer à travailler ?

L'éditeur du logiciel (ou la communauté) est il en bonne santé ?

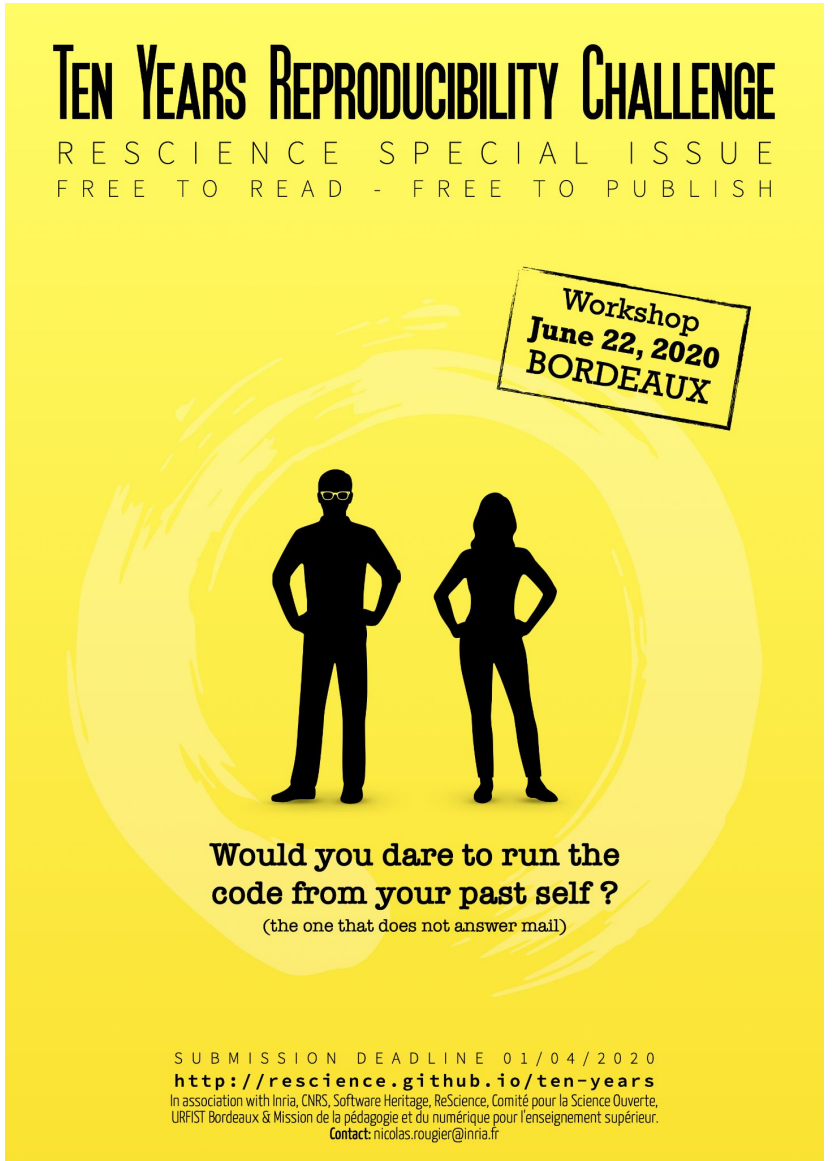
Le logiciel sera encore disponible dans 20 ans et sera-t-il encore capable de d'interpréter correctement le fichier ?

⇒ Que proposez vous pour garantir la pérennité de l'accès à vos données ?

Un cas d'usage

Quelles conséquences pour vous ?

Et pour ceux qui arriveront plus tard ?



https://github.com/ReScience/ten-years

Enjeu pour la préservation et l'exploitation des données

Formats « textuels »

- Suite d'octets représentant des caractères imprimables et affichables à l'écran
- Peuvent être lus dans un éditeur de texte
- Mais souvent besoin d'un logiciel spécifique pour interpréter la structure interne, matérialisée par certains caractères, et en donner une représentation informatique exploitable
- Caractères ordinaires + caractères ayant une valeur spéciales : < > / * etc.

HTML (avec des “balises”) conçu principalement pour représenter les pages web.

Contenu lisible dans un éditeur texte :

```
<html>
<head><head>
<body>
<p>Bonjour <span style='color:red'>tout le monde</span></p>
</body>
</html>
```

Mais « interprétable » par un logiciel dédié (navigateur web) :

Bonjour tout le monde

RTF : Rich Text Format

Contenu lisible dans un éditeur texte

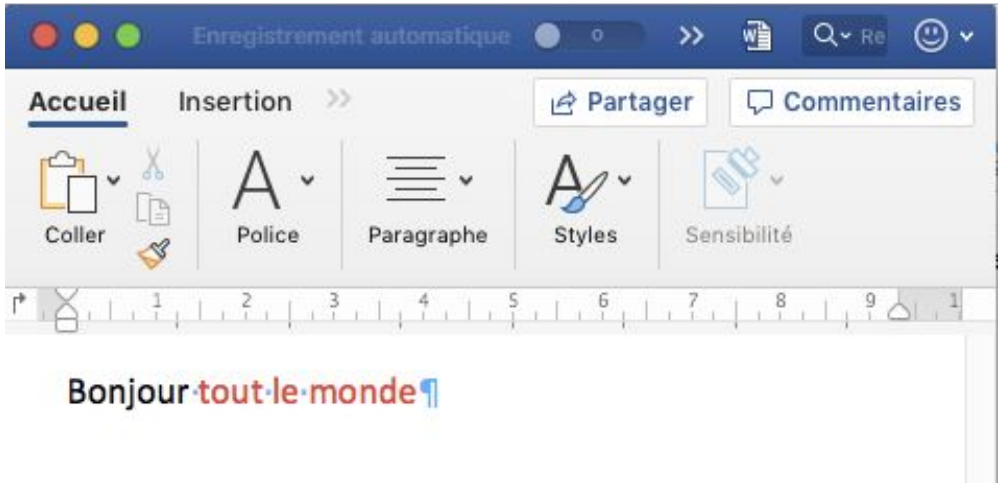
```
{\rtf1\adeflang1025\ansi\ansicpg1252\uc1\adeff0\deff0\stshfdbch37\stshfl
och37\stshfhich37\stshfbi0\deflang1036\deflangfe1036\themelang1036\theme
langfe0\themelangcs0{\fonttbl{\f0\fbidi \froman\fcharset0\fprq2{\*\panos
e 02020603050405020304}Times New Roman;}{\f34\fbidi \froman\fcharset0\fp
rq2{\*\panose 02040503050406030204}Cambria Math;}
```

```
\mlMargin0\mrMargin0\mdefJc1\mwrapIndent1440\mintLim0\mnaryLim1}{\info{\
author Mathieu Saby}{\operator Mathieu Saby}{\creatim\yr2018\mo6\dy10\hr
13\min44}{\revtim\yr2018\mo6\dy10\hr13\min44}{\version2}{\edmins1}{\nofp
ages1}{\nofwords3}{\nofchars19}
```

```
\fs24\lang1036\langfe1033\loch\af37\hich\af37\dbch\af37\cgrid\langnp1036
\langfenp1033 {\rtlch\fcs1 \af0 \ltrch\fcs0 \insrsid16651434 \hich\af37\
dbch\af37\loch\f37 Bonjour }{\rtlch\fcs1 \af0 \ltrch\fcs0 \cf6\insrsid16
651434\charrsid16651434
```

```
\hich\af37\dbch\af37\loch\f37 tout le monde}{\rtlch\fcs1 \af0 \ltrch\fcs
0 \insrsid16651434
```

Mais uniquement interprétable avec Word, Libre office ou autre traitement de texte

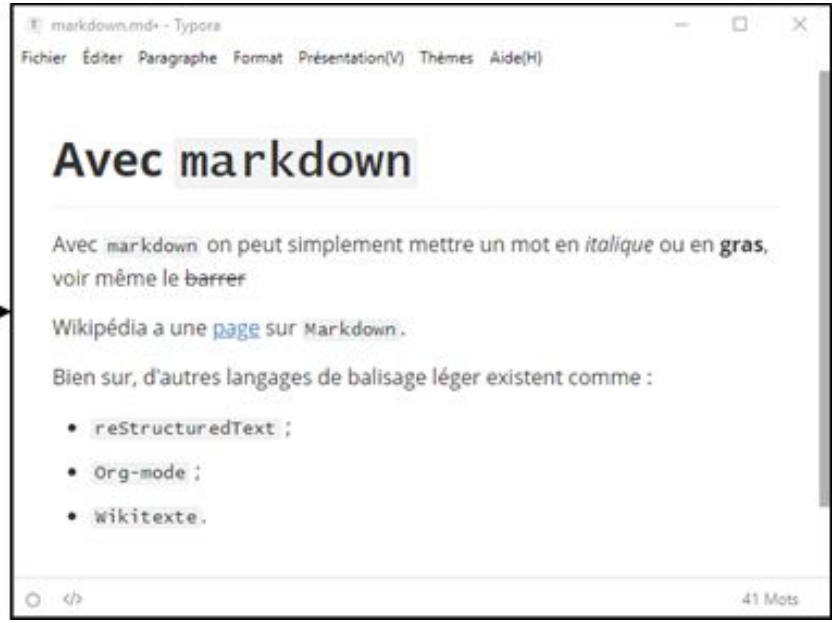


Markdown

- Langage de balisage léger
- Facile à lire et à écrire
- Facilement interprétable avec de nombreux éditeurs
- Peut être lu en l'état sans donner l'impression d'avoir été balisé ou formaté par des instructions particulières

```
# Avec `markdown`

Avec `markdown` on peut simplement mettre un mot
en _italique_ ou en __gras__,
voir même le ~~barrer~~

Wikipédia a une
[page](https://fr.wikipedia.org/wiki/Markdown) sur
`Markdown`.

Bien sur, d'autres langages de balisage léger
existent comme :

- `reStructuredText` ;
- `Org-mode` ;
- `Wikitexte`.
```

export



Formats binaires

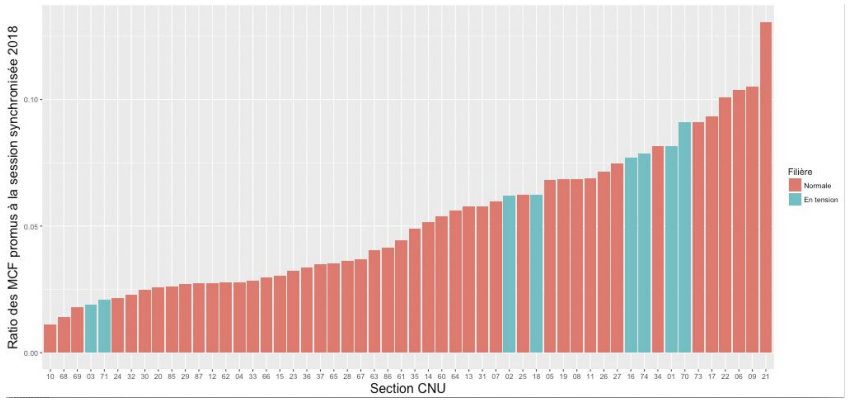
- Suite d’octets non interprétables comme des caractères imprimables ou affichables
- Structure interne opaque
- Besoin de logiciel spécifique pour les lire et les interpréter

Exemple de format binaire : **PNG** (image)

- Contenu illisible dans un éditeur texte
  (à part «?PNG » au début)

```
?PNG

?V4??????6n?I6?"?d??θ??83???OEP|1?L?? (??>?/?
%?? (>???P苦?;3?i???e?|??{?g?賺X????-2?s???=+?????WQ+]?L6O

w?[?C?{_???????F qb??

                              ???U?vz????Z?b?1@?/z??c??s>~?if?,?HUS

j??????F
```

- Uniquement lisible et interprétable avec un visionneur d’images :



Format ouvert ou fermé

- Privilégiez les **formats ouverts** afin de faciliter le partage des données
  - Définition légale du format ouvert en France (loi no 2004-575 du 21 juin 2004) :
  - On entend par standard ouvert tout protocole de communication, d'interconnexion ou d'échange et tout format de données **interopérable** et dont les spécifications techniques sont **publiques** et **sans restriction d'accès ni de mise en œuvre**.
  - ⇒ Format bien documenté et utilisable sans demander d'autorisation

Format ouvert	Format fermé
Spécifications publiques et gratuites	Spécifications non publiques
Aucune restriction légale pour l'utiliser	Des restrictions légales s'opposent à son utilisation (droit d'auteur, copyright, brevet)
Format indépendant du logiciel utilisé qui assure l'interopérabilité des données	Format lisible qu'avec un logiciel particulier
Maintenu par une organisation à but non lucratif	Format propriétaire

Note : pour que ces organisations non-lucratives puissent perdurer, il est nécessaire et important de contribuer régulièrement

Formats conseillés

Type	Formats conseillés	Formats non conseillés
Document texte	PDF, TXT, ODT	MS Word, RTF
Feuille de calcul	CSV, ODS	MS Excel, PDF, OOXML
Base de données	SQL, SIARD, DB tables (.CSV)	MS Access, dBase (.dbf), HDF5
Données statistiques	CSV, XML, TXT, RData (suivant les versions)	SAS, RData (suivant les versions)
Images	JPEG, TIFF, PNG	DICOM
Audio	BWF, MXF, Matroska (.mka), FLAC, OPUS	WAVE, MP3, AAC, AIFF, OGG
Video	MXF, MKV	MPEG-4, MPEG-2, AVI, QuickTime (.mov, .qt)
Information géographique	GML, MIF/MID	ESRI Shapefiles, MapInfo, KML
Images géoréférencées	GeoTIFF (.tif, .tiff)	TIFF World File
Raster	ASCII GRID (.asc, .txt)	ESRI GRID

Formats conseillés - Compression

CORNELL UNIVERSITY LIBRARY

File formats for digital content: Probability for full long-term preservation

Content type	High	Medium	Low
Text	• Plain text (encoding: USASCII, UTF-8, UTF-16 with BOM) • XML (includes XSD/XSL/XHTML, etc.; with included or accessible schema) • PDF/A-1 (ISO 19005-1) (*.pdf)	• Cascading Style Sheets (*.css) • DTD (*.dtd) • Plain text (ISO 8859-1 encoding • PDF (*.pdf) (embedded fonts) • Rich Text Format 1.x (*.rtf) • HTML (include a DOCTYPE declaration) • SGML (*.sgml) • Open Office (*.sxw/*.odt) • OOXML (ISO/IEC DIS 29500) (*.docx) • Microsoft Word 2007 or newer (*.docx)	• PDF (*.pdf) (encrypted) • Microsoft Word 2003 or older (*.doc) • WordPerfect (*.wpd) • DVI (*.dvi) • All other text formats not listed
Raster image	• TIFF (uncompressed) • JPEG2000 (lossless) (*.jp2) • PNG (*.png)	• BMP (*.bmp) • JPEG/JFIF (*.jpg) • JPEG2000 (lossy) (*.jp2) • TIFF (compressed) • GIF (*.gif) • Digital Negative DNG (*.dng)	• MrSID (*.sid) • TIFF (in Planar format) • FlashPix (*.fpx) • PhotoShop (*.psd) • RAW • JPEG 2000 Part 2 (*.jpf, *.jpx) • All other raster image formats not listed
Vector graphics	• SVG (no Java script binding) (*.svg)	• Computer Graphic Metafile (CGM, WebCGM) (*.cgm)	• Encapsulated Postscript (EPS) • Macromedia Flash (*.swf) • All other vector image formats not listed
Audio	• AIFF (96kHz 16bit PCM) (*.aif, *.aiff) • WAV (96kHz 24bit PCM) (*.wav)	• SUN Audio (uncompressed) (*.au) • Standard MIDI (*.mid, *.midi) • Ogg Vorbis (*.ogg) • Free Lossless Audio Codec (*.flac) • Advance Audio Coding (*.mp4, *.m4a, *.aac) • MP3 (MPEG-1/2, Layer 3) (*.mp3)	• AIFC (compressed) (*.aifc) • NeXT SND (*.snd) • RealNetworks 'Real Audio' (*.ra, *.rm, *.ram) • Windows Media Audio (*.wma) • Protected AAC (*.m4p) • WAV (compressed) (*.wav) • All other audio formats not listed
Video	• Motion JPEG 2000 (ISO/IEC 15444-4)??*.mj2) • AVI (uncompressed/native, motion JPEG) (*.avi) • QuickTime Movie (uncompressed/native, motion JPEG) (*.mov)	• Ogg Theora (*.ogg) • MPEG-1, MPEG-2 (*.mpg, *.mpeg, wrapped in AVI, MOV) • MPEG-4 (H.263, H.264) (*.mp4, wrapped in AVI, MOV)	• AVI (others) (*.avi) • QuickTime Movie (others) (*.mov) • RealNetworks 'Real Video' (*.rv) • Windows Media Video (*.wmv) • All other video formats not listed

IFB INSTITUT FRANÇAIS DE BIOINFORMATIQUE

elixir FRANCE

IBI

Format standardisé

La documentation d'un format peut devenir une norme officielle nationale ou internationale ou un standard de facto

- une norme est établie par un organisme reconnu, comme l'ISO ou l'AFNOR), qui fournit des règles. On a donc une garantie de stabilité et de pérennité
- un standard est établi par un groupe privé, pour assurer une cohérence des échanges a un moment donné.

