

# GRADUATE CAMPUS

Gestion des données de recherche pour  
jeunes chercheurs et chercheuses

[Jean-Blaise.Claivaz@unige.ch](mailto:Jean-Blaise.Claivaz@unige.ch)

[Claire.Wuillemin@unige.ch](mailto:Claire.Wuillemin@unige.ch)

[Talal.Zouhri@unige.ch](mailto:Talal.Zouhri@unige.ch)

Novembre 2021

LIBRARY



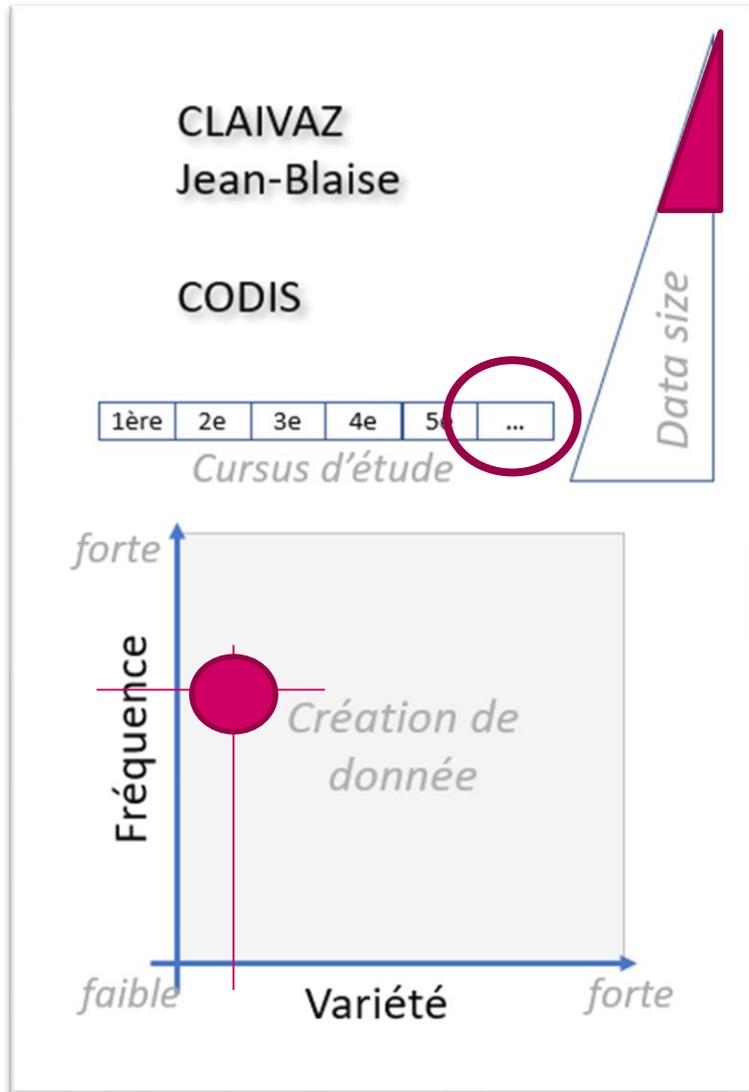
UNIVERSITÉ  
DE GENÈVE

# DÉROULÉ

- 1 Accueil et programme de la journée
- 2 Introduction générale sur l'Open Science
- 3 Données de recherche : définitions, organisation, description, formats & outils
- Repas de midi -----
- 4 Dépôts de données, copyrights et licences
- 5 Data Management Plan (DMP)
- 6 Discussions & retour sur la journée
- 7 Jeu : Escape the Lab



# ECHAUFFEMENT



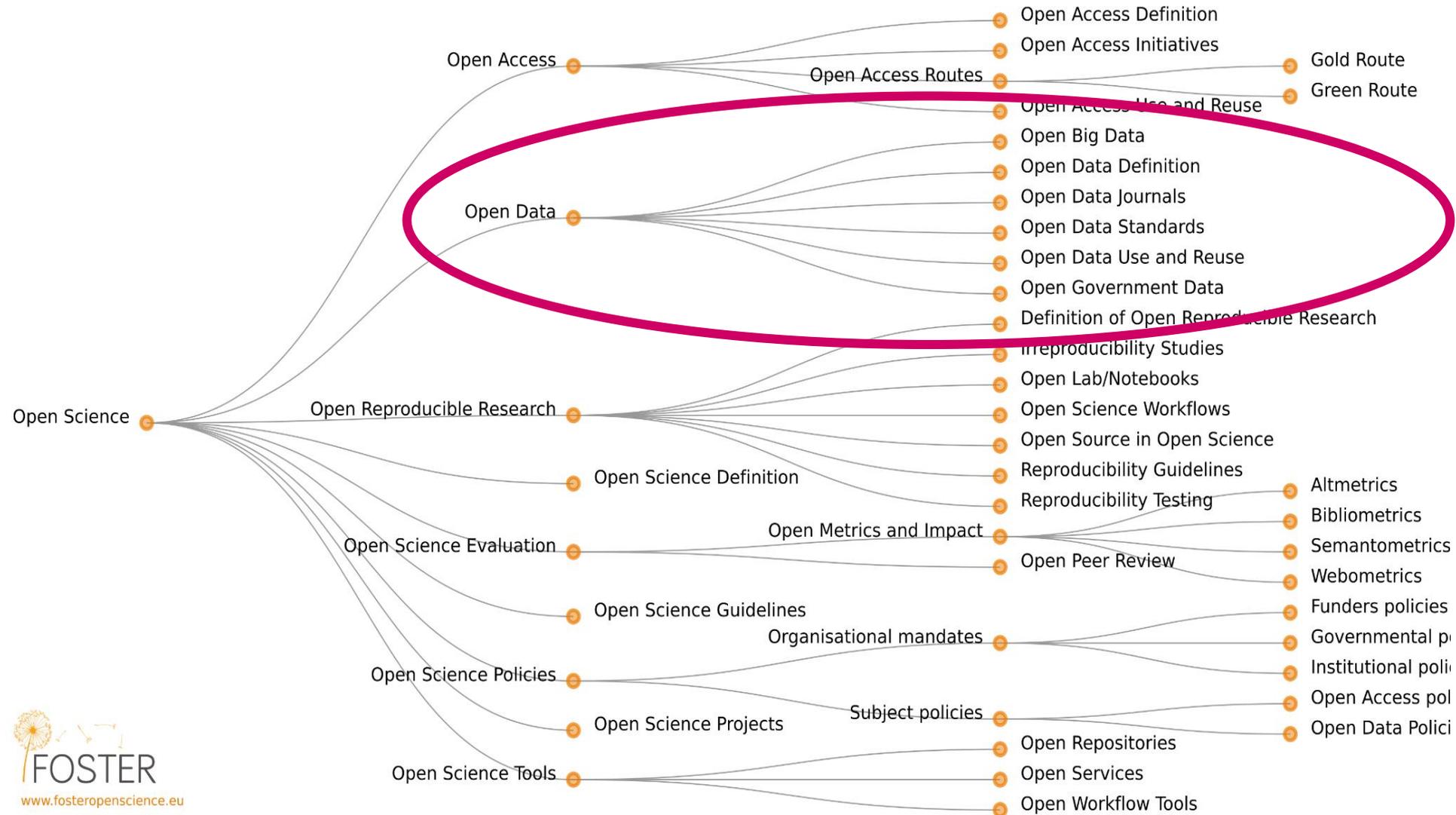
- Remplissez votre carte d'identité de *Data scientist*
  - Présentez-vous à votre voisin-e
- 
- Présentez votre voisin-e

# QU'EST-CE QUE L'OPEN SCIENCE ?

---

La **science ouverte** est le mouvement visant à rendre la **recherche scientifique** (y compris les publications, les données, les échantillons physiques et les logiciels) **accessibles** à tous les niveaux d'une société curieuse, amateur ou professionnelle. (trad. libre)

# TAXONOMIE DE L'OPEN SCIENCE



# OBJECTIFS DE L'OPEN SCIENCE

---

1. **Transparency** in experimental methodology, observation, and collection of data.
2. **Public** availability and reusability of scientific **data**.
3. **Public** accessibility and transparency of scientific **communication**.
4. Using web-based tools to facilitate scientific **collaboration**.

# LES + ET LES – DE L'OPEN SCIENCE

---

 « Reproducibility crisis »

- Beaucoup de recherches ne sont pas reproductibles

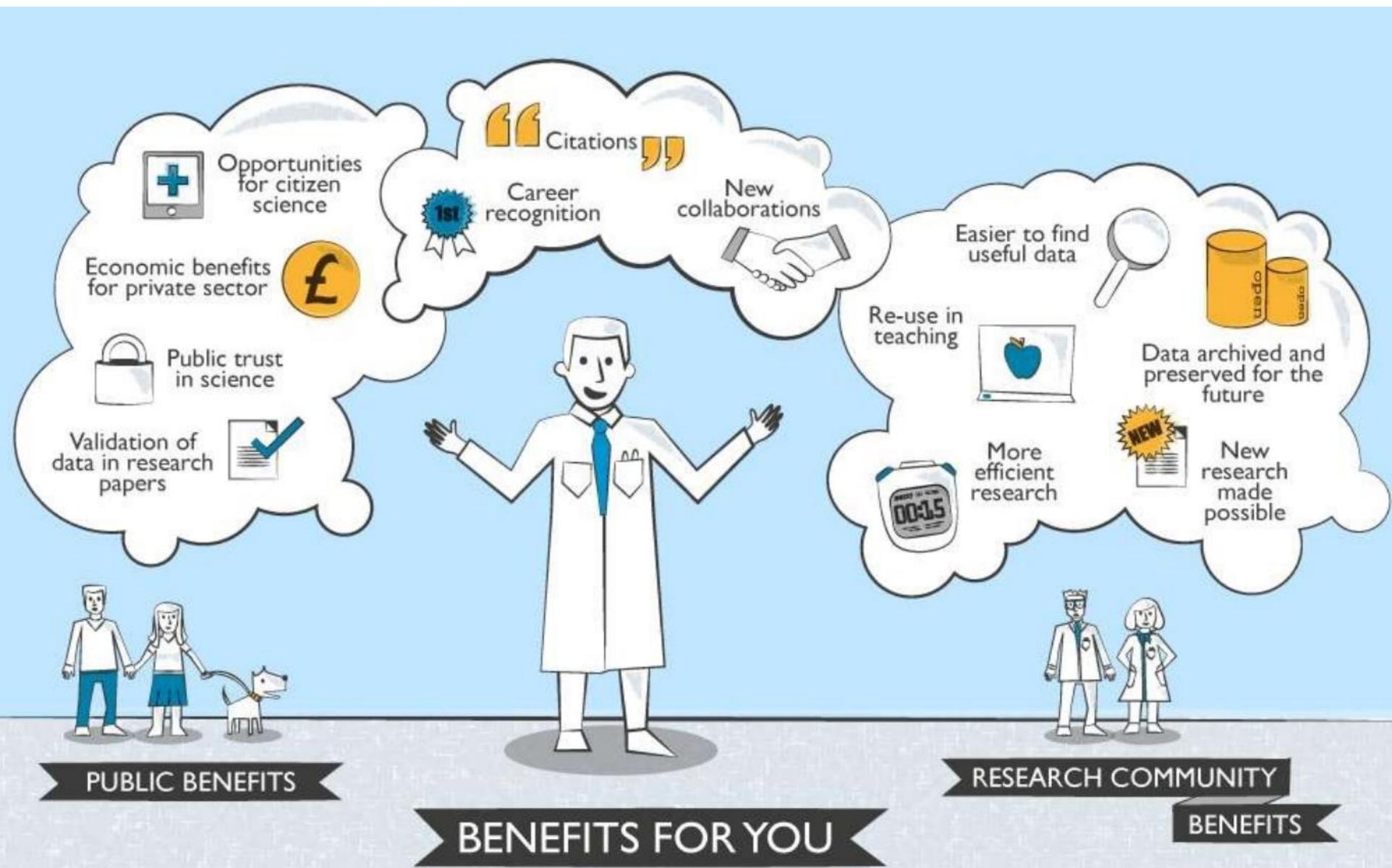
 Diffusion plus large des résultats scientifiques

