

DATA MANAGEMENT PLAN (DMP) DU FNS



Comment le remplir et quelles informations y faire figurer ?

Anouk Santos, Claire Wuillemin

Mai 2021

Plan



À propos des **données de recherche** et des exigences du FNS



Découverte du **plan de gestion des données** (Data Management Plan – DMP)



L'essentiel en bref



Nos services pour rédiger votre DMP

- ✓ **Adoption** par le rectorat, après concertation avec les décanats (juin 2018)
- ✓ **1^{ère} haute école de Suisse** avec une politique sur la gestion des données de recherche
- ✓ Définition des **responsabilités** :
 - de l'Université
 - des chercheurs et chercheuses

<https://www.unige.ch/researchdata/fr/services/all/politique/>

Données de la recherche ?



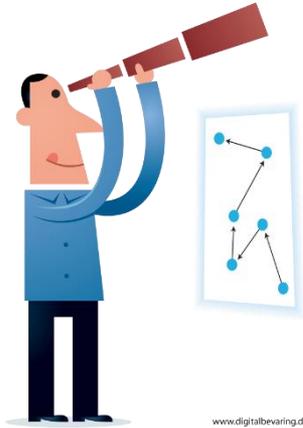
Enregistrements factuels (chiffres, textes, images et sons), qui sont utilisés comme **sources principales** pour la recherche scientifique et sont généralement reconnus par la communauté scientifique comme **nécessaires pour valider des résultats de recherche.**



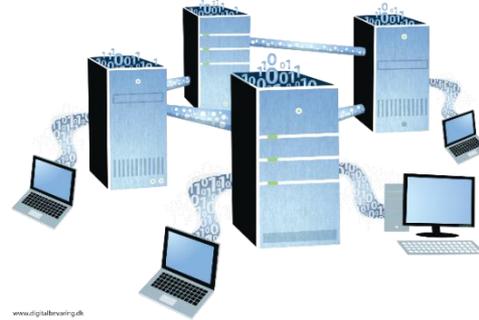
Principes et lignes directrices de l'OCDE pour l'accès aux données de la recherche financée sur fonds publics, 2007, p.18. <http://www.oecd.org/fr/sti/sci-tech/38500823.pdf>