Il existe des cas de standards qui sont devenus des normes, par exemple :

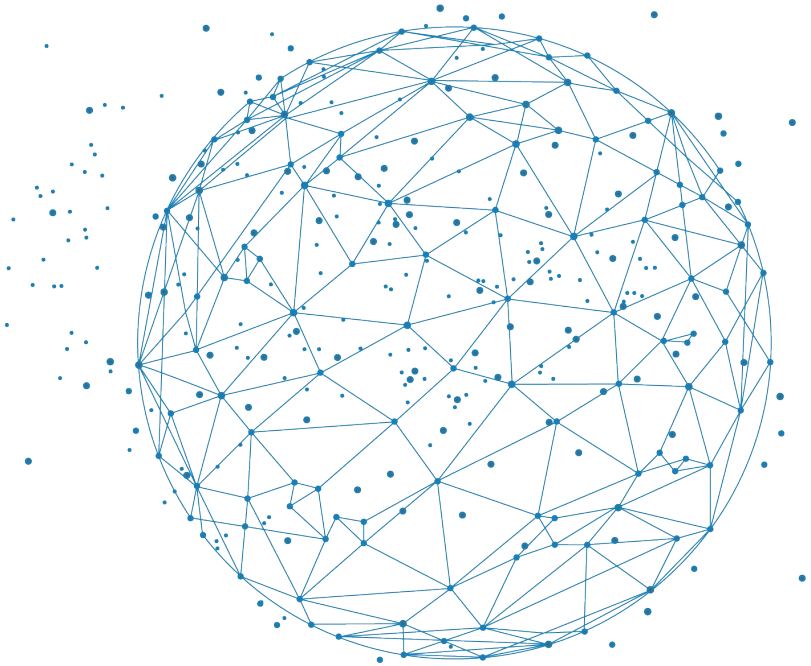
- PDF, standard Adobe devenu PDF/A1 norme ISO (ISO 19005)
- Les formats LibreOffice (ODS, ODT...) sont standardisés (ISO/IEC 26300)
- PNG, standard W3C devenu norme ISO (ISO 15948)
- Le format CSV est décrit dans la RFC 4180 (mais il s'agit plus d'un document indicatif que d'une norme, plusieurs versions existent)
- Les formats bureautique Microsoft Office (XLSX, DOCX...) sont standardisés (ISO/IEC 29500). Mais les logiciels semblent parfois s'écarter du standard
-

Formats, pour résumer …

- Utiliser des **formats ouverts** afin de faciliter le partage et l'interopérabilité
- Si on utilise des formats fermés, il faut vérifier s'assurer de la **pérennité** et si la conversion altère les informations
- Le format doit être **documenté** (standards)
- Format textuel en **UTF-8** répond à la majorité des besoins :
  - excel ⟶ csv
  - word ⟶ txt/md/html
- Mais aussi :
  - zip ⟶ tar.gz
  - images ⟶ png/svg

https://fr.wikipedia.org/wiki/Format_ouvert#Les_principaux_formats_ouverts

Organisation des données



Organisation des dossiers

- Limitez le nombre de dossiers par niveau (5 ou 6 max)
- Allez du général au spécifique
- Choisissez des noms de dossiers explicites et non redondants

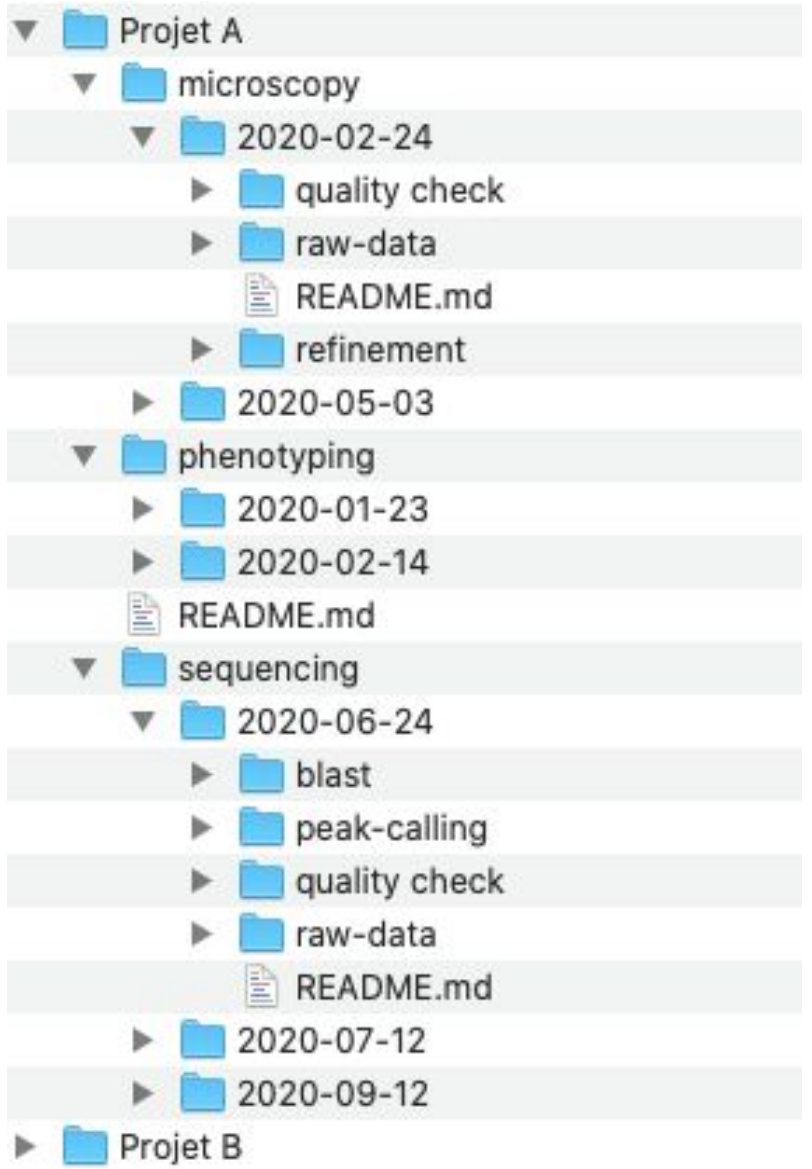
Pour chaque dossier, ajoutez un fichier README:

- Choisissez un format simple et ouvert (par exemple Markdown ou TXT)
- Indiquez un minimum de métadonnées concernant le dossier et son contenu:
  - Titre
  - Date de création / réception des données
  - Origine/Source des données
  - Version
  - Propriétaire/responsable des données
  - Organisation des données
  - Méthode de réception/téléchargement des données

Un exemple :

1. Un dossier par projet
   a. Un sous-dossier par type de manip (microscopie, séquençage, phénotypage)
      i. Un sous-dossier par date (2020-02-24, 2020-05-03)
         1. Un sous-dossier pour les données brutes
         2. Un sous-dossier par analyse (contrôle qualité, nettoyage statistique, raffinement)
      ii. Un sous-dossier par publication
         1. Un lien symbolique vers chaque dossier données ou analyse associé à la publication

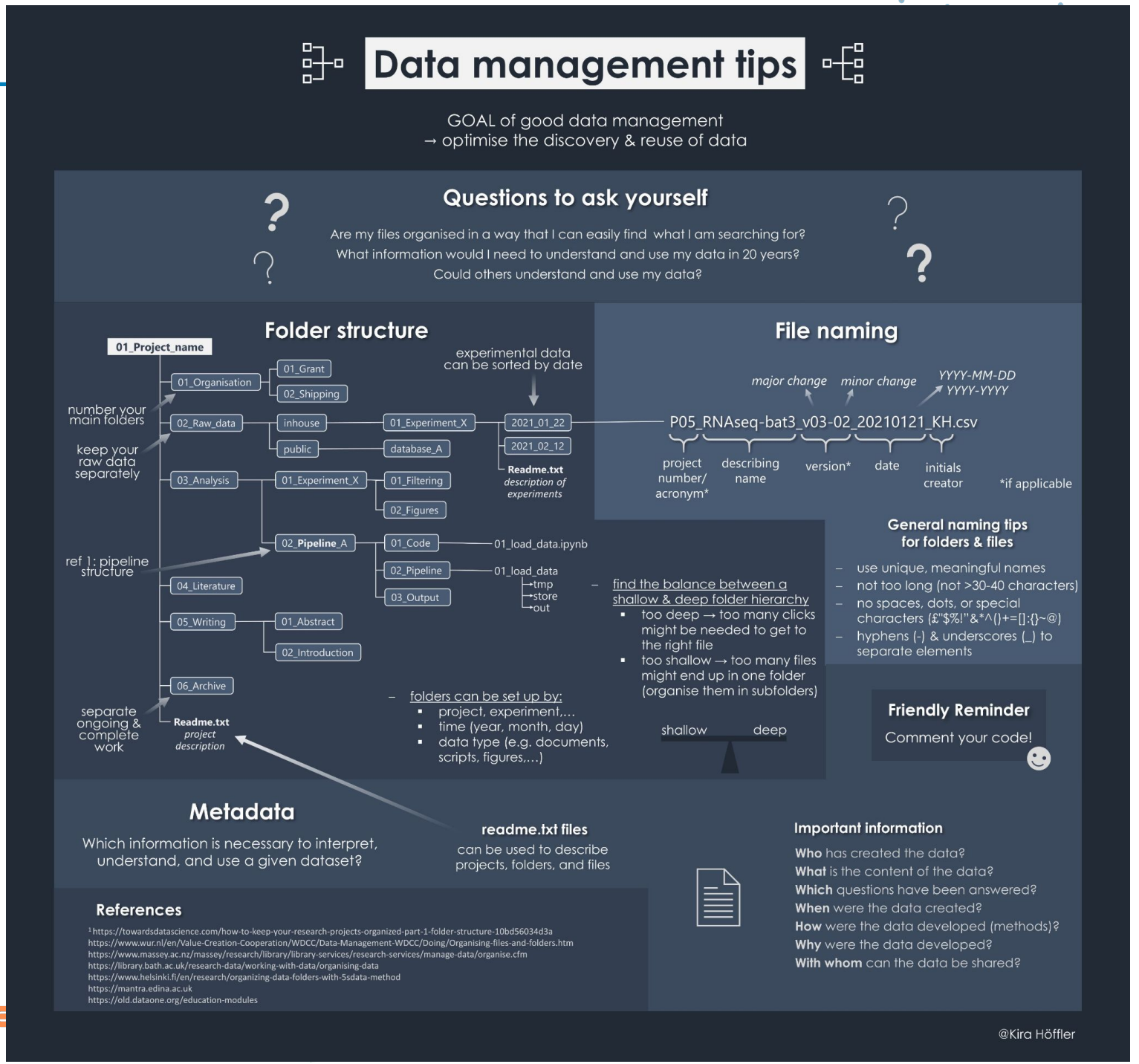
- Dispo sur zenodo : https://zenodo.org/record/4410128#.YjiRpDXjJD9



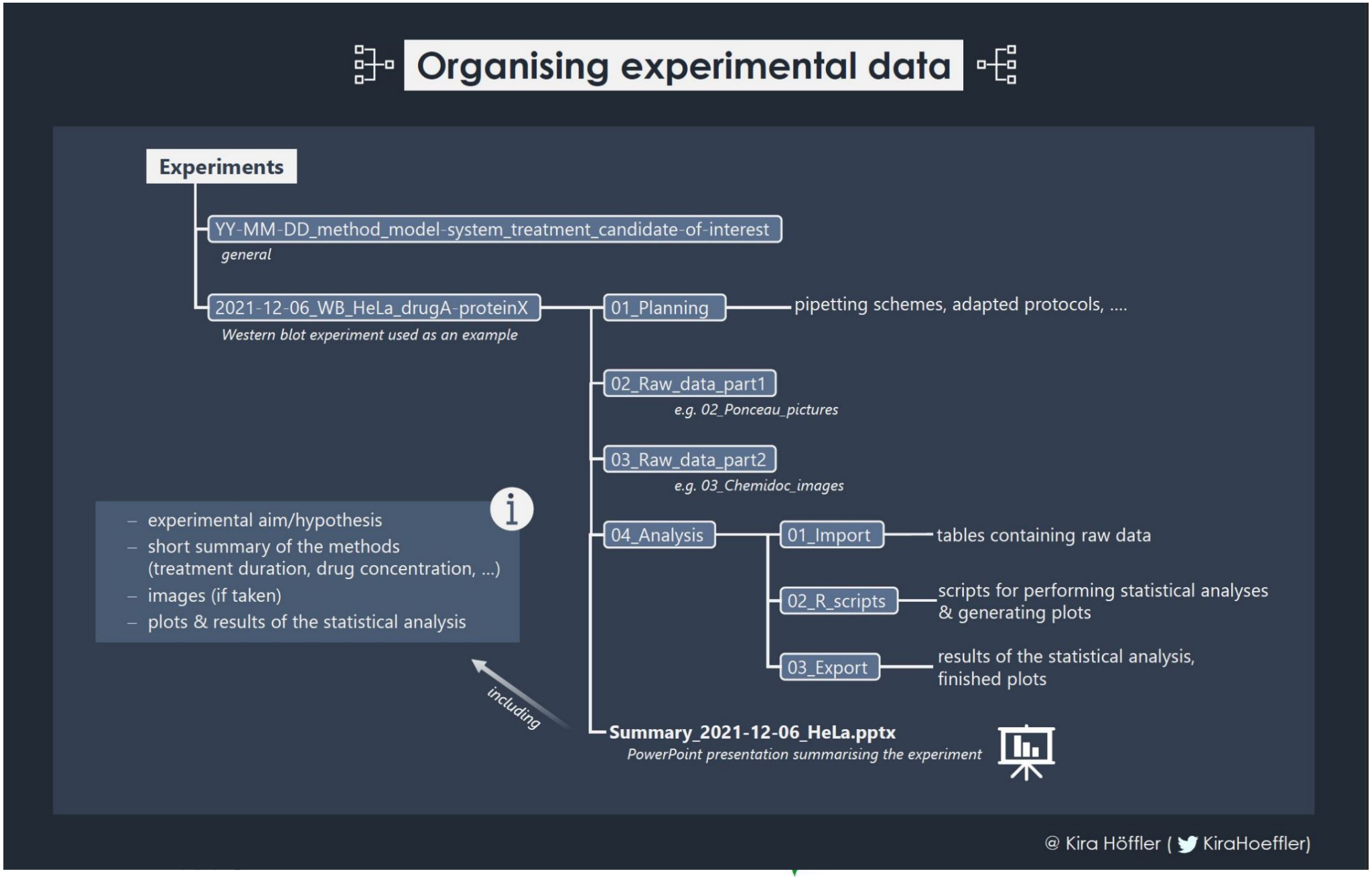
Un autre exemple :