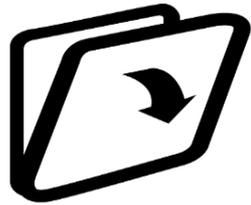
- Abonnements, licences, formatage inapproprié, formats propriétaires... freinent le partage

 Bonne utilisation de l'argent public

 En faveur des profiteurs et des prédateurs

 Bureaucratie





# **GESTION DE VOS DONNÉES DE RECHERCHE**

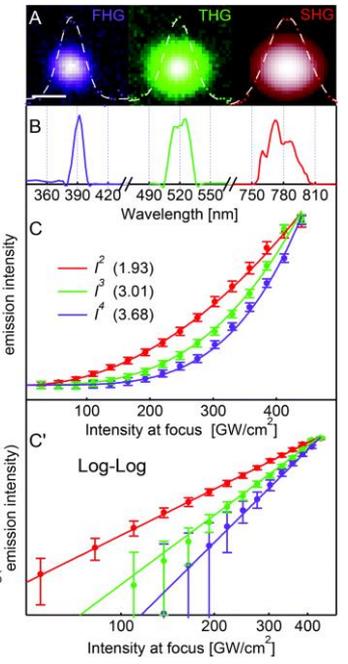
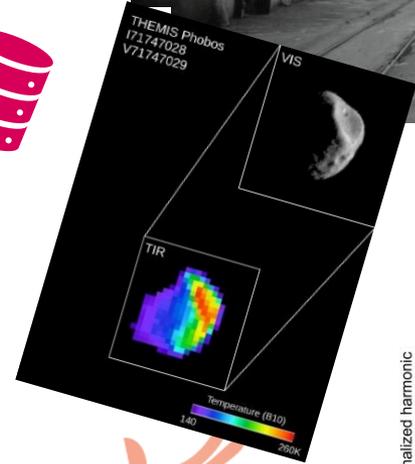
# DONNÉES DE RECHERCHE ?



**Enregistrements factuels** (chiffres, textes, images et sons), qui sont utilisés comme **sources principales** pour la recherche scientifique et sont généralement reconnus par la communauté scientifique comme **nécessaires pour valider des résultats de recherche.**

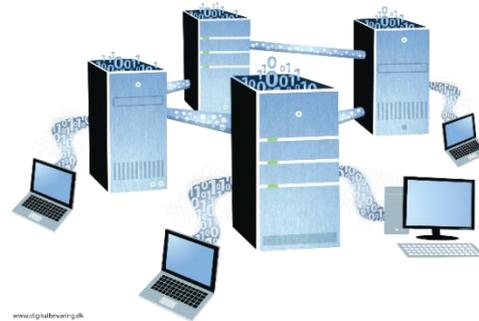
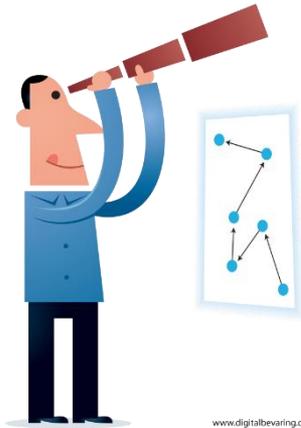


# UN CONCEPT POUR UNE PLURALITÉ DE FORMES



# CLASSIFICATION DES DR

Observationnelles



Derivées /  
compilées



Simulation /  
modèles



Experimentales

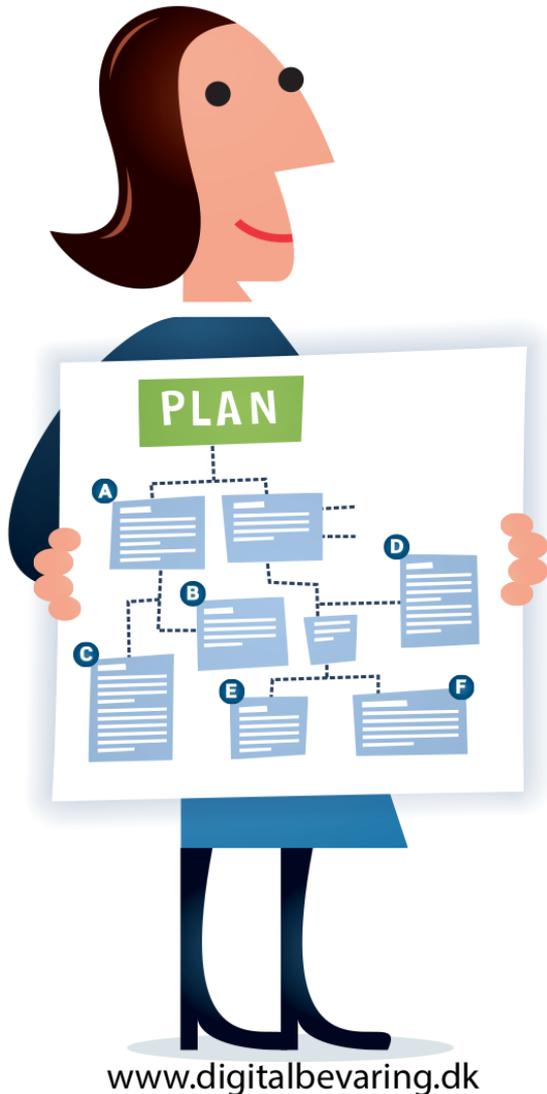
Références

# QU'EN EST-IL DE VOS DONNÉES ?



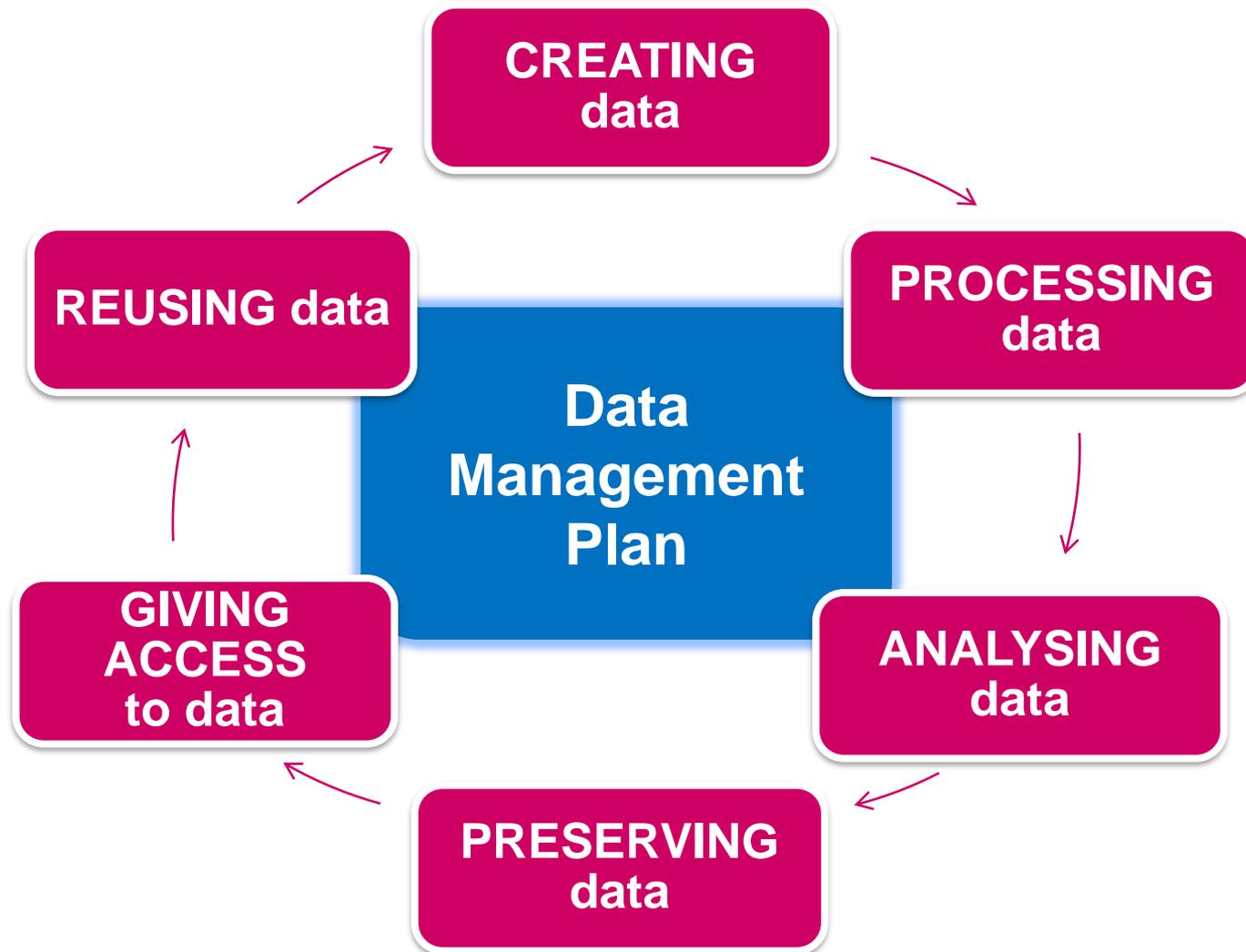
	Observationnelles	Experimentales	Référence	Dérivées/compilées	Simulation
Numérique	<ul style="list-style-type: none"> <li>Résultats des enquêtes</li> <li>Mesures de la température</li> </ul>		<ul style="list-style-type: none"> <li>Collection de document d'archive</li> </ul>		
Analogue			<ul style="list-style-type: none"> <li>Collection de document d'archive</li> </ul>		