Types de données de recherche

Observationnelles



Dérivées / compilées

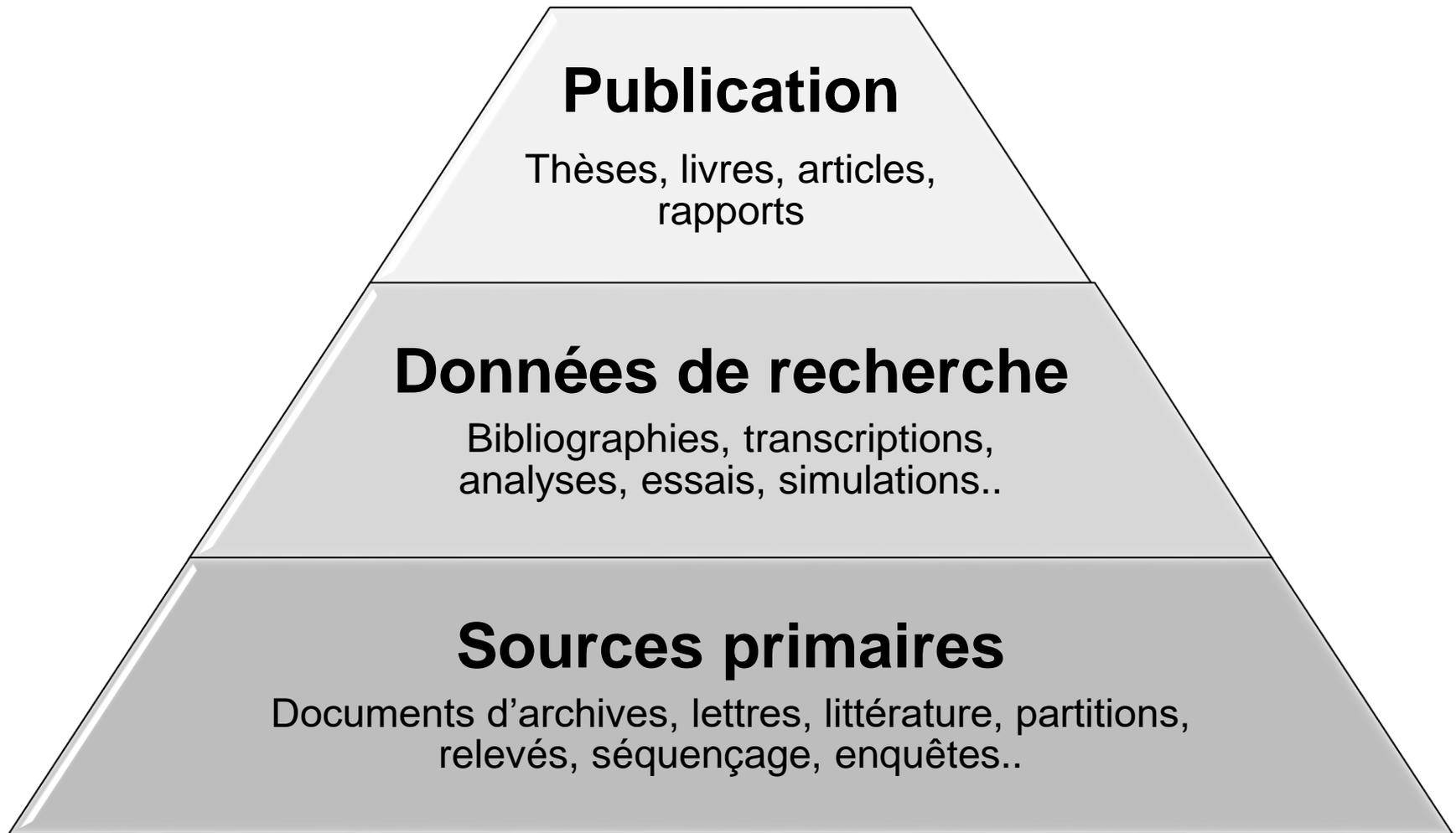


Simulées / modélisées

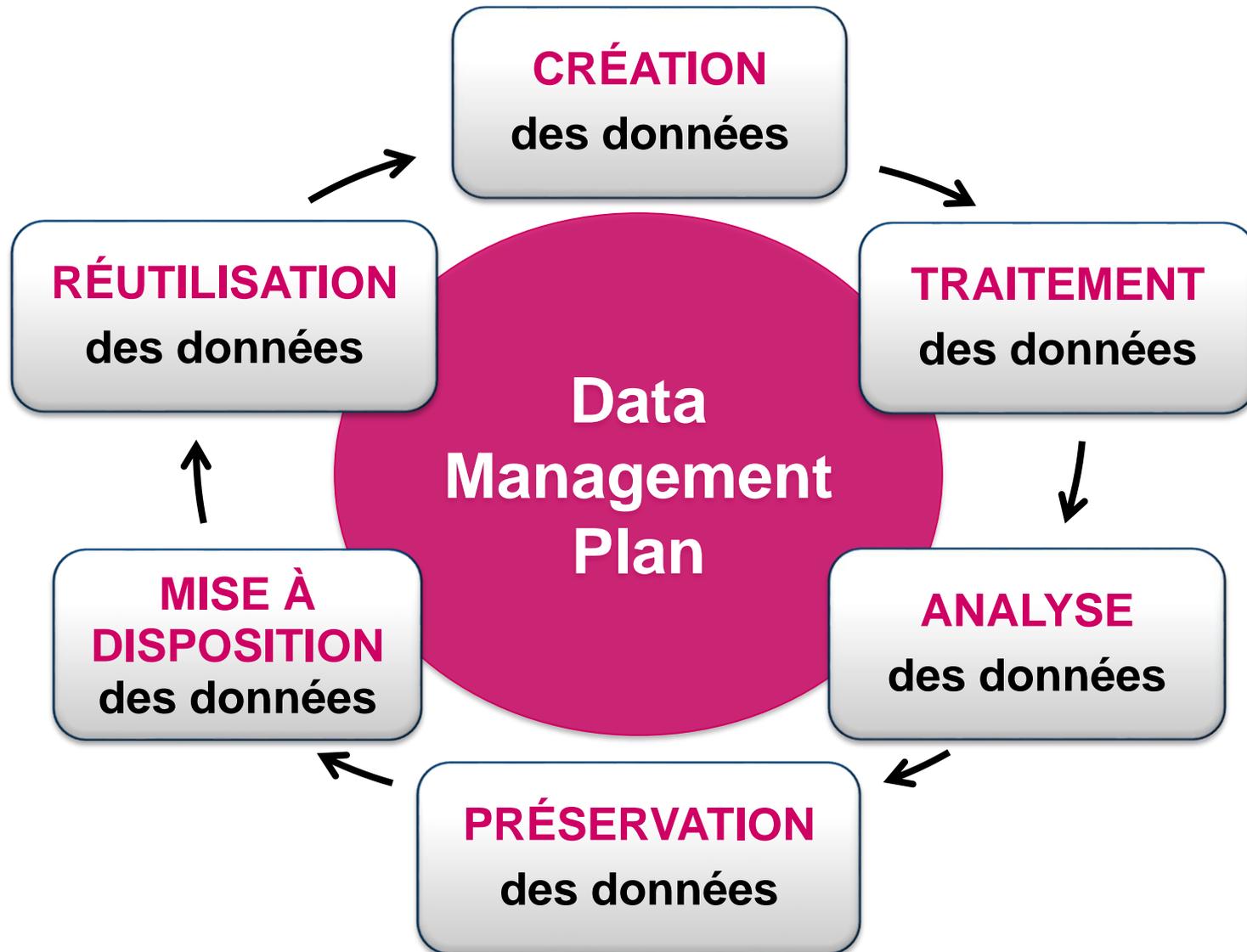


Expérimentales

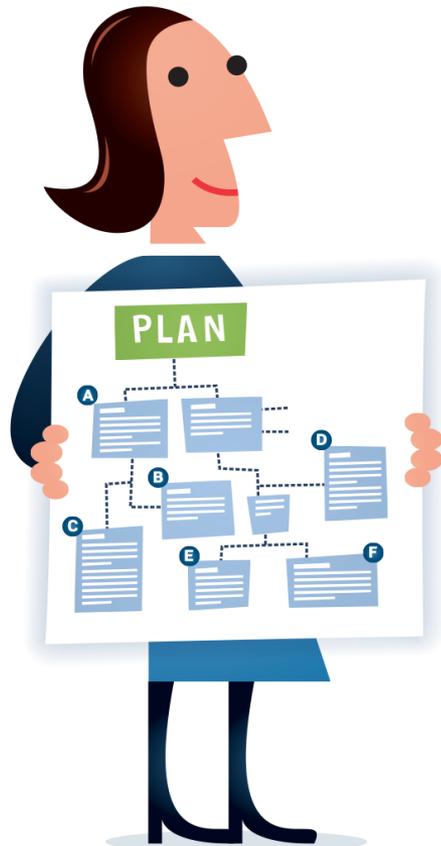
de références



Cycle de vie des données



Plan de gestion des données



www.digitalbevaring.dk

- ✓ Gagner du **temps**
- ✓ Augmenter l'**impact** de votre recherche
- ✓ Participer à la **science ouverte**
- ✓ Faciliter de **nouvelles découvertes**

Dépôt d'une requête au FNS



2. Application Data

#	No./Title	Status
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	In preparation
2.5	Continuation of	
2.6	Link to other SNSF projects	
2.7	Further requested and available funds (not from	
2.8	University or research institution	
2.9	Requested funding	
2.10	Data management plan (DMP)	
2.11	Research requiring authorisation or notification	
2.12	Exclusion of external reviewers	
2.13	General remarks on the project	

2. 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	Continuation
2.6	Relation avec d'autres projets FNS
2.7	Autres moyens demandés et disponibles (qui ne sont pas du FNS)
2.8	Haute école
2.9	Besoin financier
2.10	Plan de gestion des données (DMP)
2.11	Recherche exigeant des autorisations ou des annonces

Le DMP sur MySNF

Je ne fournis pas de DMP pour la raison suivante:

1. Collecte des données et documentation

- 1.1 Quelles données allez-vous collecter, étudier, générer ou réutiliser ?
- 1.2 Comment les données seront-elles recueillies, étudiées ou générées ?
- 1.3 Quelle documentation et quelles métadonnées allez-vous fournir avec les données ?