Il n'y a pas de solution miracle mais il est important de se mettre d'accord au sein du projet et de le documenter dans le PGD

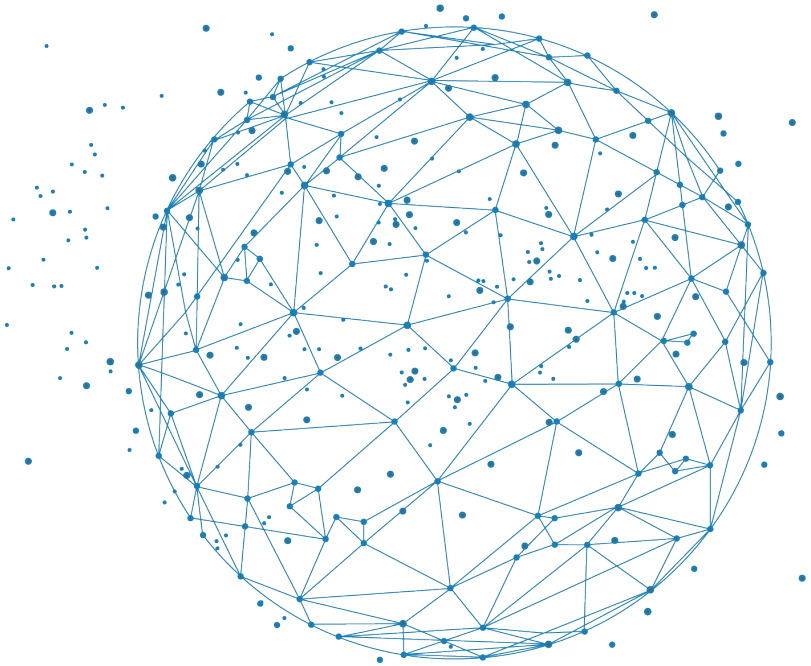
- Exemple de générateur d'arborescence https://www.tiesdekok.com/folder-structure-generator/#



Un autre exemple :



Suppression des données



Un cas d’usage

Est-ce que ces données peuvent être supprimés ?

https://www.wooclap.com/GEIRXI

Est-ce que ces données peuvent être supprimés ?

Le stockage des données a un coût financier et écologique.

- Distinguez clairement la copie principale (main) de ses dérivés
- Organisez régulièrement une revue des données
- Récupérer rapidement les données sur supports externes (disque ou clé USB)

Je veux garder mes données pour l'éternité

- Quels sont vos obligations en terme de rétention de données
- Dans quelles conditions allez-vous les archiver ?
- Avez-vous documenter clairement vos données ?
- Que se passera-t-il si vous partez (pour l'éternité) ?

Les infrastructures de stockage sont vos amies

- Politique de sauvegarde professionnel et cohérente
- Nombre de copies minimum (stratégie 3-2-1)
- Gestion claires des droits d'accès
- Haute disponibilité et accessibilités
- Sécurité

Où se situe mon fichier ?

0% ⟷ 100%

Lisible par	Moi	Mon équipe	Ma communauté	D'autres communautés	Le monde entier
Format	Propriétaire fermé		Propriétaire ouvert		Ouvert
Format	En évolution				Stable
Description	Pas de schema(.org)				schema
Langage du format	Propriétaire				Norme
Formalisation	Pas de norme		Norme propriétaire		Norme ISO

IFB
INSTITUT FRANÇAIS DE BIOINFORMATIQUE

elixir FRANCE

Module 2

Pratiques d'hygiène numérique pour la gestion des données

Outils et solutions

ARTbio IBPS

migale

INSTITUT PASTEUR

PLANT BIOINFORMATICS FACILITY

cnrs

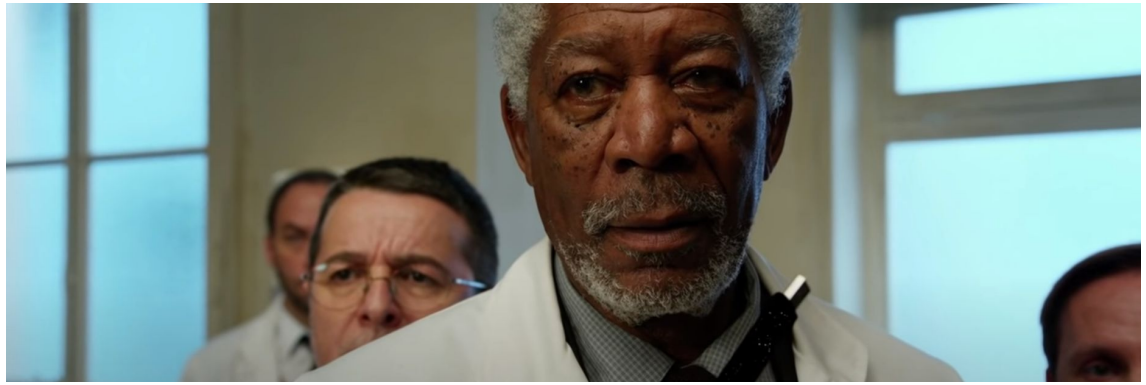
INRAE

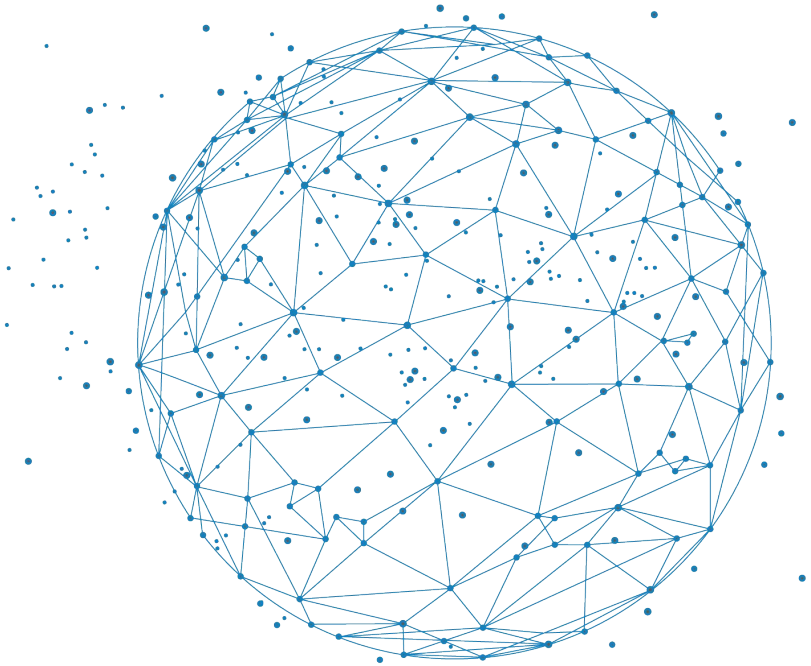
Inserm

cea

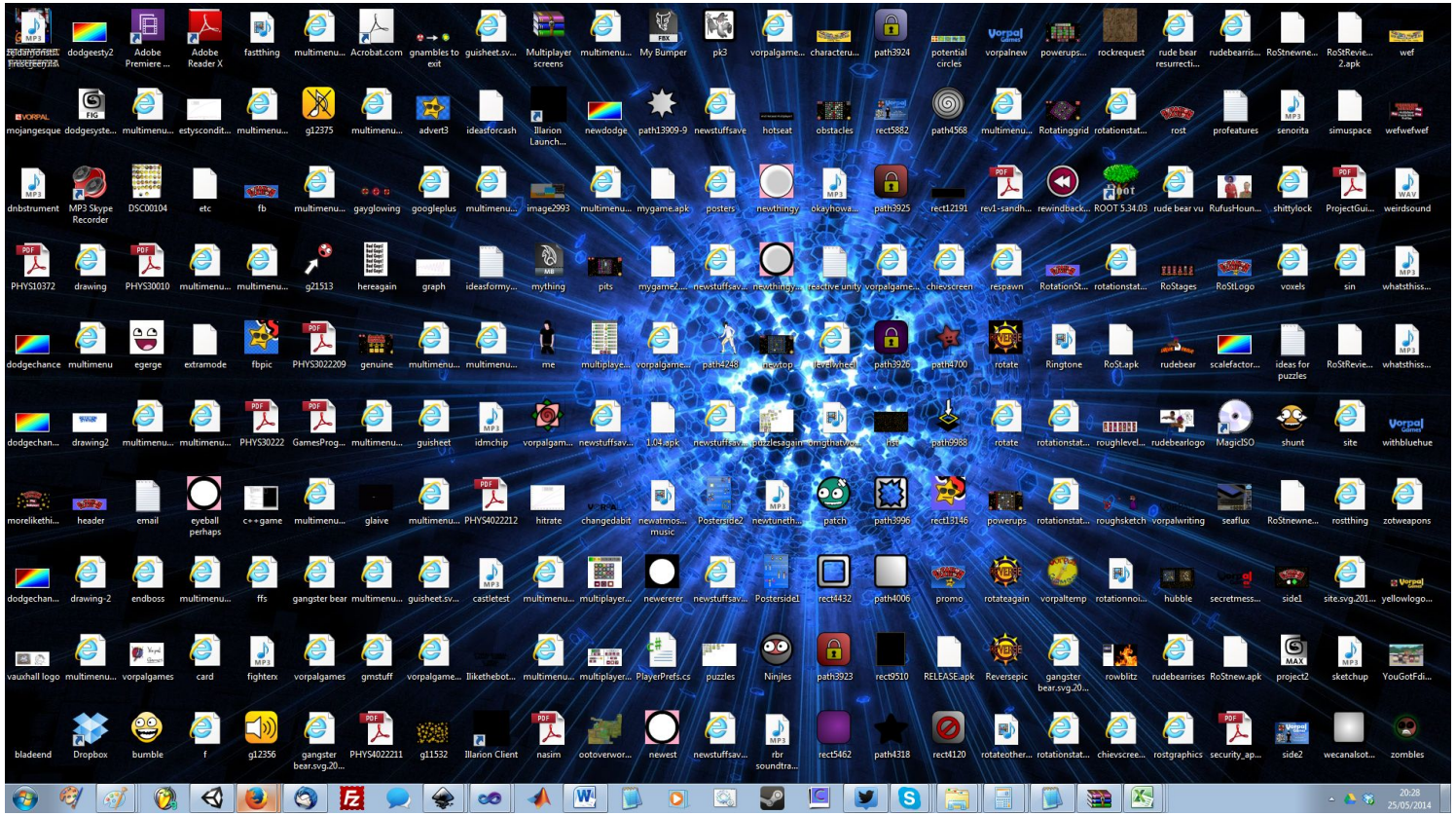
INVESTIR L'AVENIR

Gestion Électronique de Documents





Il est nécessaire d'utiliser des outils pour gérer un espace de **préservation et de partage du savoir du groupe**