# QUE FAIRE ?



Mettre en place de bonnes pratiques pour gérer les données de recherche

# CYCLE DE VIE DES DR ET LE DMP

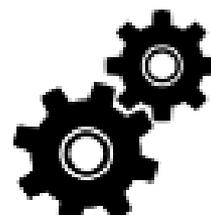


# FAIR = BONNES PRATIQUES

---

**F**indable 

**A**ccessible 

**I**nteroperable 

**R**eusable 

# EN BREF...



<https://www.youtube.com/watch?v=N2zK3sAtr-4>



*Can't find the USB  
drive. It is in a box...  
There are many boxes!  
I forgot to label the  
boxes!*

*Let's talk about :*

- *Architecture* des dossiers et leur nommage
- *Nommage des fichiers*
- *Versioning des fichiers*

*Conventions de nommage*

# YOUR COMPUTER DESKTOP

## MOST FREQUENT SHORTCUTS

INTERNET BROWSER



## SIDE PROJECTS



## TO DO LISTS



TO DO



TO REALLY DO



SUPPOSED TO HAVE DONE LAST WEEK

PIRATED MUSIC, MOVIES, COMICS (MAIN USE FOR UNIVERSITY HIGH-SPEED INTERNET CONNECTION)



PAPERS YOU'VE BEEN MEANING TO READ FOR MONTHS



DRAFT WITH ADVISOR COMMENTS ON THEM



STUFF YOU DON'T KNOW WHAT TO DO WITH BUT DON'T WANT TO DELETE BECAUSE YOU'RE OBSSIVE COMPULSIVE

INSTALL FILES FOR RANDOM PROGRAMS YOU ONLY USED ONCE AND NOW HAVE NO IDEA WHAT THEY DO.



## QUARANTINE SECTION

E-MAIL ATTACHMENTS FROM YOUR PARENTS



DRUNKEN PICTURES FROM HAPPY HOUR



## THESIS STUFF

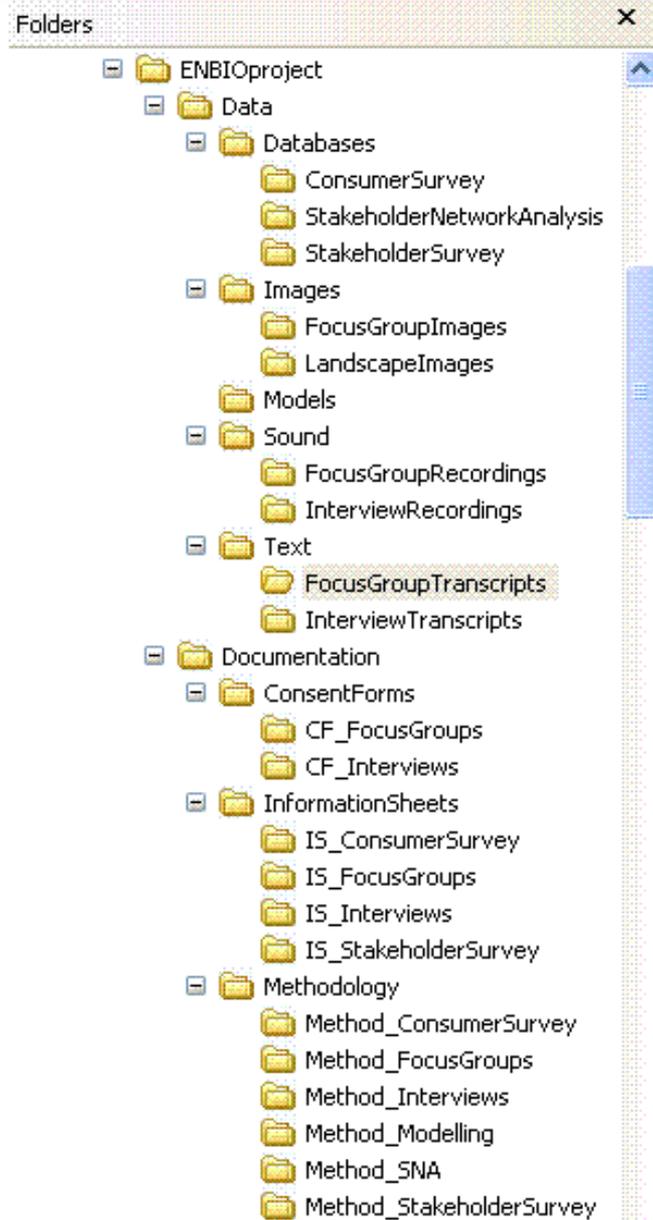


# DOSSIERS ET FICHIERS : LEUR STRUCTURE

---

**Activité: Organisez vos fichiers**

# DOSSIERS ET FICHIERS : LEUR STRUCTURE



- ✓ Évitez les chevauchements de catégories
- ✓ Ne laissez pas les dossiers devenir trop gros
- ✓ Limitez la profondeur des structures

Untitled 158.docx  
Untitled 241.doc  
Untitled 138 copy.docx  
Untitled 138 copy 2.docx  
Untitled 139.docx  
Untitled 40 MOM ADDRESS.jpg  
Untitled 242.doc  
Untitled 243.doc  
Untitled 243 IMPORTANT.doc  
Untitled 41.jpg



PRO TIP: NEVER LOOK IN SOMEONE ELSE'S DOCUMENTS FOLDER.

# NOMMAGE DES FICHIERS: EXERCICE

My passwords.doc	My data.xls
IMPORTANT.doc	My study.doc
My Thesis final final.doc	Doc.1.doc
Ma thèse version 12.doc	New doc.doc
Data 01/08/2016.xls	Int 1 (2).doc
Data 10 jan. 2016.xls	Interview 1.doc

1. En revenant à vos données dans un an, saurez-vous reconnaître le contenu de ces fichiers ?
2. Quelles informations doivent figurer dans un nom de fichier pour identifier le contenu ?
3. Que changeriez-vous dans ces noms ?



# BONNES PRATIQUES DE NOMMAGE

Best Practice	Example
Ordonner les dates en commençant par l'année pour permettre le tri par date (garder 4 chiffres pour l'année)	YYYY-MM-DD or YYYYMMDD
Limiter le nom du fichier à <b>32 caractères</b>	32CharactersLooksExactlyLikeThis.csv
Pour la numérotation séquentielle, utilisez des <b>zéros de tête</b> pour permettre des versions à plusieurs chiffres. Pour une séquence de 1 à 10 : 01-10 Pour une séquence de 1-100 : 001-010-100	<b>NON</b> : ProjID_2.csv ProjID_12.csv <b>OUI</b> : ProjID_02.csv ProjID_12.csv
N'utilisez pas de <b>caractères spéciaux</b> ou <b>d'espaces</b> & , * % # ; * ( ) ! @\$ ^ ~ ' { } [ ] ? < > -	<b>NON</b> : name&date@location.doc <b>OUI</b> : name_date_location.doc
N'utiliser <b>qu'un seul point</b> (pour l'extension du fichier)	<b>NON</b> : name.date.doc <b>OUI</b> : name_date.doc
Évitez d'utiliser des noms de fichiers génériques qui peuvent être ambigus lorsqu'ils sont déplacés.	<b>NON</b> : MyData.csv <b>OUI</b> : ProjID_date.csv

# "FINAL".doc



FINAL.doc!



FINAL\_rev.2.doc



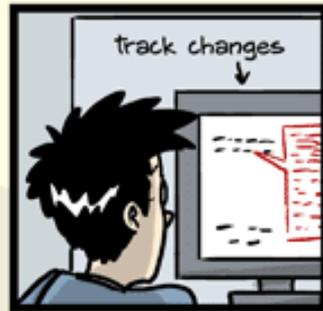
FINAL\_rev.6.COMMENTS.doc



FINAL\_rev.8.comments5.  
CORRECTIONS.doc



JORGE CHAM © 2012



FINAL\_rev.18.comments7.  
corrections9.MORE.30.doc



FINAL\_rev.22.comments49.  
corrections.10.#@\$%WHYDID  
ICOMETOGRADSCHOOL????.doc



# FILE VERSIONING: BEST PRACTICES

---

1. Évitez les étiquettes "finales" imprécises
2. Enregistrez les nouvelles versions en utilisant une convention cohérente :
  - Changements majeurs: v1; v2...
  - Changements minor: v1\_1; v1\_2
3. Documentez votre convention
4. Tenez compte de vos besoins en matière de contrôle des versions

# EXEMPLE DE CONVENTION DE NOMMAGE

## TILS Document Naming Convention

Document naming for the TILS Division should follow this convention:

GDL\_TILSDocNaming\_V1\_20090612.docx

A prefix shows the document type

The document title describes the content

The version number

The date in the format yyyymmdd

File names created from the TILS document naming convention are made up of four parts joined together with an underscore character ( \_ ). There should **not be any spaces** in the file name.

# PRINCIPLES CLÉS D'ORGANISATION

---

1. En consacrant un peu de temps **au départ**, vous gagnerez beaucoup de temps par la suite.
2. Soyez **réaliste** : trouvez un équilibre entre trop et pas assez d'efforts.
3. Il n'y a pas une seule bonne façon de procéder ; établissez un **système qui vous convient**.
4. Réfléchissez aux personnes **pour lesquelles votre système doit fonctionner** : Vous seul ? Vous et votre groupe de laboratoire ? Des collaborateurs ?