2. Questions éthiques, légales et de sécurité

- 2.1 Comment les questions éthiques seront-elles abordées et traitées ?
- 2.2 Comment seront gérés l'accès aux données et la sécurité ?
- 2.3 Comment allez-vous gérer les questions de droits d'auteur et de propriété intellectuelle ?

3. Stockage et préservation des données

- 3.1 De quelle manière vos données seront-elles stockées et sauvegardées au cours de la recherche ?
- 3.2 Quel est votre plan en matière de conservation des données ?

4. Partage et réutilisation des données

- 4.1 De quelle manière et où seront partagées les données ?
- 4.2 Y a-t-il des restrictions nécessaires pour protéger les données sensibles ?
- 4.3 Je confirme que je choisirai exclusivement des bases de données (repositories) numériques conformes aux FAIR Data Principles
- 4.4 Les bases de données (repositories) choisies pour le dépôt des données sont gérées par une organisation à but non lucratif.

1. Collecte des données et documentation

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

Liste des données :

→ *Type*

→ *Format*

→ *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 générera 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 en 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.

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

Qualitative (e.g. phenotyping data for cell lines), **quantitative** (e.g. cell counts), mass spectrometry and **image** data will be generated. Raw data will be analysed and expressed as graphs, tables and annotated images.

Data generated will be in **various formats and sizes of datasets**. They include:

1. **Cell images** e.g. phase and fluorescence, and electron micrographs (**~5,000 images** over project). Software used includes OpenLab, Softworx and IN Cell Investigator, with data saved as **software-specific files e.g. liff and lg3 files**, as well as generic formats such as **jpeg, tiff** etc.
2. **mass spectrometry spectra** (from **<50 samples**). MS data will be analysed using Bruker Data Analysis or Thermo Excalibur software (generating **xml and raw files**) and proteins will be matched to the *T. brucei* genome dataset using the Matrix Science Mascot search engine. Each LC-MS data file is between **1-2GB**.
3. Cell line phenotyping data including growth curves and DAPI counts (**Excel and GraphPadPrism files**) and flow cytometry data (FlowJo and **jpeg/tiff files**) (~200 data sets).
4. ...

([source](#))

1. Collecte des données et 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 ?

All samples on which data are collected will be prepared according to **published standard protocols in the field**.

Quality of analytical data will be guaranteed through **calibration of devices** and comparison with internal standards. Appropriate experimental design, data recording and data validation (controls, randomization/ blinding, sampling/replicates, experimental versus hypothesis driven-protocol) will be used, ensuring internal validity.

Methods and materials will be recorded using an **Electronic Lab Notebook (ELN)**. All experimental data will be automatically imported into a **Laboratory Information Management System (LIMS)** from the measurement device.

Files will be named according to a pre-agreed **convention** including a short, unique identifier for the project, a summary of the content, the date (YYYYMMDD format) and the document version. Ex. : **YYYYMMDD_ProjectIDX_ContentSummary_V01**

Peer review will be employed as an additional step of quality control in the form of **regular supervision and lab meetings**.

(eg from [DMP Canva Generator](#), VitalIT)