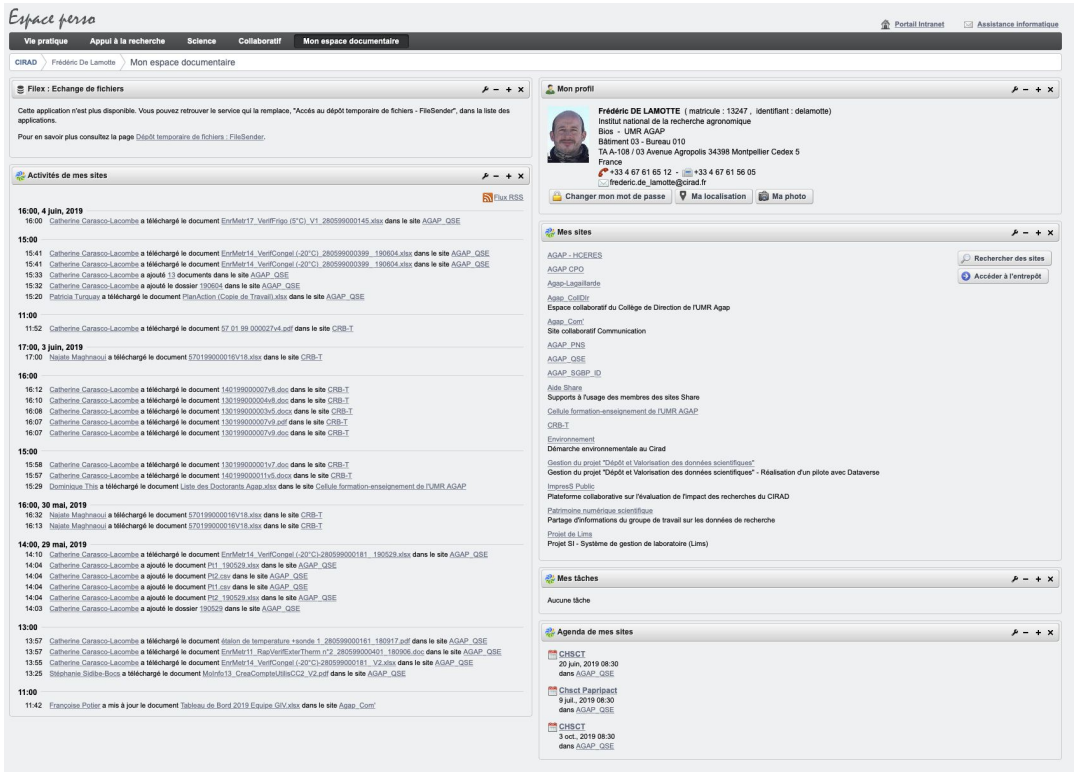


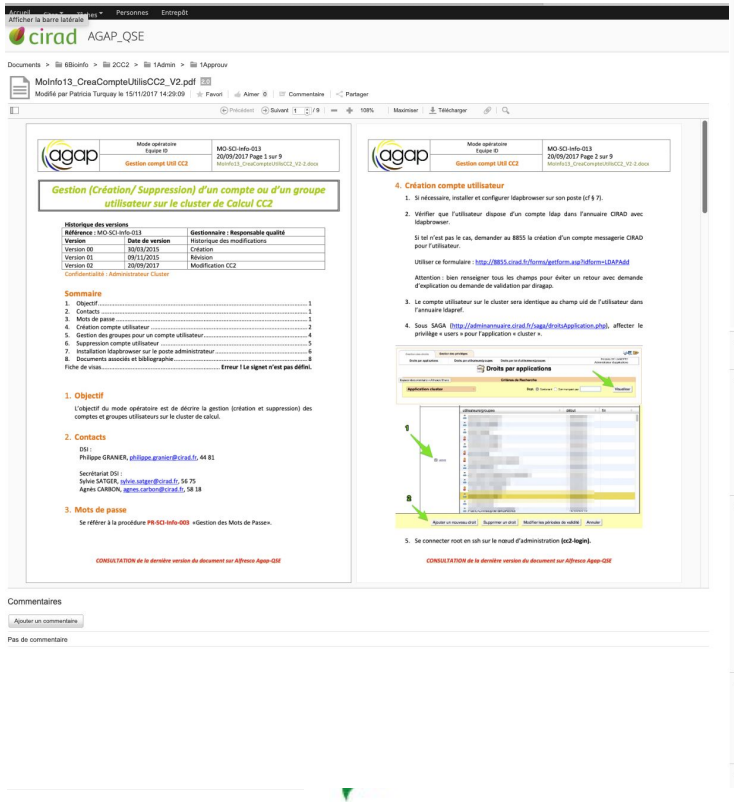
Simulation du bureau d'Hélène

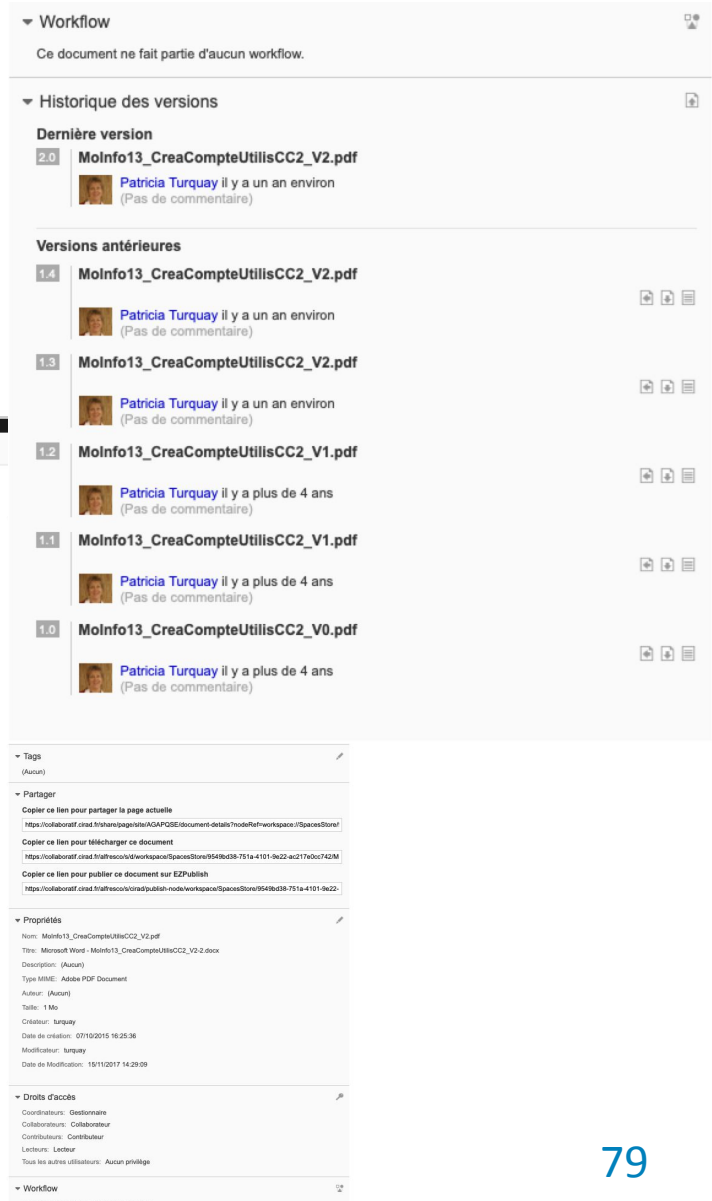
GED

De nombreux outils sont dispo :

- Google drive (ne faites pas ça !!)
- OSF
- Alfresco





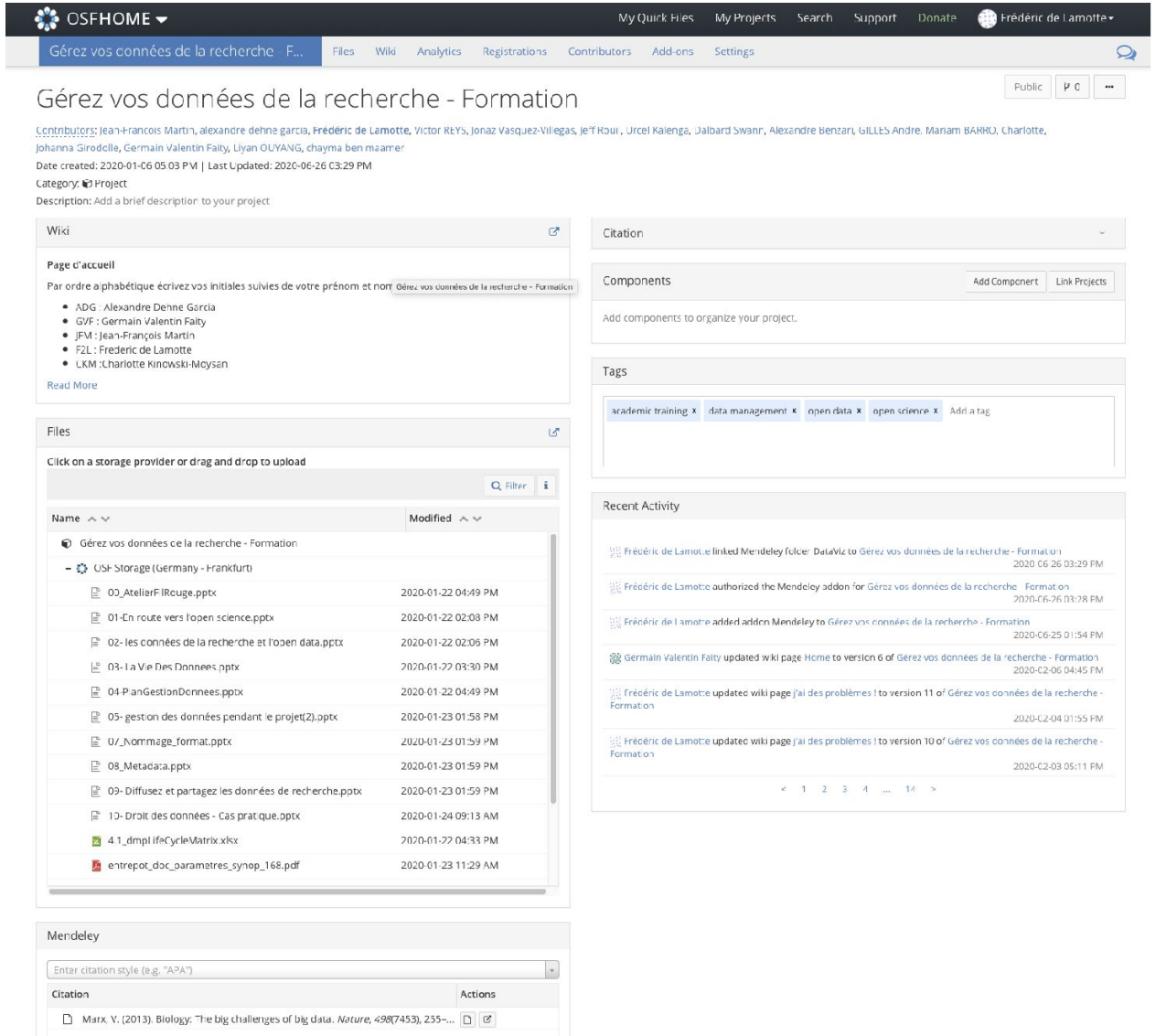


GED

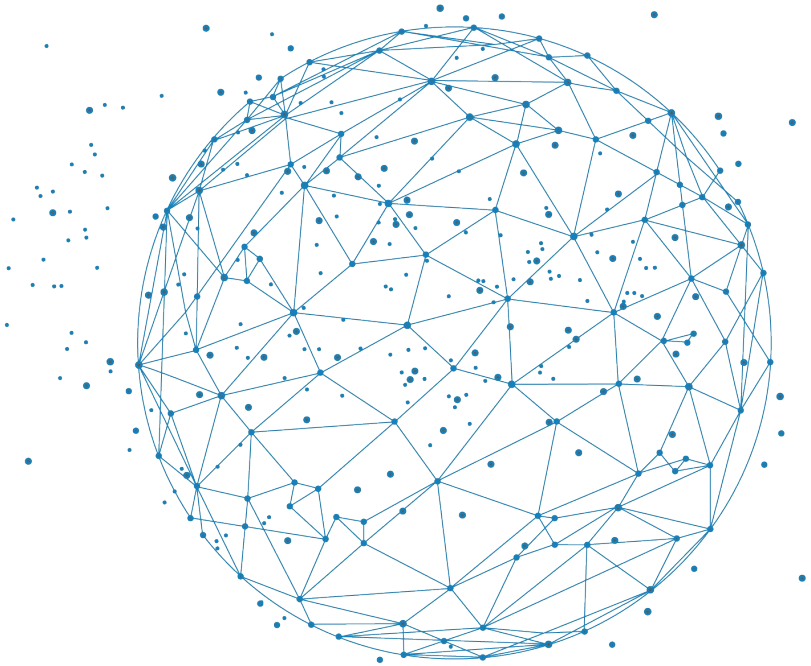
Avoir une GED c’est bien.

Bien s’en servir, c’est mieux !

- Où sont localisé les serveurs ?
- Sauvegardes ?
- Qui gère les droits ?
- Comment est organisé la GED ?
- Edition synchrone et gestion des conflits ?
- Partage avec l’extérieur ?
- gestion des métadonnés
- workflow de validation



Document computationnel

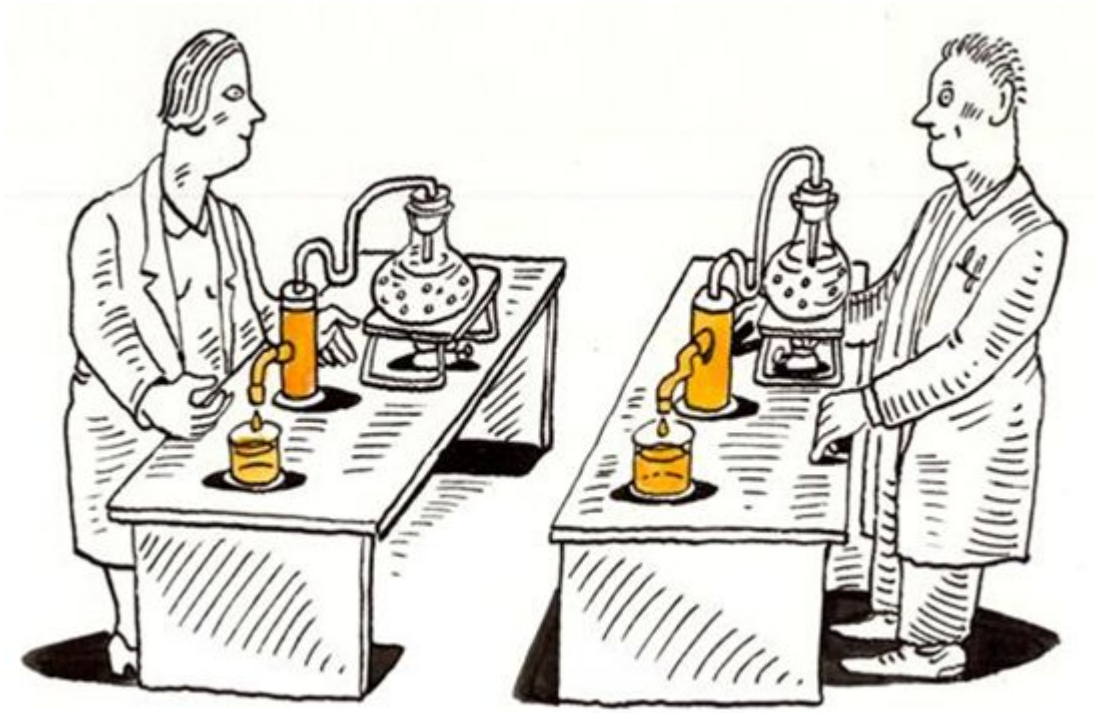


Documents computationnels

- Il faut se donner les moyens pour qu'autrui puisse inspecter nos analyses
- Expliciter augmente les chances de trouver les erreurs et de les éliminer

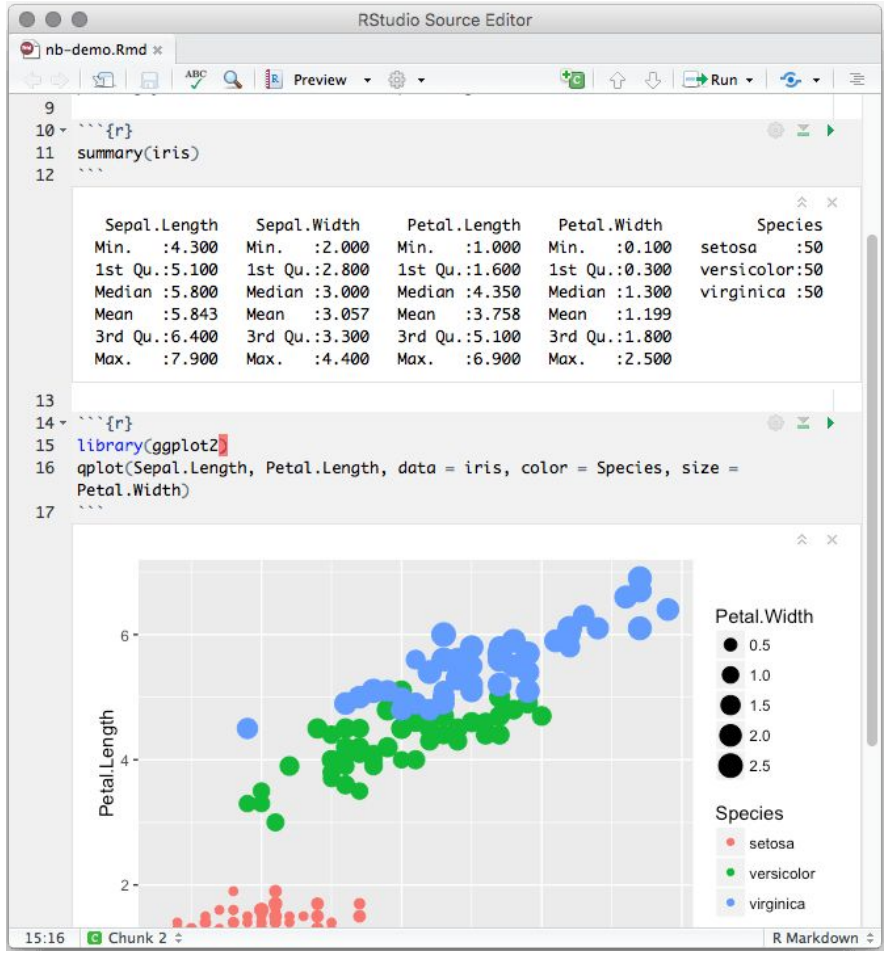
"Inspecter pour justifier et comprendre"

"Refaire pour vérifier, corriger et réutiliser"