*“Make a system. Share the system. Follow the system.”*

([MIT libraries](#))



*I opened the data  
and I could not  
understand it !*



**Everything you  
need to know  
about the data is  
in the article**

*Let's talk about :*

- *Metadonnées & documentation: Documenter le format, outils, champs, etc.*
- *Documentez vos variables, toutes les abréviations, etc. dans un fichier **readme***

# METADONNÉES

Une recherche doit être documentée à différents niveaux :

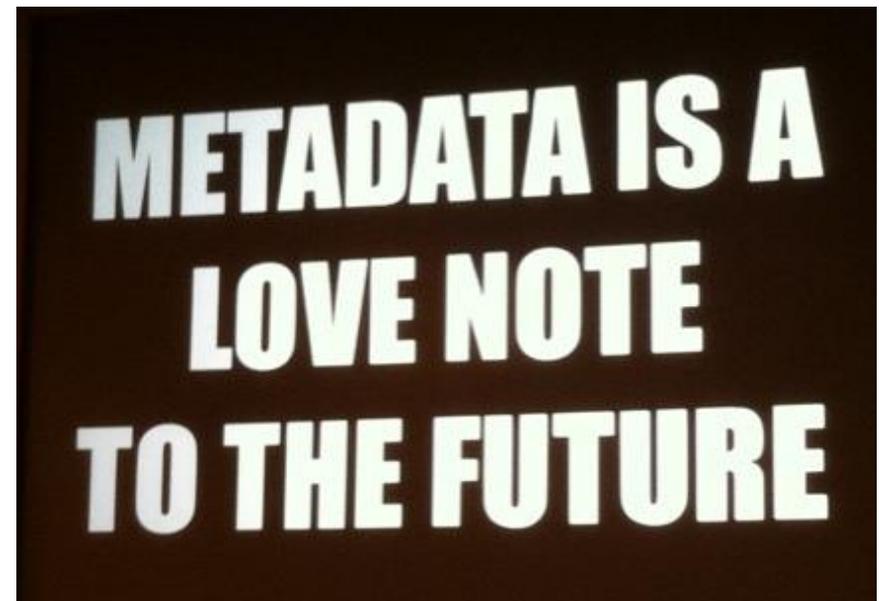
- Niveau du projet
- Niveau du fichier ou de la base de données
- Niveau de la variable ou de l'élément

L'objectif est de la rendre

- Trouvable
- Compréhensible
- Réutilisable

Cf. FAIR data 😊

→ Eviter l'ambiguïté et les interprétations erronées



# 3 TYPES PRINCIPAUX DE METADONNÉES

---

- Métadonnées **descriptives** permettent la découverte, l'identification et la sélection des ressources
  - Title/unique ID, auteur, résumé, mots-clés, financeur, date, ...
- Métadonnées **administratives** facilitent la gestion des ressources
  - Rights statement, date, format de fichiers, version, préservation, ...
- Métadonnées **structurelles** décrivent comment les relations entre différents composants d'un ensemble de données
  - Dictionnaire de données, liste de variables, schéma de base de données, taxonomie, abréviations, ...



NEWS • 25 OCTOBER 2019

# Venice 'time machine' project suspended amid data row

Disagreements among international partners leave plans to digitize the Italian city's history in limbo.

Davide Castelvecchi



Historians want to use archive documents to create a virtual time machine for Venice, pictured here in the 18th century. Credit: DEA/Getty

PDF version

## RELATED ARTICLES

**The 'time machine' reconstructing ancient Venice's social networks**

**Saving Venice**

## SUBJECTS

Databases

History

# IMPORTANCE DES MÉTADONNÉES

---

◆ **2012** – Le projet est officiellement **lancé** :

Venice's State Archive + Ca' Foscari Univ. + EPFL (DHLAB)

**2014** – Un accord non contraignant est signé. Mais ... il ne précise pas les licences qui régiront l'utilisation des données numérisées par les chercheurs.

**2017** – Enjeu: 1000 ans de documents sous forme numérique dynamique : des scanners spéciaux à grande vitesse, des milliers d'images HD par heure.

**2019** – La numérisation de ~190,000 documents (**8 TB**) n'aurait pas suivi **une politique commune en matière de métadonnées** : [archival-science guidelines](#) (qui exige par exemple l'enregistrement de la provenance de chaque document)

▼ **Maintenant** – ... **la collecte des données a été interrompue, en raison des doutes sur l'utilité des données déjà recueillies !**

# README.TXT

---

Fichier de métadonnées accompagnant l'ensemble de données et fournissant toutes les informations nécessaires à son sujet.

- Pour **vous**, afin de vous tenir à vos standards
- Pour **vous**, pour comprendre vos noms de fichiers dans 15 ans (ou 15 jours !)
- Pour les collègues, pour enregistrer correctement les fichiers dans vos dossiers
- Pour d'autres personnes, pour utiliser vos fichiers

# README.TXT : BONNES PRATIQUES

---

- Placez ce fichier à **la racine** de l'ensemble de données
- **Son titre** doit indiquer qu'il faut le lire (en premier)
- Incluez **tout ce que vous jugez utile de savoir** pour les utilisateurs potentiels de vos données
  - Métadonnées descriptives, administratives et structurelles
- Rédigez votre document sous forme d'un **fichier texte brut** (ou PDF/A)
- Utilisez un modèle pour vous assurer que vous n'avez rien omis d'important
- → Des fichiers supplémentaires peuvent être créés dans les sous-dossiers si nécessaire
  - Formatez plusieurs fichiers readme de manière identique

# DÉCRIRE LES VARIABLES

Dans le Codebook ou dans le fichier readme

- Inclure des références aux normes
- Incluant les unités de mesure

```
78      SP      Species
79      REF      Reference number - This is my maintenance code for updates.
80      NEWREF   New Reference number - These are the refs numbers in the
81                files REFS and SUMMARY.
82      FAMLAB   Family Label - An 8-character label for family.
83      GENLAB   Genus Label - An 8-character label for genus.
84      SPLAB   Species Label - An 8-character label for species.
85      COD      Species code - A 5-character code for the species.
86      DISPCAT  Disperser type category - BIRDS, MIXED, MAMMALS.
87      DISP     Disperser type - Finer categorization. Not yet completed.
88                Needs revision.
89      MEGAFUNA Whether the fruit species is associated with dispersal
90                of seeds by megafauna.
91      AREA     Geographic area - Major geographic areas of the data
92                sources.
93                MEurope: Mediterranean Europe (also includes Israel and
94                Morocco.
95                NEurope: Temperate and Northern Europe.
96                NAmerica: North America, excl. Southern Mexico.
97                NTAmerica: Neotropical America, incl. Southern Mexico.
98                Africa: Africa, south of Sahara.
99                Australasia: Australia, Malaysia, New Guinea, and whole
100               SouthEast Asia.
101      HABIT    Growth habit - Tree, Shrub, Liana, Herb.
102      FRUIT    Fruit type - Type of fruit. Not completed yet.
103                Needs revision.
104      COLOR    Fruit color -
105                Black, Blue, Red, Orange, Yellow, Green, Brown, White
106      COMPCOLOR Whether the fruit is monocolored, bicolored, or multicolored.
```

<https://doi.org/10.5061/dryad.9tb73/1>

Fichier Edition Format Affichage ?

This DATSETNAMEreadme.txt file was generated on YYYYMMDD b

-----  
GENERAL INFORMATION  
-----

Title of Dataset:

Author Information (Name, Institution, Address, Email)

Principal Investigator:  
Associate or Co-investigator:  
Alternate Contact(s):

Date of data collection (single date, range, approximate d

Geographic location of data collection: <City, State, Coun

Information about funding sources or sponsorship that supp

-----  
SHARING/ACCESS INFORMATION  
-----

Licenses/restrictions placed on the data, or limitations o

Recommended citation for the data:

Citation for and links to publications that cite or use th

Links to other publicly accessible locations of the data:

Links/relationships to ancillary or related data sets:

-----  
DATA & FILE OVERVIEW  
-----

File list (filenames, directory structure (for zipped fil

Relationship between files, if important for context:

-----  
DATA & FILE OVERVIEW  
-----

File list (filenames, directory structure (for zipped files) and brief desc

Relationship between files, if important for context:

Additional related data collected that was not included in the current data

If data was derived from another source, list source:

If there are there multiple versions of the dataset, list the file updated,

-----  
METHODOLOGICAL INFORMATION  
-----

Description of methods used for collection/generation of data: <Include lin  
or other documentation containing experimental design or protocols used in

Methods for processing the data: <describe how the submitted data were gene  
data>

Software- or Instrument-specific information needed to interpret the data,  
version numbers:

Standards and calibration information, if appropriate:

Environmental/experimental conditions:

Describe any quality-assurance procedures performed on the data:

People involved with sample collection, processing, analysis and/or submis:

-----  
DATA-SPECIFIC INFORMATION <Create sections for each datafile or set, as app  
-----

Number of variables:

<https://data.research.cornell.edu/content/>

# STRUCTUREZ VOS DONNÉES

Utilisez **les normes de votre discipline** (si elles existent !) pour faciliter l'interopérabilité.

Chaque norme est clairement définie et documentée



Chaque ensemble de données utilise les mêmes termes / est organisé de manière cohérente



Partage facile des données entre les institutions, les applications et les disciplines

Pour trouver des standards:

<https://fairsharing.org/standards/>

**FAIR**sharing.org  
standards, databases, policies

*But surely, you saved your data ?*

*- I did, I saved it on a USB drive*

*Let's talk about :*

- *Stockage des données*
  - *Back-ups*
  - *Sécurité*

*... I will need that back when you are finished, this USB is my only copy of my data*

# CASH REWARD

for returning my lost backpack



365 Adventure.com

- Black [AK] Burton Rucksack
- Lost on Friday 15. July at 8 pm in the Panton Arms pub 43, Panton St. Cambridge
- Containing a laptop (white MacBook), a black external hard drive and scientific research documents

The external hard drive is VERY important to me as it contains 5 years of research data which are crucial for my PhD thesis!!!

If you found it, I would be extremely grateful if you could return it to the Panton Arms or contact me on:  
(ar456@cam.ac.uk)

Thank you!!

# PERDU !

## POCHETTE NOIRE AVEC DISQUE DUR

Ce lundi 07 octobre matin, j'ai perdu mon disque dur professionnel aux alentours de la rue de carouge. Il est extrêmement important pour moi.

Si vous l'avez trouvé, merci de me contacter au :

**Vous serez récompensé/e financièrement.**

Merci infiniment !!!

**RECOMPENSE  
FINANCIÈRE !!!**



environ 10 x 15cm

Credit: Peter Murray-Rust,

<http://blogs.ch.cam.ac.uk/pmr/2011/08/01/why-you-need-a-data-management-plan/>, August 2011, CC-BY

# STORAGE & BACK-UP VS ARCHIVE

---

## **Stockage & Backup**

- Récupérable
- Copie
- Modification
- Court terme

## **Archive**

- Accessible
- Déplacée
- Inactive
- Long terme

# AVOIR UN PLAN DE SAUVEGARDE SYSTÉMATIQUE

---

- Pendant la **collecte**, les données doivent être stockées sur des **ordinateurs locaux ou des serveurs institutionnels**.
- Pendant **l'analyse**, une copie brute des données doit être conservée et les données doivent être sauvegardées à **d'autres emplacements** :
  - **Idéalement: 2 copies sur-site + 1 hors-site**
- **Ensuite**, les données peuvent être téléchargées dans un **repository** pour une **préservation à long terme** (et un partage, si possible).

# BONNES PRATIQUES POUR LE STOCKAGE

→ Stocker **3 copies** de vos données:



1. L'original
2. Une copie conservée sur un dispositif externe local
3. Une copie conservée sur un dispositif externe à un autre endroit.

# NAS DE L'UNIGE

- Network-attached Storage connecté aux réseaux d'ordinateurs
- Dédié aux données de recherche
- Snapshots toutes les 4 hours (données récupérables jusqu'à 6 semaines)
- Sauvegarde hors site(Campus Biotech)
- Gratuit jusqu'à 50 GB
- CHF 75.- / TB / année



# CLOUD - ONEDRIVE



UNIVERSITÉ  
DE GENÈVE

Intranet STIC

Espace utilisateurs-trices

Vous êtes ici : [Accueil](#) / [Espace utilisateurs-trices](#) / [Logiciels](#) / OneDrive

## Comment installer et démarrer avec OneDrive sur son poste de travail

### Contexte et prérequis

La solution OneDrive Entreprise offre à tous les étudiant-es et collaborateurs/trices de l'UNIGE disposant d'un compte ISIs un espace cloud de stockage d'1 teraoctet (To). Cet espace est synchronisé avec une copie locale de vos fichiers sur votre poste de travail, facilitant ainsi leur utilisation, leur mise à jour et leur partage, particulièrement dans les contextes d'accès à distance et de télétravail.



### Environnement

- Solution : OneDrive (Microsoft)
- OS : Windows\*, macOS, Android, iOS (iPhone/iPad),  
\*OneDrive est déjà installé sous Windows 10.

### Procédure

1. Accéder au site Microsoft et télécharger OneDrive
2. Installer OneDrive avec ses identifiants UNIGE (ISIs)
3. Pour les premiers pas, consulter la documentation Microsoft

Free up to 1 TB

<https://plone.unige.ch/distic/pub/logiciels/onedrive/comment-installer-onedrive>

# SWITCHdrive

Store, synchronise, share and edit files quickly and reliably - all in the secure SWITCH cloud.

## Facts at a glance

- Standard 50GB online storage per user
- Easy collaboration and protected access thanks [AAI-Login](#)
- Synchronize files & folders across multiple devices
- Access via desktop client, browser or mobile application



## Your data at SWITCH

- All data is stored on SWITCH servers in Switzerland.
- Full compliance with Swiss data protection regulations
- No data and metadata exchange with other Office companies - not even for diagnoses or future extensions.

## New collaboration functions thanks to OnlyOffice integration

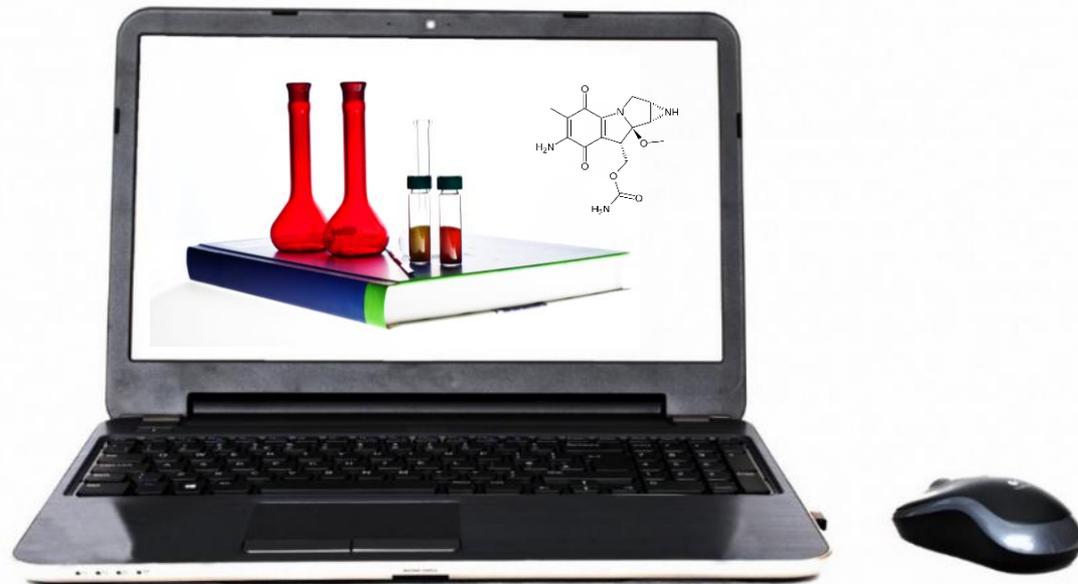
- Real-time collaboration on shared documents - work with others on a document and track all changes live
- No local software installation required
- Access via desktop client, browser or mobile application

<https://drive.switch.ch/index.php/login>



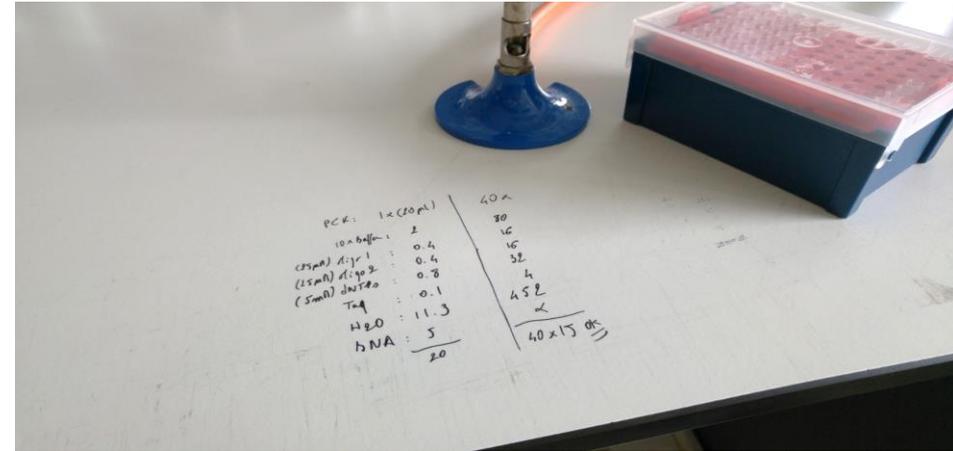
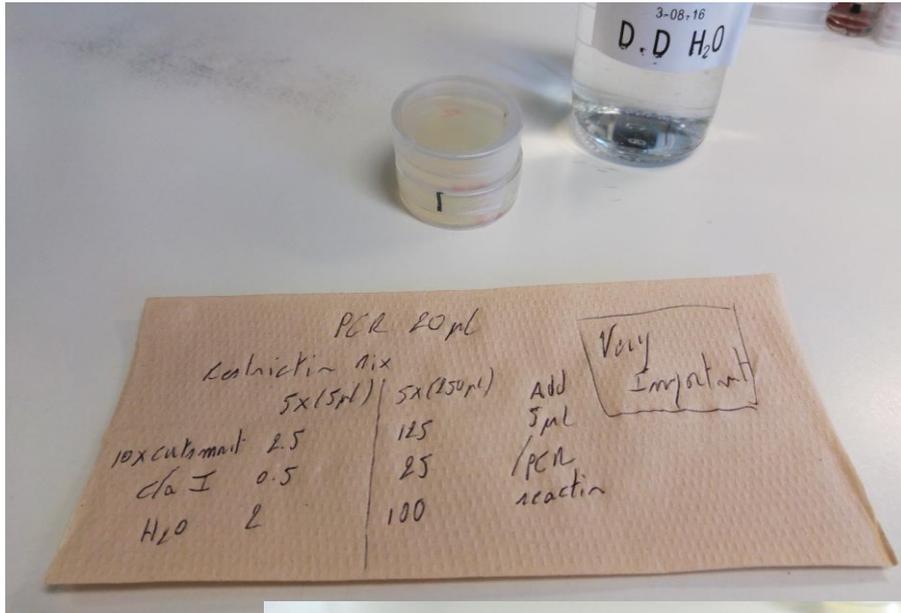
# ELECTRONIC LAB NOTEBOOKS

---



Un autre **outil** pour sauvegarder & partager des données et des informations au sein de votre laboratoire

# ALTERNATIVES AUX LAB NOTEBOOKS...



Vincent Gaggioli

# ENVOYER DES FICHIERS

SWITCH

Services ▾

Stories ▾

About us ▾

Services → Send files

## SWITCHfilesender: send files via the SWITCH cloud

Some files are just too big to send in an e-mail. The best way round this is SWITCHfilesender. The service runs in the protected SWITCH cloud, and it's free.

SWITCHfilesender is the best way to send large files up to 50 GB. The web-based service couldn't be simpler:

1. Start SWITCHfilesender in your browser.
2. Log in (using your AAI or voucher login).
3. Upload your file to the protected SWITCH cloud.
4. Enter the recipient's e-mail address.
5. Send the file.

The recipient then gets an e-mail containing the download link.

All university members can send external users a voucher for a single use of SWITCHfilesender. The voucher is valid for a maximum of 20 days and permits the holder to send one file.

Unlike comparable web-based services, SWITCHfilesender runs in the protected SWITCH cloud. Files are stored exclusively at SWITCH's data centre in Switzerland and deleted after 20 days at the very latest. If you want to store files in the SWITCH Cloud with no time limit, [SWITCHdrive](#) is the service for you.

# NOTE SUR LES DONNÉES PERSONNELLES

- **Données personnelles:** “toutes les informations qui se rapportent à une personne identifiée ou identifiable;”
- **Données sensibles:** “les données personnelles sur:
  1. les opinions ou activités religieuses, philosophiques, politiques ou syndicales,
  2. la santé, la sphère intime ou l'appartenance à une race,
  3. des mesures d'aide sociale,
  4. des poursuites ou sanctions pénales et administratives; ”
- **Données confidentielles / «critiques»**

Traitement de ce type  
de données



[Cantonal Data Protection Law](#)

[Loi fédérale sur la protection des données \(LPD\)](#)

# SÉCURITÉ DES DONNÉES

Pendant la recherche, pour garantir l'intégrité de vos données, vous devez bien gérer l'accès aux données : **qui** a accès à **quoi** ?

- Accès physique
- Accès numérique



Attention et des réglementations supplémentaires lors du traitement des données personnelles ("Keep it in Switzerland").

NAS académique  
Educloud (filr)  
OneDrive  
SWITCHdrive  
SWITCHfilesender  
Locally hosted ELN

Dropbox  
Google Drive  
Wetransfer

...

*I can't read hexadecimal.*

You will need the program that created the hexadecimal file

Yes I will. What is the name of the program. Do you have a copy of the program?

I do not use this program anymore because the company that made it went bankrupt.