Documents computationnels - **rmarkdown**

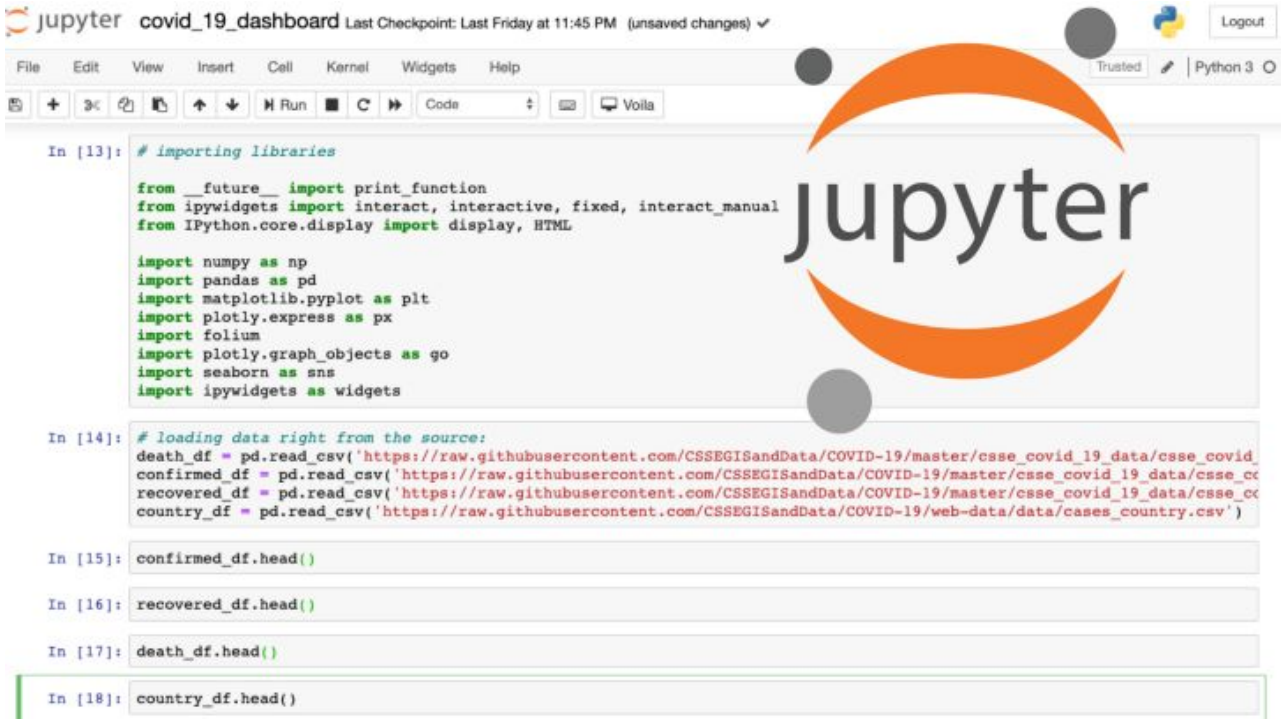
- Regrouper dans un unique document
  - Les informations de contexte
  - Les versions des outils et des banques
  - Le code
  - Les calculs et résultats
  - L'interprétation
- Assurer la cohérence des analyses et améliorer la traçabilité
- Génère un document exportable (ex : html) pour une meilleure portabilité et lisibilité.



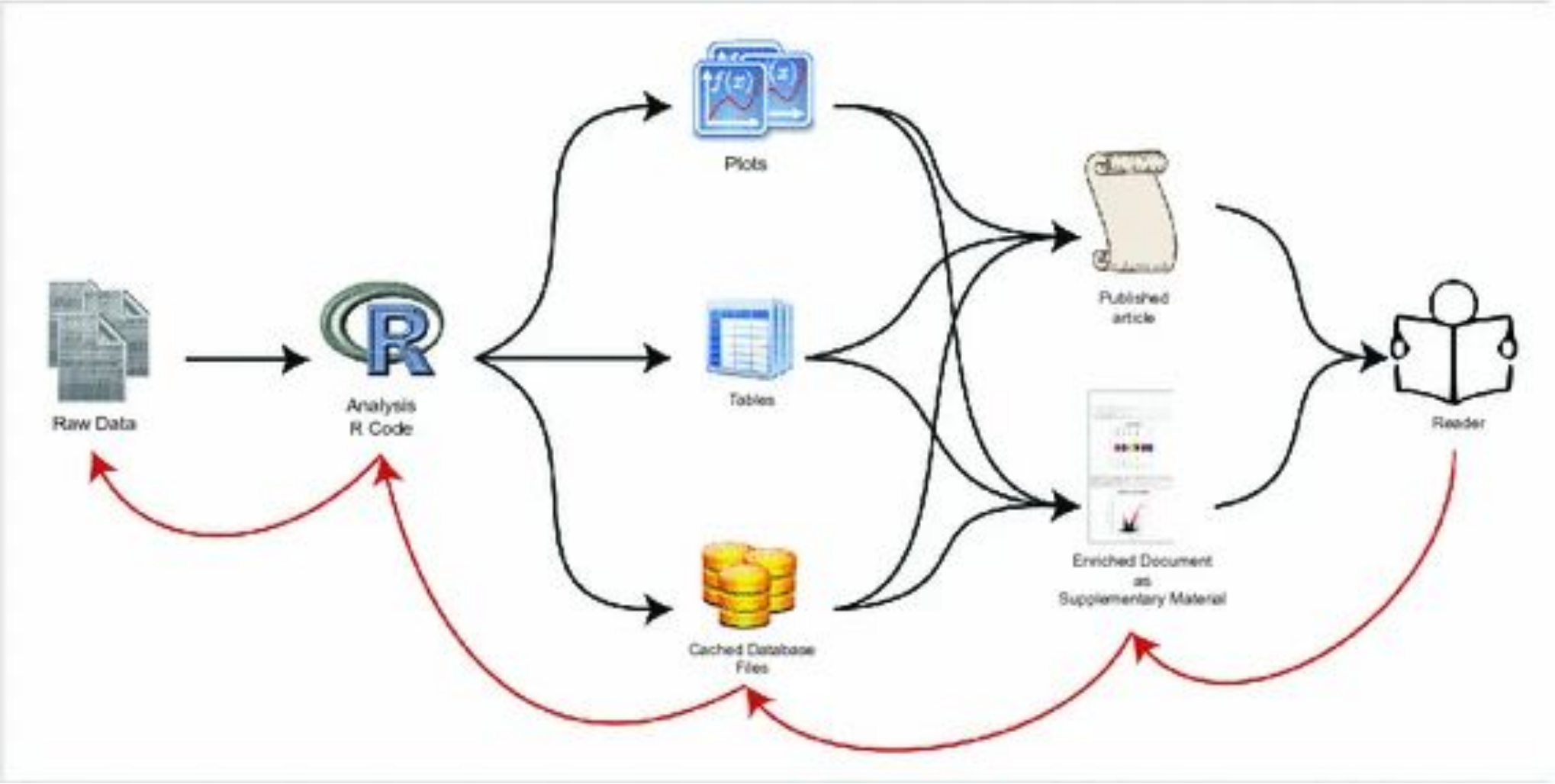
https://rmarkdown.rstudio.com/

Documents computationnels - **jupyter notebook**

- Regrouper dans un unique document
  - Les informations de contexte
  - Les versions des outils et des banques
  - Le code
  - Les calculs et résultats
  - L'interprétation
- Assurer la cohérence des analyses et améliorer la traçabilité
- Génère un document exportable (ex : html) pour une meilleure portabilité et lisibilité.



https://jupyter.org/

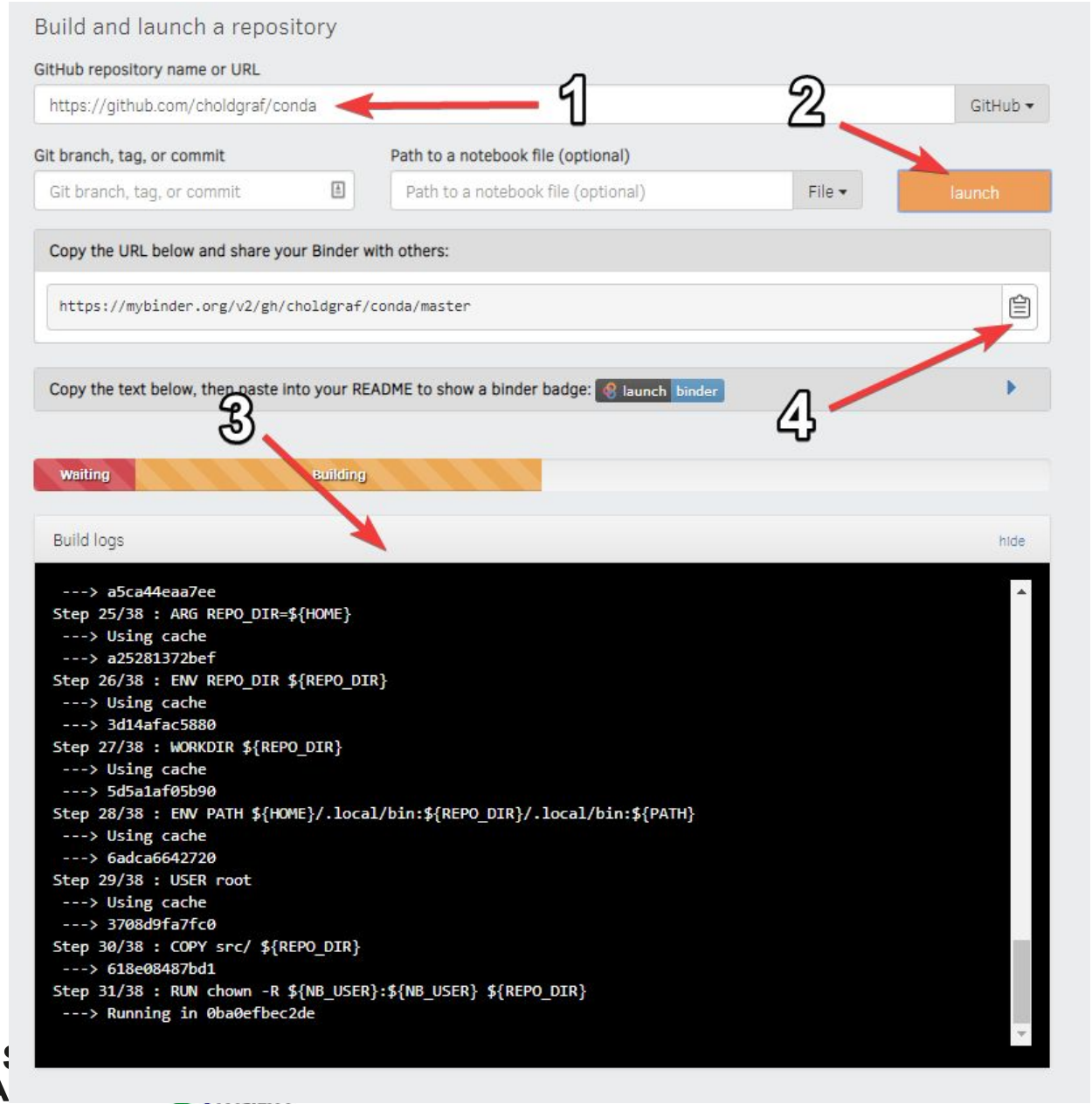


doi.org/10.1007/978-3-319-44332-4_19

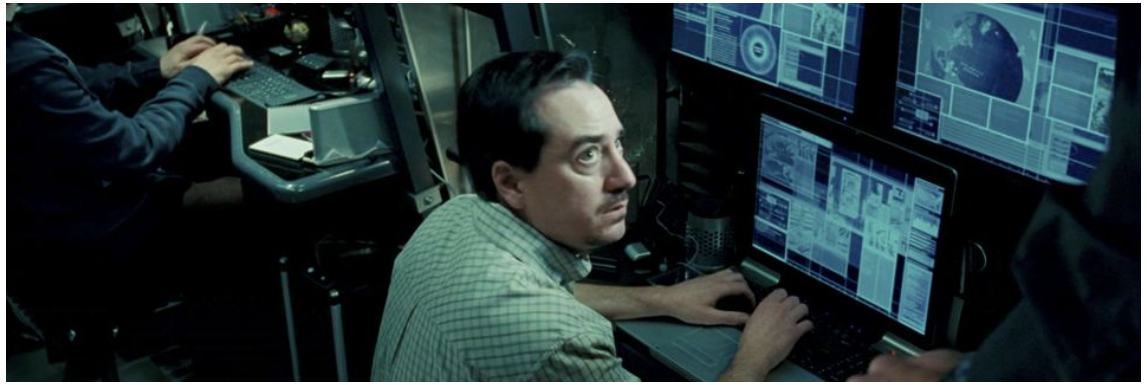
mybinder.org

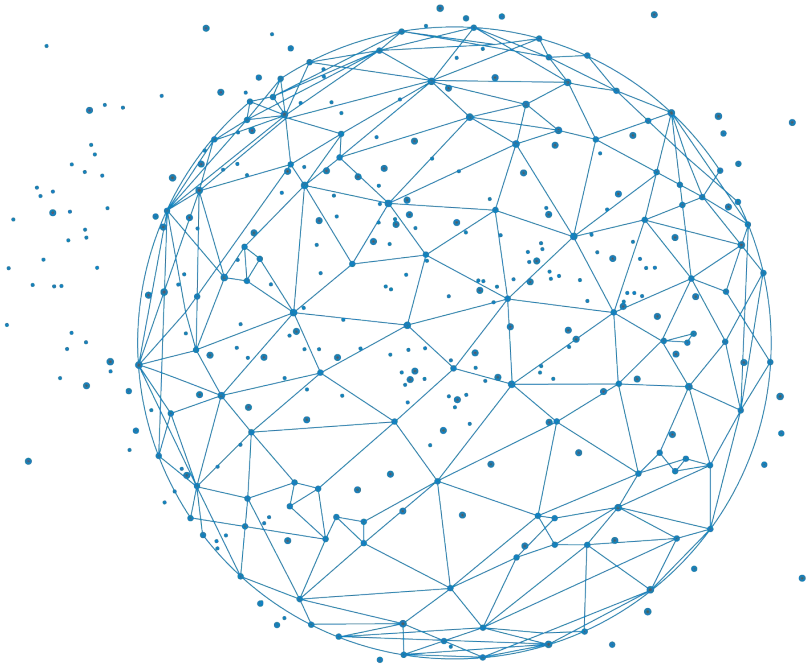
- *Turn a Git repo into a collection of interactive notebooks*

- Utilise les outils d'intégration continue
- Ré-exécute le notebook à chaque commit
- Assure la cohérence des résultats