*Let's talk about :*

- *Conversion des données: formats ouverts pour la préservation*
- *Préserver le logiciel en même temps que les données ?*

# BACK-UP VS ARCHIVE

---

## Stockage & Backup

- Récupérable
- Copie
- Modification
- Court terme

## Archive

- Accessible
- Déplacé
- Inactif
- Long terme

# KEEP RAW DATA RAW

---

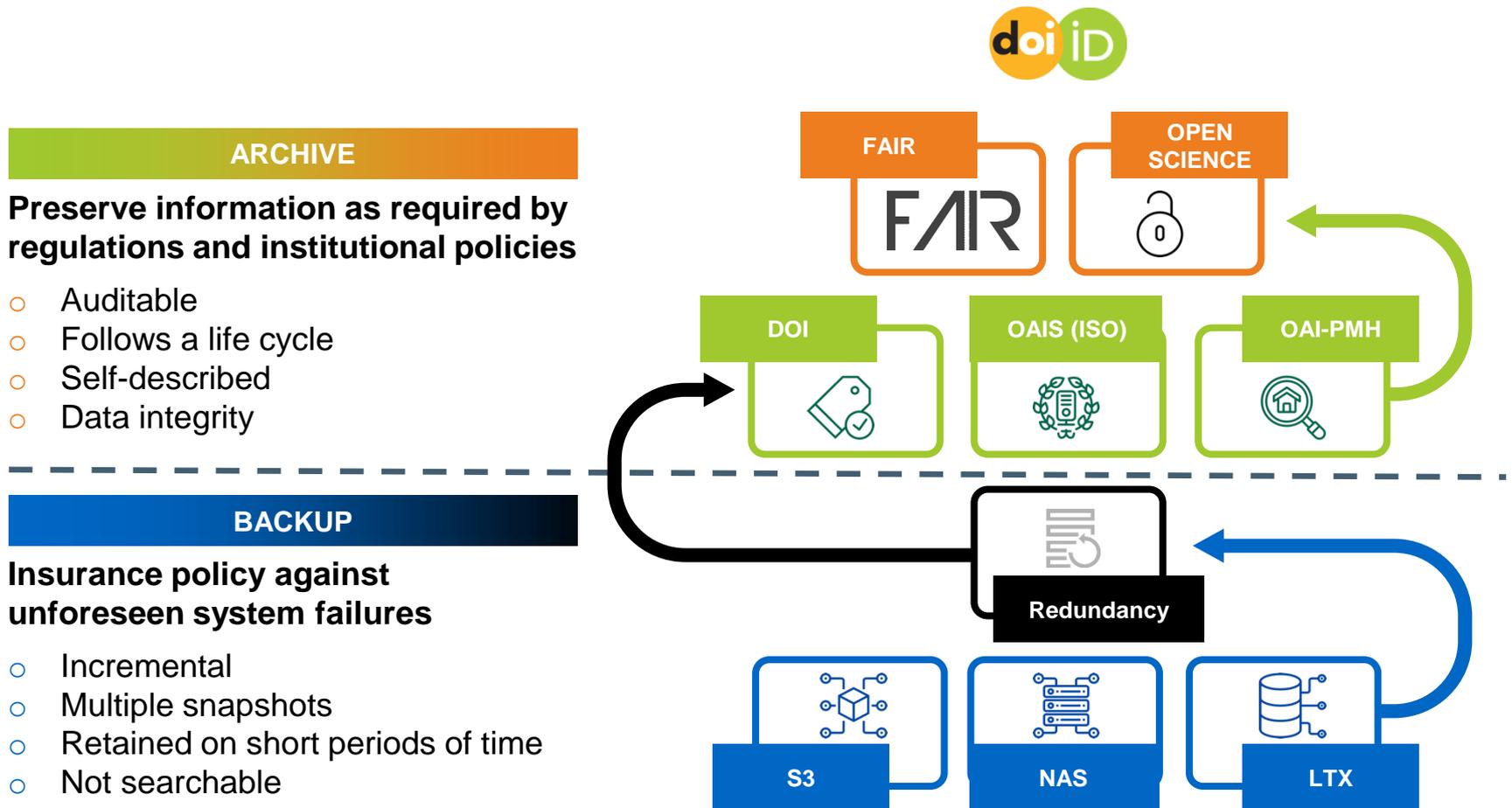
- Permet une **nouvelle analyse** avec les anciennes données et la **transparence**
- Dans la mesure du possible, utilisez un **hachage cryptographique** pour garantir que l'ensemble de données n'a pas subi de corruption et/ou de manipulation silencieuse.

# PRIVILÉGIER DES FORMATS OUVERTS

---

- Le traitement quotidien peut utiliser des formats fermés
- Mais à des fins **d'archivage**, il faut maximiser **l'accessibilité** et la **valeur à long terme** en convertissant les documents en formats ouverts ou portables qui ne sont pas dépendants d'un logiciel.
- Par exemple, préférer:
  - **.csv** plutôt que **.xlsx**
  - **.svg** plutôt que **.ai**
  - **PDF/A** plutôt que **PDF**
- Voir les guides de [ETH](#), [UK data service](#), [Harvard](#)

# THE WHOLE PICTURE



*Thank you for  
sending me a copy of  
your data on a USB  
drive*

*... I will need that  
back when you are  
finished, this USB is  
my only copy of my  
data*

*Let's talk about :*

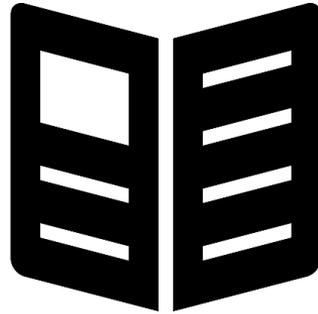
- *Archivage & partage*
  - *Mise en garde concernant les données spéciales*
- *Repositories & Licenses*

# EXIGENCES POUR LE PARTAGE DES DONNÉES

*Your work was funded by NIH and published in Science. I am requesting your data!*



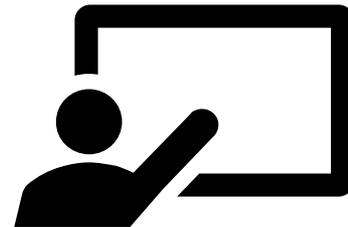
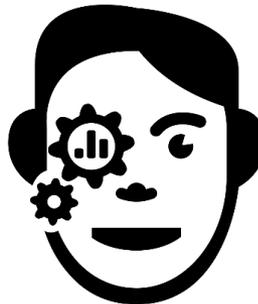
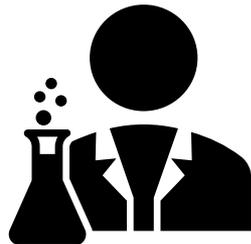
Bailleurs de fonds



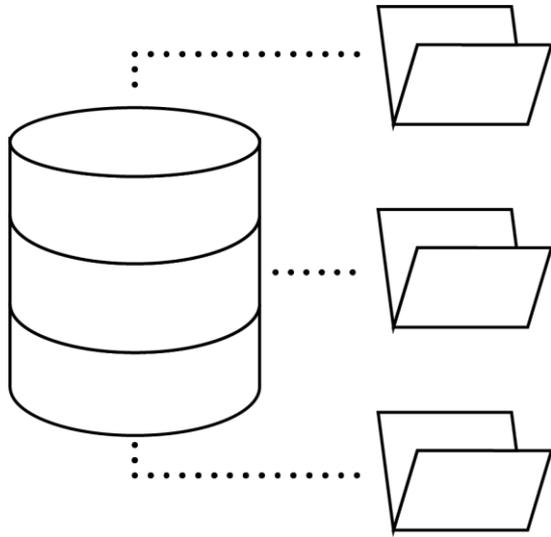
Revue



Institutions



# ARCHIVAGE ET DISSEMINATION



- ✓ Multi-disciplinaire (ex: Zenodo)
- ✓ Disciplinaire (ex: GenBank)
- ✓ Institutionnel (ex: Yareta)

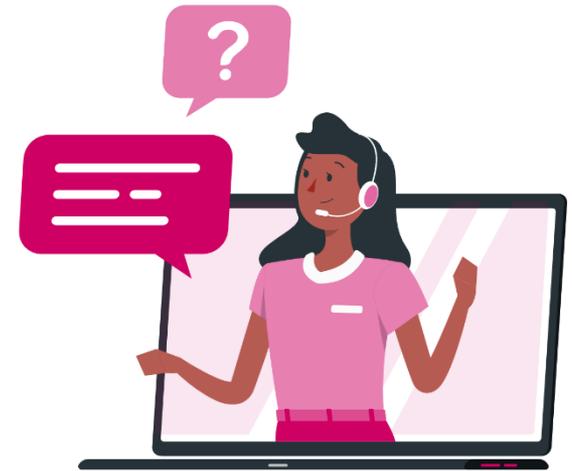


- 💣 service proposé par une revue
- 💣 serveur d'un projet ou groupe, site personnel

# TROUVER ET CHOISIR UN REPOSITORY



**Consultez les collègues  
de votre discipline**



**Demandez à vos  
bibliothécaires**



**Explorez Re3data**

# UN ANNUAIRE DE REPOSITORIES

Search

Browse ▾

Suggest

Resources ▾

Browse by subject

Browse by content type

Browse by country

re3data.org

REGISTRY OF RESEARCH DATA REPOSITORIES

Search...



Search

# NOTE SUR LES DONNÉES PERSONNELLES

- **Données personnelles:** “toutes les informations qui se rapportent à une personne identifiée ou identifiable;”
- **Données sensibles:** “les données personnelles sur:
  1. les opinions ou activités religieuses, philosophiques, politiques ou syndicales,
  2. la santé, la sphère intime ou l'appartenance à une race,
  3. des mesures d'aide sociale,
  4. des poursuites ou sanctions pénales et administratives; ”

Traitement de ce type  
de données



[Cantonal Data Protection Law](#)  
[Loi fédérale sur la protection des données \(LPD\)](#)

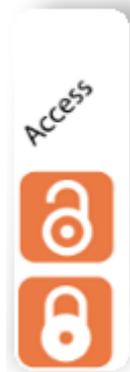
# MON REPOSITORY EST-IL FAIR ?

re3data.org  
REGISTRY OF RESEARCH DATA REPOSITORIES

**F**indable

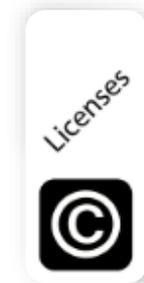


**A**ccessible



**I**nteroperable

**R**eusable



Au moins une icône  
bleue est présente

Présence d'une icône  
cadenas orange

Présence d'une icône  
licence

# CREATIVE COMMONS LICENCES

Most open



CC0

Cette licence est l'option "pas de droit d'auteur" des Creative Commons - l'auteur a renoncé à tous les droits d'auteur et autres droits similaires détenus sur une œuvre et ainsi a placé son œuvre dans le domaine public.



BY

Cette licence autorise à partager, copier, reproduire, distribuer, communiquer, traduire, réutiliser, adapter l'œuvre, même à des fins commerciales, pourvu que la création originale est créditée.



BY



SA

Avec cette licence, l'utilisateur peut partager, copier, reproduire, distribuer, communiquer, traduire, réutiliser, adapter l'œuvre, même à des fins commerciales, pourvu que la création originale est créditée.



BY



ND

Cette licence autorise à partager, copier, reproduire, distribuer, communiquer, traduire, réutiliser, adapter l'œuvre, même à des fins commerciales, pourvu que la création originale est créditée.



BY



NC

L'utilisateur peut ainsi : partager, copier, reproduire, distribuer, communiquer, traduire, réutiliser, adapter l'œuvre, mais pas dans un but commercial. La nouvelle œuvre créée peut néanmoins être diffusée elle-même sous n'importe quelle licence.



BY



NC



SA

Cette licence vous permet de remixer, de modifier et de vous baser sur l'œuvre originale à des fins non commerciales, à condition de citer l'œuvre originale et d'accorder la même licence à vos nouvelles créations.



BY



NC



ND

Cette licence est la plus restrictive. Elle permet uniquement de télécharger et de partager l'œuvre originale avec d'autres, tant que l'œuvre originale est mentionnée. L'œuvre originale ne peut pas être modifiée quelque manière que ce soit ni utilisée à des fins commerciales.

**! Certains repositories impose l'utilisation d'une certaine licence ! (eg. Dryad = CC0)**

Least open

# UN REPOSITORY LOCAL:



- ✓ Repository pour les chercheurs/euses genevois-es
- ✓ Serveur (original et copy) **en Suisse**
- ✓ Satisfait aux principes **FAIR** & attribue des **DOI**
- ✓ Solution de preservation long-terme  
(vous définissez la durée de conservation des données)
- ✓ **Gratuit** jusqu'à 50 Go  
(au dessus: 100 CHF / Tb / années de conservation)

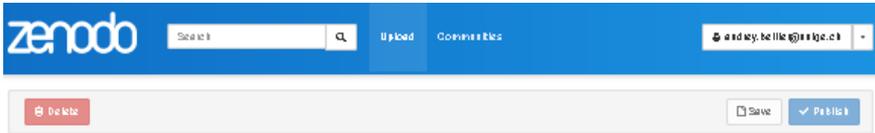
<https://yareta.unige.ch>

# UN AUTRE REPOSITORY PUBLIC:



- ✓ Catch-all repository
- ✓ g er  par le CERN
- ✓ Acc s et d p t libres et gratuits

<https://www.zenodo.org/>



# ZENODO SANDBOX VERSION

<https://sandbox.zenodo.org/>

Si vous voulez faire un  
dépôt fictif

# GESTION DES DONNÉES EN BREF

---

- ✓ Vérifier les exigences **légales** et **réglementaires**
- ✓ **Anticipez** vos besoins en matière de données (volume, sensibilité, formats,...) et sélectionnez les options appropriées en conséquence
- ✓ Veillez à ce que vos données puissent être **réutilisées** (formats ouverts, dénomination correcte, ajout de métadonnées, etc.)
- ✓ Solutions de UNIGE : **NAS, SWITCH, Yareta**

# DATA MANAGEMENT PLAN (DMP)



[www.digitalbevaring.dk](http://www.digitalbevaring.dk)

# EXIGENCES ÉLARGIES



Department  
for International  
Development



European Research Council  
Established by the European Commission

MELINDA  
GATES foundation



European  
Commission



GenomeCanada



National Science Foundation  
WHERE DISCOVERIES BEGIN



ACADEMY OF FINLAND



Vetenskapsrådet



National Institutes of Health  
*Turning Discovery Into Health*



Deutsche  
Forschungsgemeinschaft



National Institute for  
Health Research



For research on  
diseases of poverty  
UNICEF • UNDP • World Bank • WHO

# EXIGENCES DU FNS

## 2. Application data



#	No./Title	
2.1	Basic data I	In preparation
2.2	Basic data II	In preparation
2.3	Use-inspired project	In preparation
2.4	Re-submission	
2.5	Link to other SNSF projects	
2.6	Available or requested funds	
2.7	University or research institution	
2.8	Collaboration (national and international)	
2.9	Information on additional funds	
2.10	Requested funding	
2.11	Data management plan (DMP)	
2.12	Research requiring authorisation or notification	
2.13	Children you are obliged to support	
2.14	Diplomas / certificates / extension eligibility period	
2.15	General remarks on the project	

### .. Données relatives à la requête

#	No./titre
2.1	Données de base I
2.2	Données de base II
2.3	Projet orienté vers l'application
2.4	Resoumission
2.5	Relation avec d'autres projets FNS
2.6	Moyens disponibles ou requis
2.7	Haute école
2.8	Collaboration (nationale et internationale)
2.9	Moyens additionnels
2.10	Besoin financier
2.11	Plan de gestion des données (DMP)
2.12	Recherche exigeant des autorisations ou des annonces

# LE DMP SUR MYSNF

I do not submit a DMP for the following reason:

## 1. Data collection and documentation

1.1 What data will you collect, observe, generate or reuse?

1.2 How will the data be collected, observed or generated?

1.3 What documentation and metadata will you provide with the data?

## 2. Ethics, legal and security issues

2.1 How will ethical issues be addressed and handled?

2.2 How will data access and security be managed?

2.3 How will you handle copyright and Intellectual Property Rights issues?

## 3. Data storage and preservation

3.1 How will your data be stored and backed-up during the research?

3.2 What is your data preservation plan?

## 4. Data sharing and reuse

4.1 How and where will the data be shared?

4.2 Are there any necessary limitations to protect sensitive data?

4.3 All digital repositories I will choose are conform to the FAIR Data Principles.

4.4 I will choose digital repositories maintained by a non-profit organisation.

# 1. Collecte et dcumentation des données

---

▾  1.1 Quelles données allez-vous collecter, étudier, générer ou réutiliser ?

→ *Type(s)*

→ *Format(s)*

→ *Volume*

## ▾ □ 1.1 Quelles données allez-vous collecter, étudier, générer ou réutiliser ?

Les données existantes qui seront réutilisées dans le cadre du projet incluent des **transcriptions de textes**, des **images en fac-similé**, et des **fichiers audio** qui font partie de la base de données Australian Drama Archive : <https://www.austlit.edu.au/australiandramaarchive>

Le projet lui-même génèrera les types de données suivants :

- Des **transcriptions de textes de pièces de théâtre** plus anciennes que celles actuellement dans la base Australian Drama Archive. Celles-ci seront produites au **format XML ouvert .docx** en cours de projet.
- Des **images**. Quand pertinent, des **photographies** seront faites pour rendre compte de certaines caractéristiques des textes originaux ou publications éphémères identifiés au cours des recherches menées dans divers services d'archives. Environ **500 photographies seront réalisées** en haute résolution (2000 x 3000 pixels) au **format .jpg**
- **Tableaux et données chiffrées** au **format .xlsx**, relevant diverses caractéristiques des monologues trouvées dans les pièces de théâtre plus anciennes.

On estime à ce stade que le volume total de données générées n'excédera pas **20 GB** mais ce chiffre sera affiné en fin de phase de recherches dans les divers services d'archives, en fin de première année.

# 1. Data collection and Documentation

---

▾ □ 1.2 Comment les données seront-elles recueillies, étudiées ou générées ?

## **Processus de traitement:**

→ *Standards*

→ *Méthode*

→ *Versions et noms des fichiers*



## 1.2 Comment les données seront-elles recueillies, étudiées ou générées ?

Des cerveaux larvaires écrasés et colorés au Giemsa seront utilisés pour la microscopie à lumière transmise et des cerveaux larvaires entiers immunocolorés seront utilisés pour la microscopie confocale.

Tous les échantillons sur lesquels des données sont collectées seront préparés conformément aux protocoles standard publiés dans le domaine. Tous les microscopes utilisés pour l'examen des échantillons sont entretenus et recalibrés régulièrement. Toutes les lignées de drosophiles utilisées dans les expériences sont examinées périodiquement pour détecter les marqueurs phénotypiques. Les drosophiles sont maintenues en culture vivante selon les méthodes standards du domaine.

Les fichiers seront nommés selon une convention convenue au préalable :

pour les images issues de la microscopie à lumière transmise (TLM) et de la microscopie confocale (COM) et pour les quantifications (QUA), on ajoutera la date (YYMMDD), les initiales du chercheur (XY) et le numéro de version V1, V2... ) :

TLM\_YYMMDD\_TLM\_XY\_V1.tif

COM\_YYMMDD\_TLM\_XY\_V1.tif

QUA\_YYMMDD\_TLM\_XY\_V1.csv

# 1. Data collection and Documentation

---

▾  1.3 Quelle documentation et quelles métadonnées allez-vous fournir avec les données ?

→ *Infos nécessaires pour futurs utilisateurs*

→ *Annotations*

### 1.3 Quelle documentation et quelles métadonnées allez-vous fournir avec les données ?

Un **fichier de métadonnées**, intitulé **README**, sera sauvegardé au premier niveau de l'arborescence du dossier 'Etude FNS'. Il contiendra les informations suivantes :

- Les **noms des chercheurs** ayant participé à l'étude
- Une description des **méthodes employées** pour collecter les données
- Les **dates** auxquelles la collecte a eu lieu
- Une description de l'**arborescence du dossier**, des **types de données** qui se trouvent dans chaque sous-dossier et de la **convention de nommage** des fichiers
- Une **explication** des intitulés et conventions utilisés dans les fichiers
- Les **conditions d'accès** aux données (licence), durant l'étude et après celle-ci

## 2. Questions éthiques, légales et de sécurité

---

### 2.1 Comment les questions éthiques seront-elles abordées et traitées ?

- *Protection des données*
- *Confidentialité*
- *Permissions*
- *Gestion des données sensibles*

## 2.1 Comment les questions éthiques seront-elles abordées et traitées ?

La collecte des données s'effectuera dans le cadre des **règlements éthiques** de l'UNIGE et de la Faculté ; elle sera validée par la **Commission universitaire d'éthique**.

Un **formulaire de consentement** imprimé sera rempli par chaque participant·e aux entretiens enregistrés. Il indiquera précisément comment et sous quelle forme le contenu des entretiens sera utilisé et diffusé.

En préambule du questionnaire en ligne figurera une demande de consentement de la participante ou du participant quant à l'utilisation de ses réponses. Si elle ou il n'accorde pas son consentement, il ne sera pas possible d'accéder à la suite du questionnaire.

Toutes les données seront **anonymisées** lors de leur préparation pour une diffusion plus large, en fin d'étude. Si, pour une raison ou une autre, il ne devait pas être possible d'anonymiser une partie des données, celle-ci ne sera pas diffusée.

## 2. Questions éthiques, légales et de sécurité

---

▾  2.2 Comment seront gérés l'accès aux données et la sécurité ?

### **Gestion des risques :**

→ *Sécurité des données*

→ *Numérique*

→ *Physique*

→ *Droits d'accès*

## 2.2 Comment seront gérés l'accès aux données et la sécurité ?

Les **principaux risques** pour la sécurité des données sont la perte ou la détérioration des cahiers de laboratoire et la perte ou la corruption des données électroniques. Les données seront protégées par les mesures suivantes :

1. Les données contenues dans **les carnets de laboratoire** seront également enregistrées sous forme électronique et sauvegardées quotidiennement afin de se prémunir contre la perte ou la détérioration du carnet.
2. L'accès aux données électroniques (avant publication) sera limité aux membres du groupe de recherche et aux collaborateurs concernés grâce à la **régulation d'accès aux lecteurs partagés sur le serveur de l'Université**.
3. L'accès aux laboratoires et aux bureaux est contrôlé par une **carte d'accès** afin de réduire la probabilité de perte ou de dommage malveillant ; tous les ordinateurs utilisés dans le cadre de ce projet fonctionneront sous le système Standard Staff Desktop, grâce auquel **les pare-feu et les logiciels antivirus** sont automatiquement mis à jour ; **les membres du personnel verrouilleront leur poste de travail** lorsqu'ils ne s'en serviront pas.