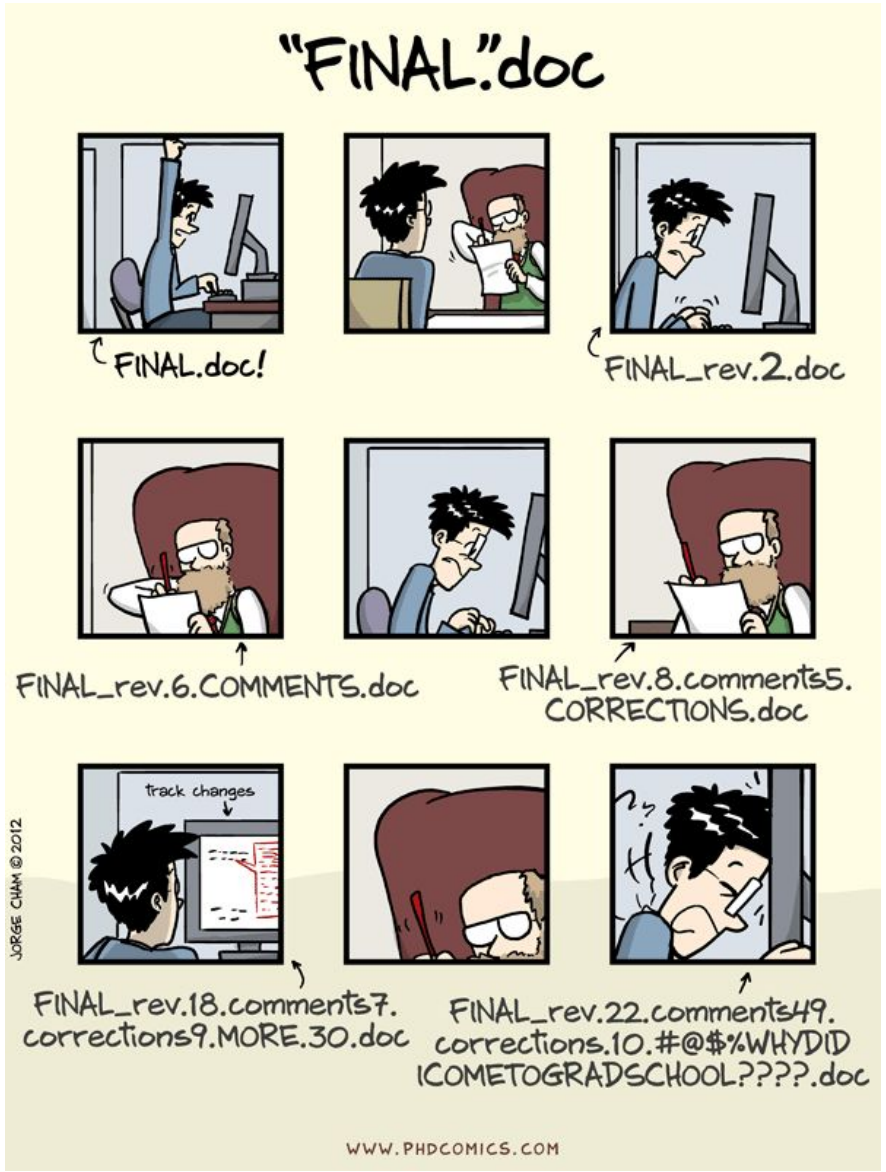


Gestion des codes sources





Gestion des codes sources

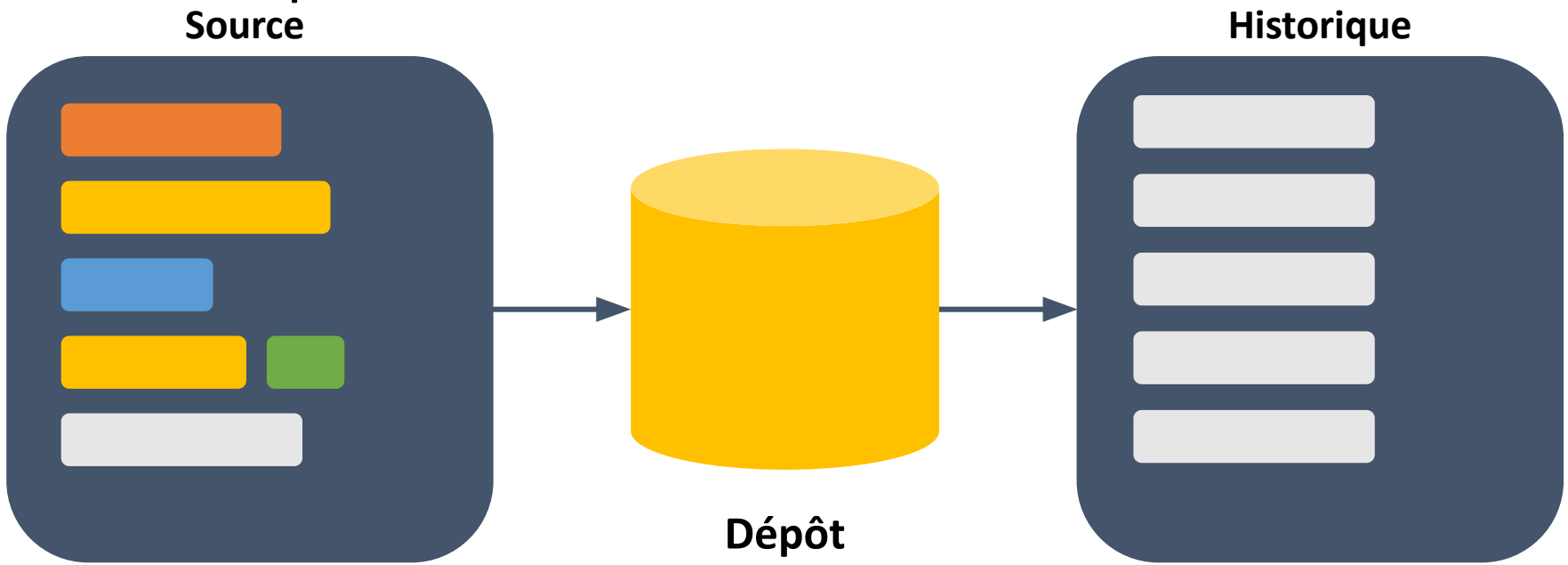


PhD Comics

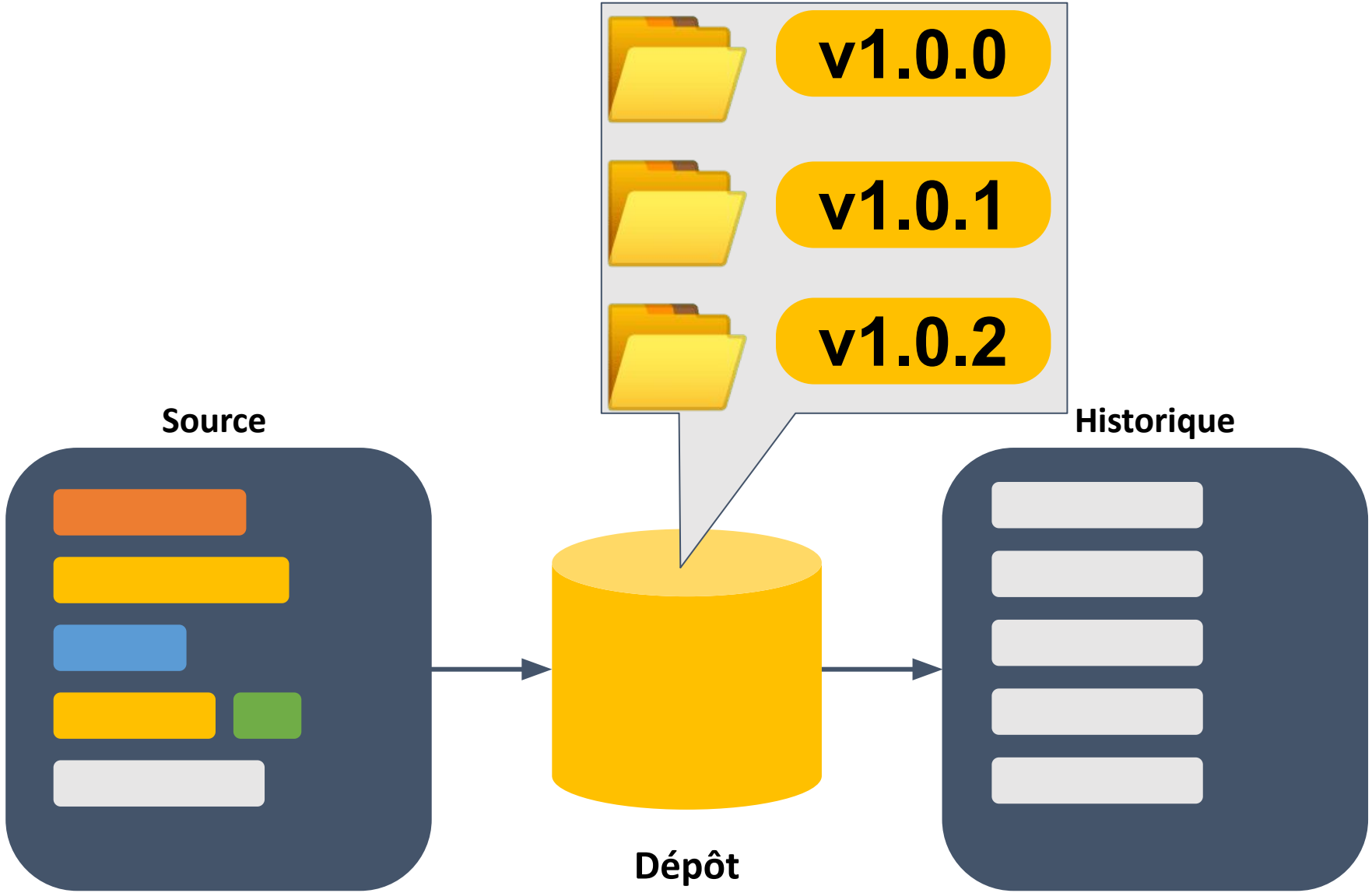
Git est un système de gestion de versions

Permets de :

- Suivre l'évolution des fichiers
- Faciliter le développement collaboratif
- Revenir à une version précédente



C’est quoi Git ?



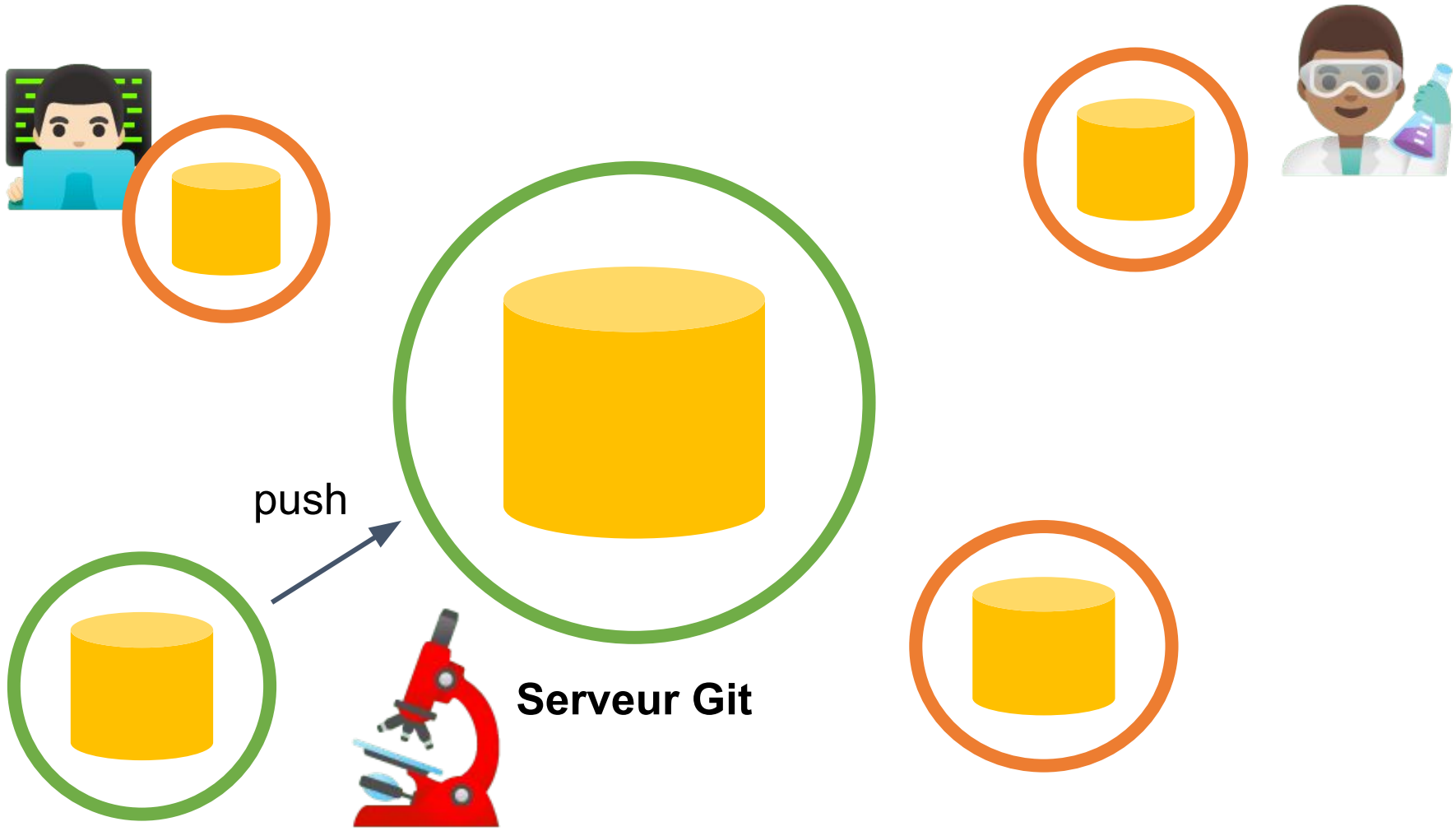
C’est quoi Git ?