([source](#) of this ex., adapted)

## 2. Questions éthiques, légales et de sécurité

---

### 2.3 Comment allez-vous gérer les questions de droits d'auteur et de propriété intellectuelle ?

→ *Qui est le propriétaire des données ?*

→ *Quelle licence appliquer ?*

→ *Quelles restriction s'appliquent à la réutilisation de données de tiers*

## 2.3 Comment allez-vous gérer les questions de droits d'auteur et de propriété intellectuelle ?

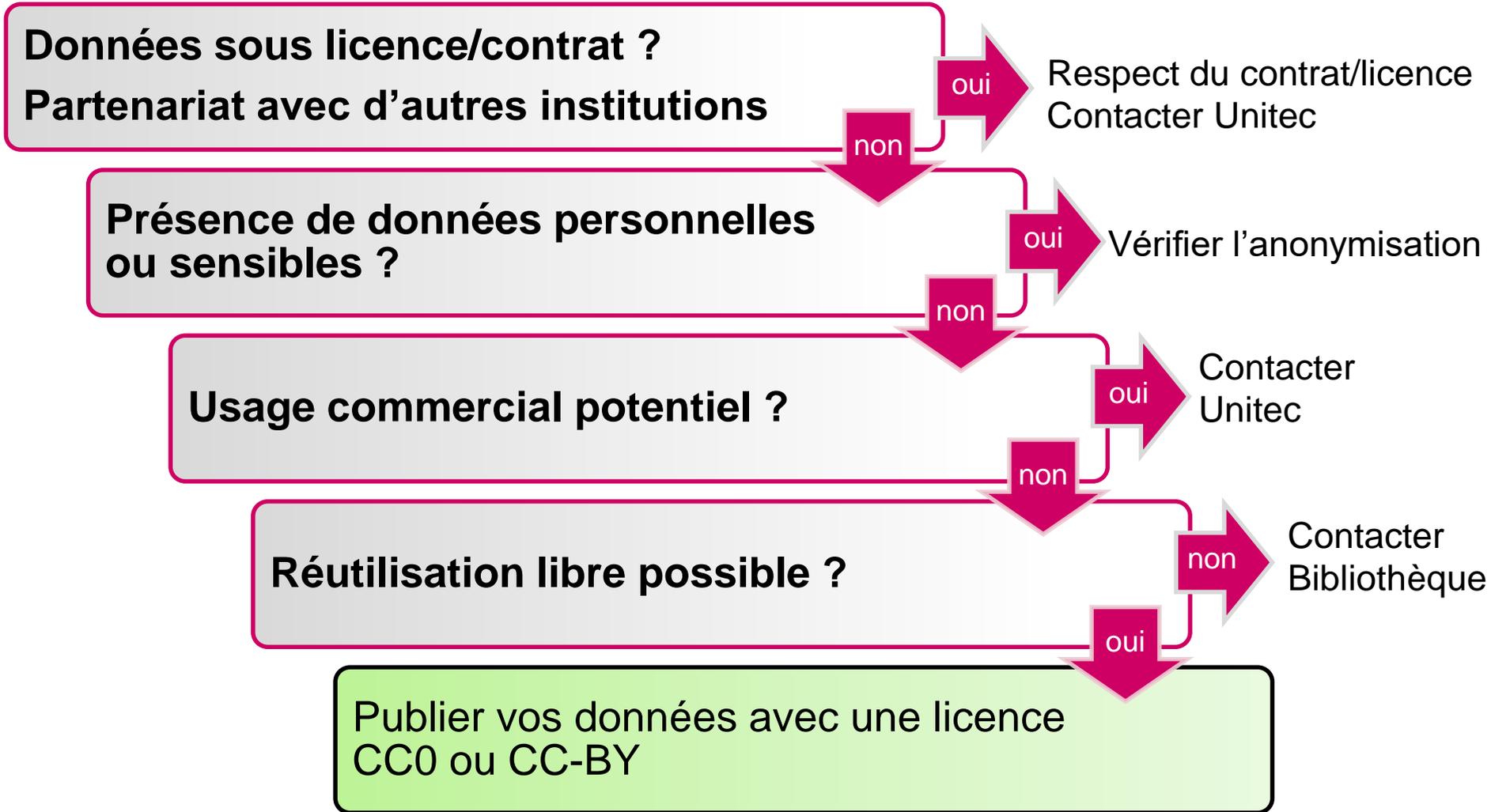
Les éditions spécifiques de certains textes analysés dans le cadre du projet sont **soumises à des droits d'auteur, et ne pourront être diffusés.**

Dans ces quelques cas, **des accords seront établis avec les propriétaires des collections concernées** permettant une diffusion au grand public d'un nombre limité d'images et d'extraits de texte sur le site web et les autres interfaces de diffusion.

**Les services compétents de l'Université seront sollicités sur ce point pour apporter un conseil expert et valider les accords.**

De nouvelles **données dérivées** (tableaux, diagrammes, compilations de données chiffrées, annotations...) de l'analyse de textes qui est au cœur du projet **seront librement partagées sous licence CC-BY.**

# CHOISIR UNE LICENCE



Voir aussi : <https://creativecommons.org/choose/>

# 3. Stockage et préservation des données

---

▾ □ 3.1 De quelle manière vos données seront-elles stockées et sauvegardées au cours de la recherche ?

**Pendant le projet :**

→ *Capacité et lieu de stockage*

→ *Sauvegardes*

3.1 De quelle manière vos données seront-elles stockées et sauvegardées au cours de la recherche ?

Au cours de la recherche, le dossier décrit au point 1.2 sera enregistré sur le **disque dur du poste de travail du PI**. Il sera sauvegardé

- chaque semaine sur un **disque dur externe**, stocké dans le coffre-fort de ce même bureau ; cette sauvegarde sera effectuée par le PI
- quotidiennement sur le **serveur NAS de l'UNIGE**, suivant une procédure automatique

Les **formulaires de consentement** imprimés seront également conservés dans le coffre-fort du bureau.

# 3. Stockage et préservation des données

---

  3.2 Quel est votre plan en matière de conservation des données ?

## **Après le projet :**

→ *Données à conserver, sauver*

→ *Formats de fichiers pour préservation*

3.2 Quel est votre plan en matière de conservation des données ?

Les **enregistrements audio** des entretiens et leurs transcriptions seront conservés en format .wav et .pdf pour une **période minimale de 10 ans** sur un **disque dur** se trouvant dans le coffre-fort du bureau du PI.

Les **formulaires de consentement** seront conservés **10 ans** dans le même coffre-fort.

Les **synthèses d'entretiens** et les **résultats des questionnaires**, après anonymisation, seront diffusés avec une licence CC-BY, en **format .csv**.

Les données qu'il n'aura pas été possible d'anonymiser ne seront pas diffusées.

# 4. Partage et réutilisation des données

---

▣ 4.1 De quelle manière et où seront partagées les données ?

## **Archivage et publication :**

→ *Quel dépôt pour le partage ?*

→ *Comment les utilisateurs vont trouver mes données ?*

▾ □ 4.1 De quelle manière et où seront partagées les données ?

Les données seront partagées sur **Yareta**, le data repository de l'Université de Genève. Elles seront accompagnées du fichier de métadonnées README décrit au point 1.3.

Une licence CC-BY sera attribuée.

Les données seront conservées au moins 10 ans après le dernier accès.

# 4. Partage et réutilisation des données

---

▾  4.2 Y a-t-il des restrictions nécessaires pour protéger les données sensibles ?

→ *Conditions de mise à disposition*

→ *Timing*

→ *Délais*

▼  4.2 Y a-t-il des restrictions nécessaires pour protéger les données sensibles ?

Les données individuelles des sujets de recherche ne peuvent pas, légalement ou éthiquement, être mises à la disposition de personnes non autorisées (LRH, cf. §2.1). Seuls le bailleurs de fonds, l'équipe de recherche, les examinateurs, les auditeurs et les autorités de contrôle sont autorisés à accéder à ces données.

**Aucune donnée personnelle ou susceptible d'identifier facilement les sujets ne sera fournie**, conformément à la loi suisse sur la recherche sur l'être humain (Loi fédérale sur la recherche impliquant des êtres humains (LRH)) et à son ordonnance applicable ClinO/KlinV/OClin/OSRUm.

(eg from [DMP Model](#), HUG CRC)

## 4. Partage et réutilisation des données

---

▼  4.3 Je confirme que je choisirai exclusivement des bases de données (repositories) numériques conformes aux FAIR Data Principles

→  *[case à cocher]*

▼  4.4 Les bases de données (repositories) choisies pour le dépôt des données sont gérées par une organisation à but non lucratif.

→ *Oui / non [radio button]*



# L'ESSENTIEL EN BREF

---

- ✓ Un DMP plausible est souvent une condition pour le déblocage des fonds
- ✓ DMPs sont modifiable tout le long de la période de financement (doivent être mis à jour)
- ✓ Une fois le financement du FNS terminé et le rapport scientifique final approuvé, le DMP ne peut plus être modifié.

# WWW.UNIGE.CH/RESEARCHDATA



UNIVERSITÉ  
DE GENÈVE

UNIVERSITÉ

FACULTÉS

ÉTUDIANTS

COLLABORATEURS

SERVICES

ALUMNI

Portail | | FR EN |

## DONNÉES DE RECHERCHE



Planifier

Collecter & Organiser

Stocker

Préserver

Partager

Services & Formations

À propos

DataFORUM

Ecosystème Data UNIGE

Relecture de votre DMP

Rendez-vous de l'info scientifique

PLANIFIER

COLLECTER & ORGANISER

STOCKER

PRÉSERVER

PARTAGER



✓ IDENTIFIER LES DONNÉES DE RECHERCHE

✓ À PROPOS DU DMP

✓ REMPLIR LE DMP DU FNS

# MERCI POUR VOTRE ATTENTION

[www.unige.ch/biblio](http://www.unige.ch/biblio)

<https://www.unige.ch/researchdata/fr/>

Toutes les images non citées ici sont dans le domaine public



Bibliothèque de l'UNIGE, 2021

Ce document est sous licence Creative Commons Attribution - Partage dans les mêmes conditions 4.0 International : <http://creativecommons.org/licenses/by-sa/4.0/deed.fr>.

LIBRARY



UNIVERSITÉ  
DE GENÈVE