Serveur Git

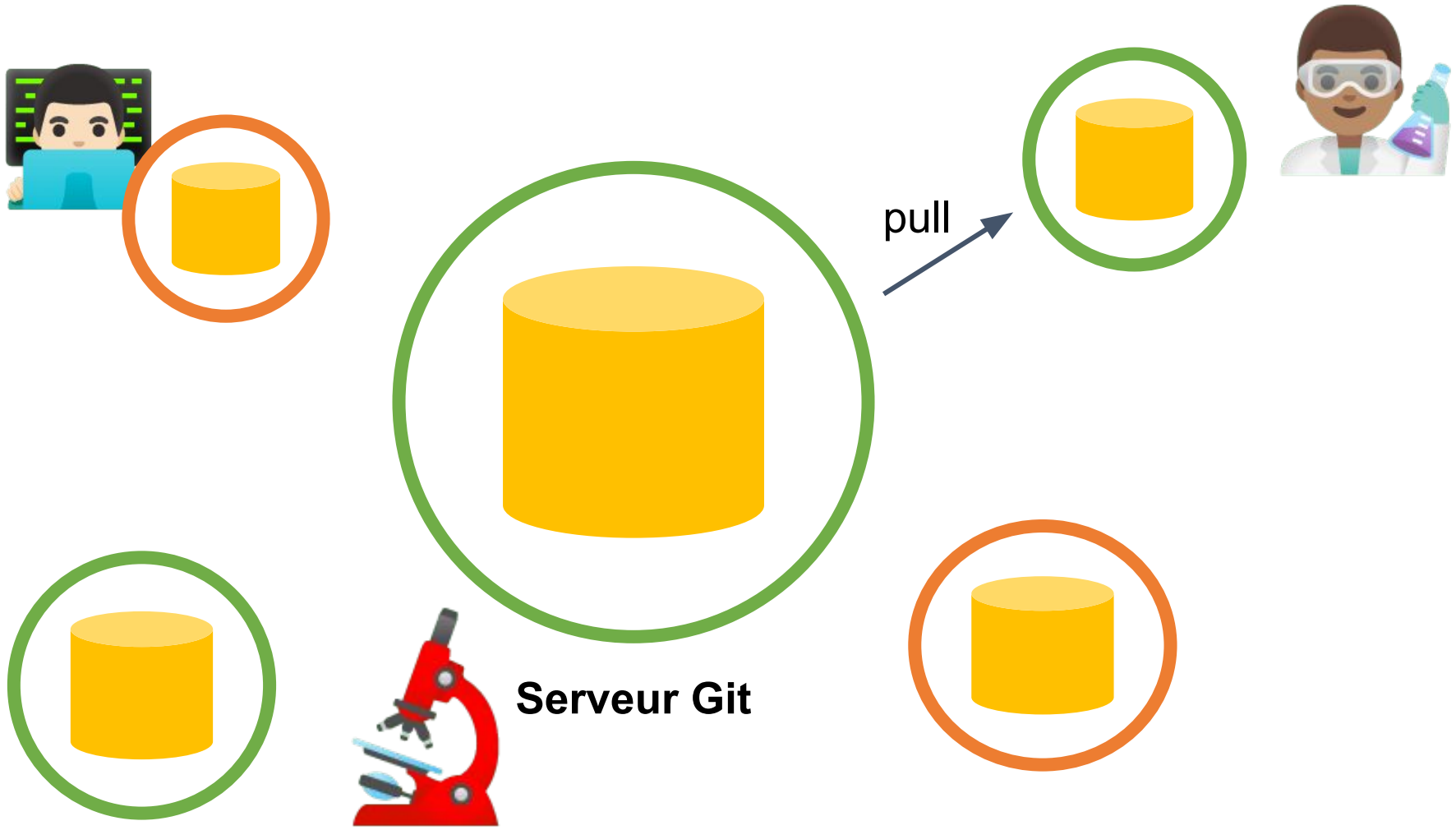
GitLab

GitHub

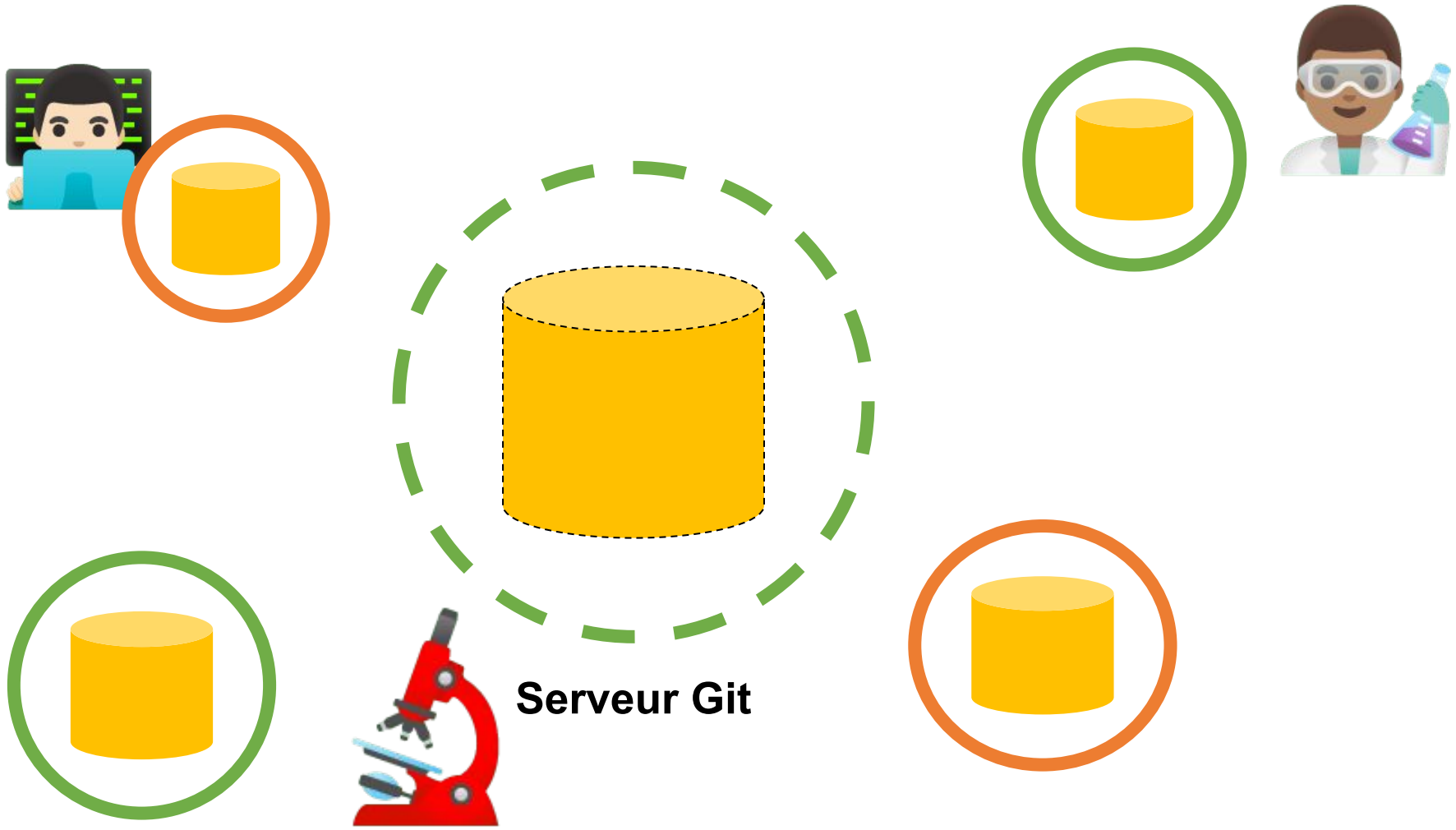
C’est quoi Git ?



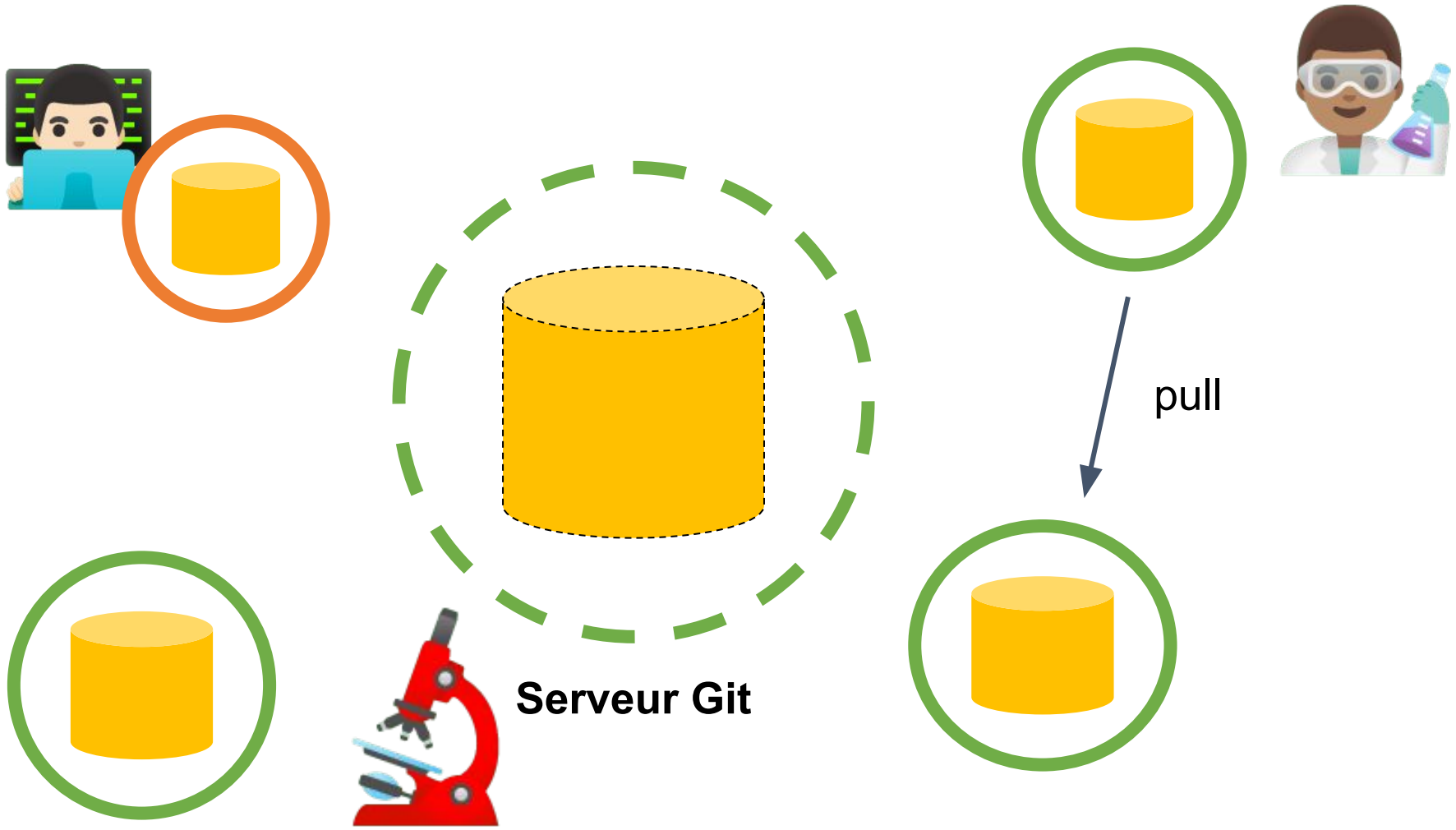
C’est quoi Git ?

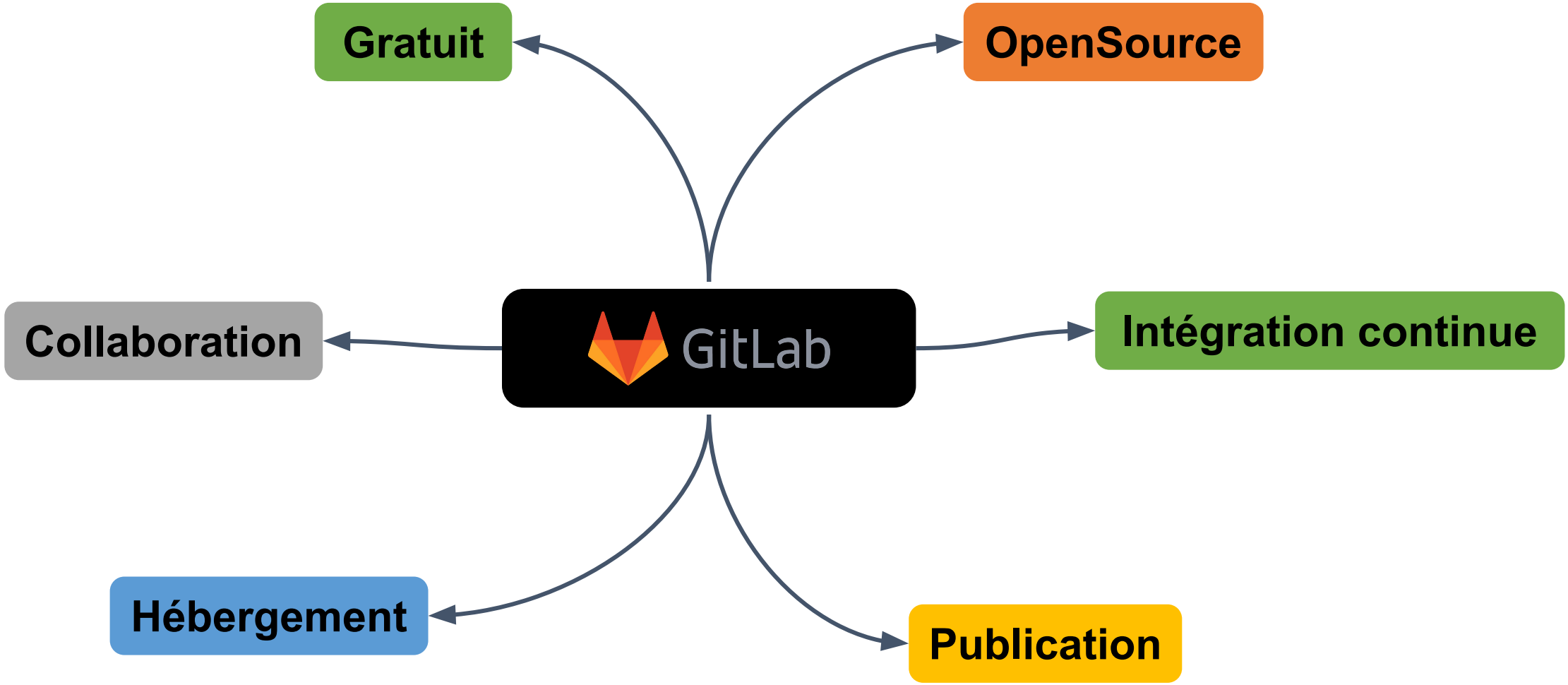


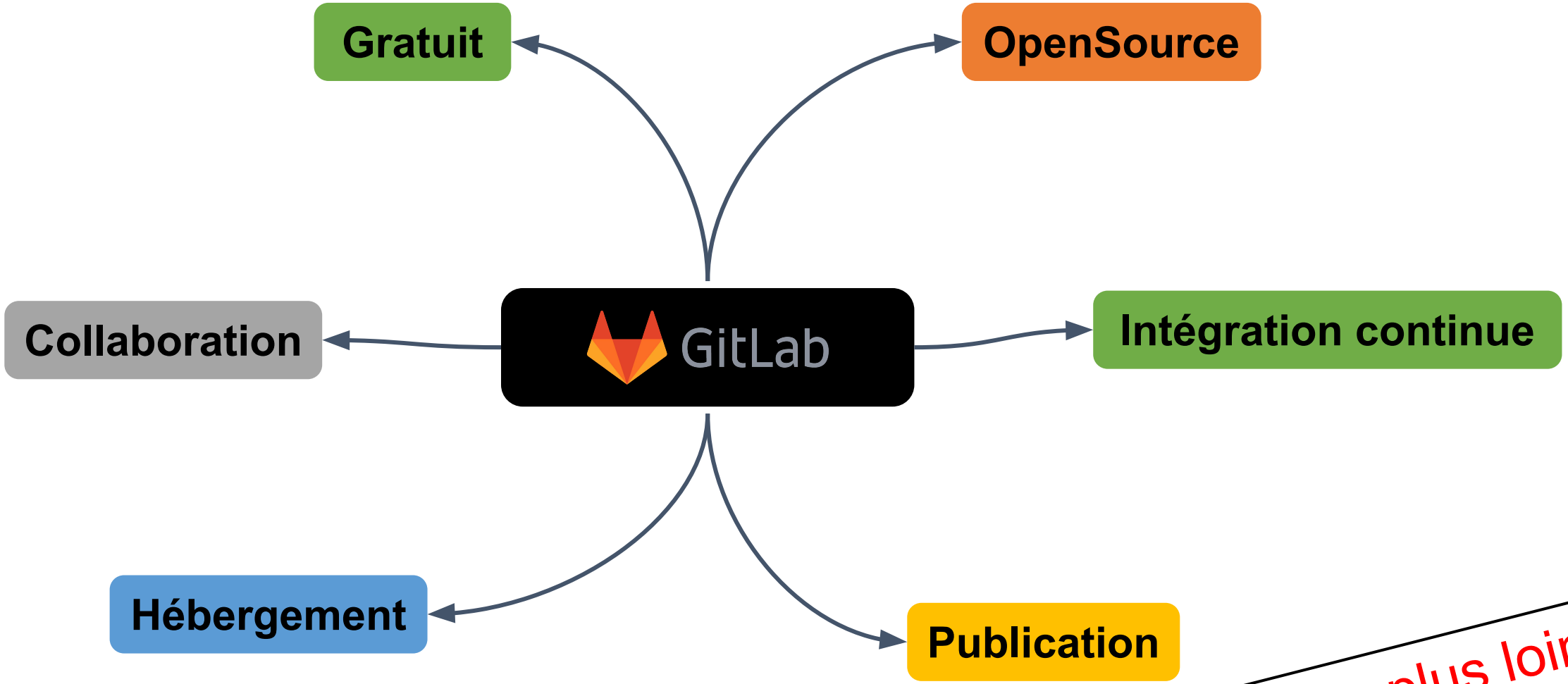
C’est quoi Git ?



C’est quoi Git ?



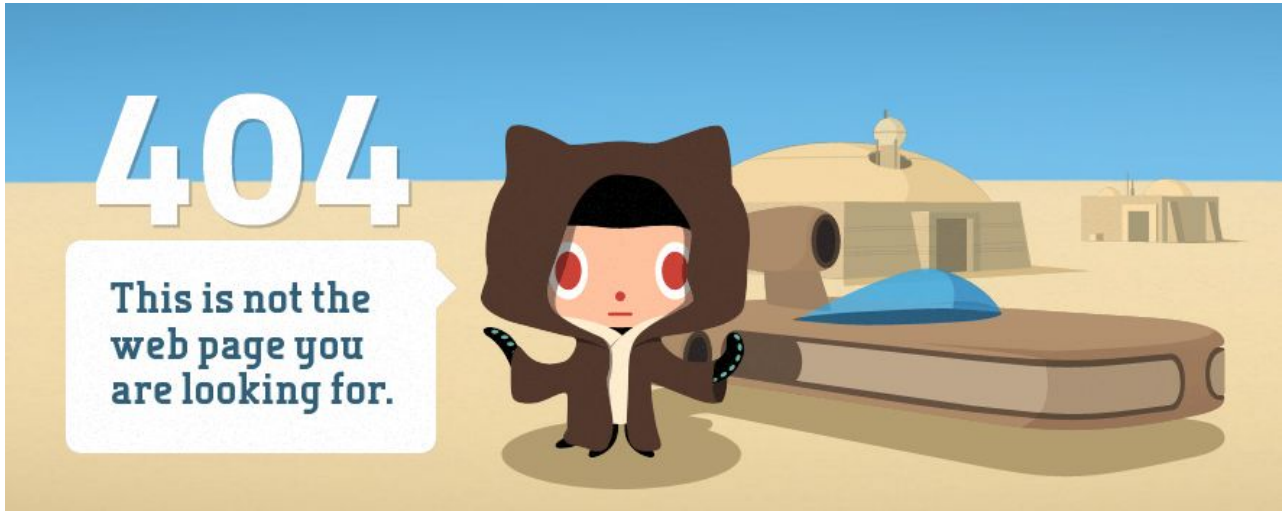

Gratuit
OpenSource
Collaboration
GitLab
Intégration continue
Hébergement
Publication



Pour aller plus loin :
IFB – FAIR-Bioinfo

Software Heritage

- Attention GitHub et GitLab ne sont pas des outils d'archivage du code au long terme.
- Un dépôt peut-être supprimé



- Software Heritage a pour ambition est de **collecter**, **préserver** et **partager** tous les logiciels disponibles publiquement sous forme de code source.

Software Heritage

IFB
INSTITUT FRANÇAIS DE BIOINFORMATIQUE

elixir FRANCE

Module 2

Pratiques d'hygiène numérique pour la gestion des données

Un dernier exercice pour la route

ARTbio IBPS

migale

INSTITUT PASTEUR

PLANT BIOINFORMATICS FACILITY

cnrs

INRAE

Inserm

cea

INVESTIR L'AVENIR

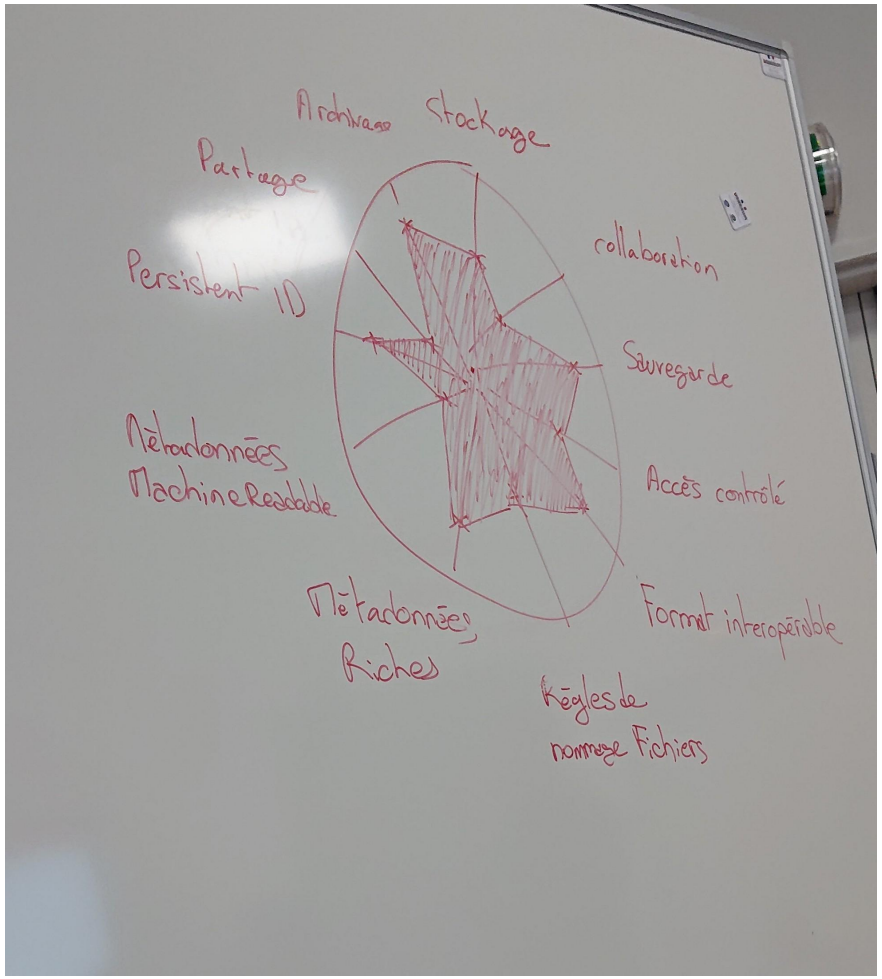
FAIR_des_crepes.zip (*ya un jeu de mot subtile …*)

Ce weekend j'ai fait des crêpes, vous trouverez le projet ici.

- Lister tout ce qui ne va pas
- Faites des suggestions d'amélioration

Télécharger la matrice Excel
modèle radar.xlsx sur osf.io

Donnez une note de 0 à 5 pour
chaque critère pour votre fichier



IFB
INSTITUT FRANÇAIS DE BIOINFORMATIQUE

elixir FRANCE

Module 2

Pratiques d'hygiène numérique pour la gestion des données

Sources

ARTbio IBPS

migale

INSTITUT PASTEUR

PLANT BIOINFORMATICS FACILITY

cnrs

INRAE

Inserm

cea

INVESTIR L'AVENIR

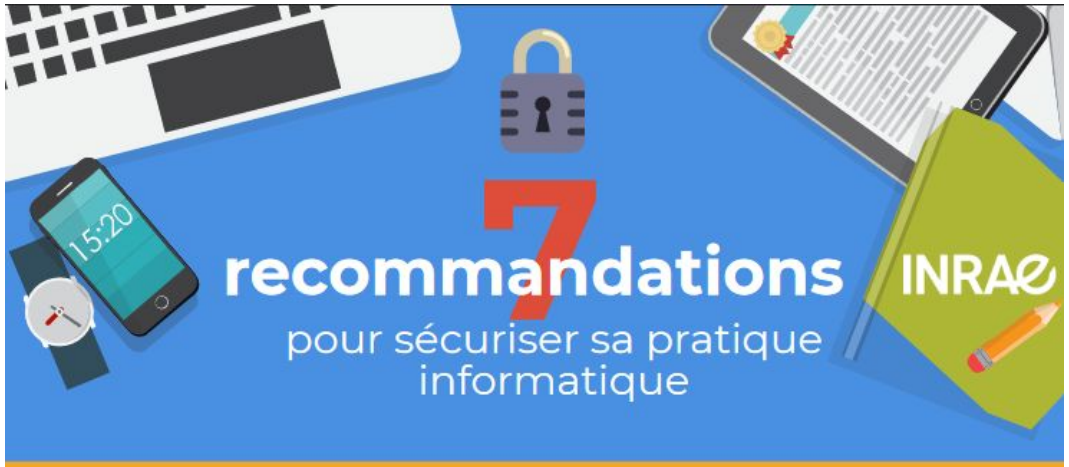
Suppléments

- Conseils d'hygiène numérique de l'ANSII
  - https://www.ssi.gouv.fr/administration/bonnes-pratiques/

- 7 recommandations cyber INRAE
  - https://intranet.inrae.fr/cybersecurite

- Ten simple rules for :
  - Digital Data Storage 10.1371/journal.pcbi.1005097
  - providing effective bioinformatics research support 10.1371/journal.pcbi.1007531
  - to make your computing more environmentally sustainable 10.1371/journal.pcbi.1009324
  - 5 reasons why researchers should learn to love the command line 10.1038/d41586-021-00263-0

- https://happygitwithr.com/



7 recommandations

pour sécuriser sa pratique informatique

INRAE

1 Mots de passe : faites preuve d'imagination et diversifiez

Un bon mot de passe contient au minimum 12 caractères (minuscule, majuscule, chiffres et caractères spéciaux), et doit être propre à chaque compte utilisé. Pour les mémoriser, utilisez un gestionnaire de mots de passe.

- **Choisir avec soin ses mots de passe** https://www.cnil.fr/fr/les-conseils-de-la-cnil-pour-un-bon-mot-de-passe
- **Utiliser le gestionnaire KeePass**

2 Mises à jour : n'attendez plus

Les mises à jour corrigent les vulnérabilités appréciées des attaquants. Dès que cela vous est demandé, procédez à la mise à jour de sécurité de vos équipements : systèmes d'exploitation, applications mais aussi les composants.

Assurez-vous que l'Anti-Virus fonctionne et qu'il est bien à jour.

- **Rapprochez-vous de votre informaticien ou de votre équipe informatique de centre pour plus d'informations en cas de besoin**

3 Sauvegardes : l'atout sérénité

Veillez à ce que des sauvegardes régulières de vos données soient effectuées. Une fois les données perdues ou détruites, sans sauvegarde, il est trop tard...

- **Utilisez les solutions de stockage ou de synchronisation fournies par l'Institut ou effectuez des sauvegardes régulières sur un support externe déconnecté comme par exemple un disque dur externe**

4 Attention au clic de trop

Messagerie : méfiez-vous des apparences... Les pièces jointes ou les liens contenus dans certains courriels réservent parfois de mauvaises surprises. Les incohérences de fond ou de forme, comme les requêtes indiscrètes sont à prendre avec des pincettes

- **Passez le curseur de votre souris sur le lien hypertexte, sans cliquer, l'adresse apparaîtra. Si le nom de domaine vous est inconnu, c'est probablement une tentative d'escroquerie**

Téléchargement : Doutez de toute demande d'installation impromptue : mise à jour de flash, ajout d'extensions, plugins, mise à jour du lecteur vidéo... Ces sollicitations spontanées ne sont jamais sans arrière-pensée.

- **Téléchargez le logiciel ou l'application uniquement depuis le site de l'éditeur du logiciel ou demandez conseil à votre informaticien**

5 Séparation des usages : évitez la contagion

Pour limiter la propagation d'une action malveillante, séparez vos usages professionnels et personnels (messagerie, équipements...).

- **N'utilisez pas votre messagerie professionnelle à des fins personnelles, pensez à créer des adresses différentes, en fonction de chaque usage**
- **N'utilisez pas de services de stockage en ligne personnels type « Cloud gratuit » à des fins professionnelles, ou du moins pas sans autorisation de l'Institut et sans avoir pris les mesures de sécurité qui s'imposent**
- **Ayez une utilisation appropriée d'Internet au travail et, d'une manière générale, soyez attentifs à vos propos sur les réseaux sociaux. Une fois sur Internet, vos données vous échappent et font le bonheur des adeptes de « l'ingénierie sociale » (usurpation d'identité, espionnage...). Pour en savoir plus et réagir en cas de problème** https://www.cnil.fr/fr/maitriser-mes-donnees

6 Nomadisme : faites rimer mobilité et sécurité

En déplacement, attention et discrétion doivent guider vos usages : gardez vos appareils, supports et fichiers avec vous. Evitez d'utiliser les services en libre-service (bornes électriques, wifi gratuit d'hôtel...) ou les objets qui vous sont offerts (clés USB, goodies connectés...), ils peuvent avoir été configurés à des fins malveillantes.

- **Lors de vos missions à l'étranger, utilisez de préférence et si possible un ordinateur dédié**
- **Plutôt que de se connecter sur la borne wifi d'un hôtel, partagez la connexion 4G de votre forfait mobile et donnez l'accès seulement à votre ordinateur**
- **Utilisez le service VPN de l'Institut qui permet de se raccorder à une connexion sécurisée et chiffrée et d'accéder aux ressources internes de l'INRA**

7 En cas d'incident, quelques réflexes à avoir...

- si la machine a un « comportement étrange » pouvant laisser croire à un piratage ;
- si vous vous êtes fait voler votre matériel ou si vous l'avez perdu ;
- s'il y a eu un clic de trop, si vous avez donné votre mot de passe trop vite, si vous avez un doute...

- **Rapprochez-vous immédiatement de votre informaticien ou de votre équipe informatique de centre**
- **Quoi qu'il arrive par précaution : déconnectez la machine du réseau pour stopper l'attaque, n'éteignez pas votre ordinateur et changez vos mots de passe**

Pour en savoir plus ...

https://intranet.inrae.fr/cybersecurite/

SSI
SÉCURITÉ DES SYSTÈMES D'INFORMATION

Merci !
mais au fait…
vous les avez tous ?

Imitation game

La vie aquatique

Contagion

Alien, le retour

Indiana Jones

Bruce Tout Puissant

Merlin, l'enchanteur

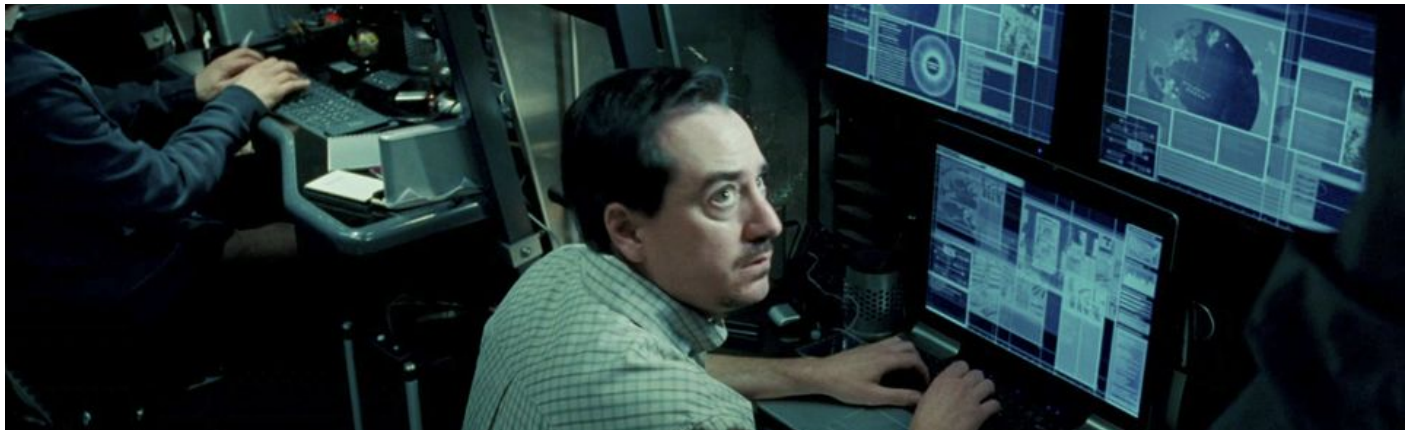
Pulp Fiction

La Momie

Matrix, reloaded

Wall-E

Lucy



Die Hard, retour en enfer

Harry Potter and the Prisoner of Azkaban