1. Collecte des données et 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

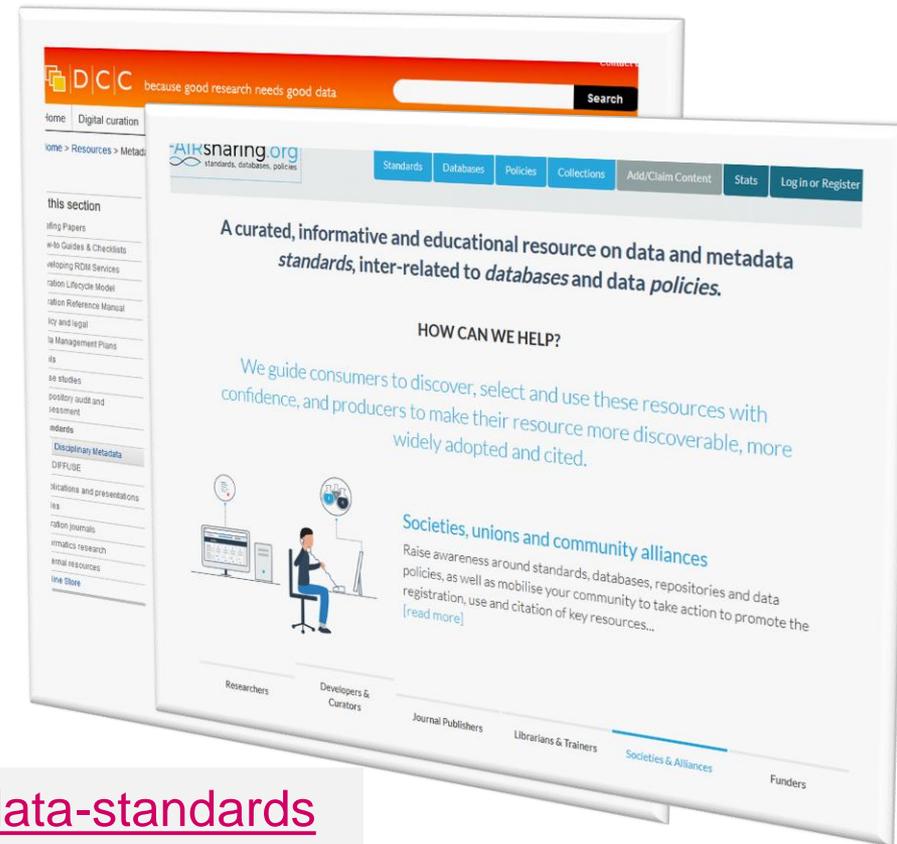
- Utilisez dans la mesure du possible des **standards** de métadonnées existants
- **Documentez votre schéma** de métadonnées le plus tôt possible
- Ex. de métadonnées : *auteur, titre, date, type et marque des appareils...*

<http://www.dcc.ac.uk/resources/metadata-standards>

<https://fairsharing.org/>

- Conseils et modèles d'un fichier README.txt

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



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.

Quelle commission d'éthique ?

Si votre recherche implique :

- des personnes vivantes ou décédées;
- des embryons et des fœtus;
- du matériel biologique;
- des données personnelles liées à la santé.

Alors elle relève de la LRH*,
et vous devez vous adresser à

Commission cantonale
d'éthique et de recherche
(CCER)

* [Loi relative à la recherche sur l'être humain](#)

Si votre recherche implique des participant·e·s humain·e·s
mais n'entre pas dans le champ d'application de la LRH,

vous devez vous adresser à

Commission universitaire
d'éthique
(CUREG)

- Charte d'éthique et de déontologie des hautes écoles universitaires et spécialisées de Genève

« Toute personne engagée dans la recherche répond à une exigence d'intégrité dans sa quête de la connaissance, dans l'interprétation et l'application des résultats et dans la mise en forme de sa recherche ». (Article 1, lettre c.)

- Directive sur l'intégrité de la recherche

« Le responsable de projet doit veiller à ce que les données de base soient conservées en sécurité pendant cinq ans au moins après le terme de la recherche. » (point 2.6)

- Loi fédérale relative à la recherche sur l'être humain (LRH)
- Commission cantonale d'éthique de recherche (CCER)
- Commission universitaire d'éthique (CUREG 2.0)

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 ?

The **main risks** to data security are loss or damage to laboratory notebooks and loss or corruption of electronic data. Data will be safeguarded by the following measures:

1. Data in **lab notebooks** will also be recorded in electronic form that is backed up daily to secure against loss or damage of the notebook.
2. Access to electronic data (prior to publication) will be limited to the members of the research group and relevant collaborators via **limiting access to shared drives on the University server**.
3. Access to laboratories and offices are controlled by **card access** to reduce the likelihood of malicious loss/damage; all computers used in this project will run Standard Staff Desktop, whereby **firewalls and antivirus software** are automatically upgraded; staff will **lock their workstation** whenever they are away from it.

([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 ?

Aspects juridiques :

- *Qui est le propriétaire des données ?*
- *Quelle licence appliquer ?*

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

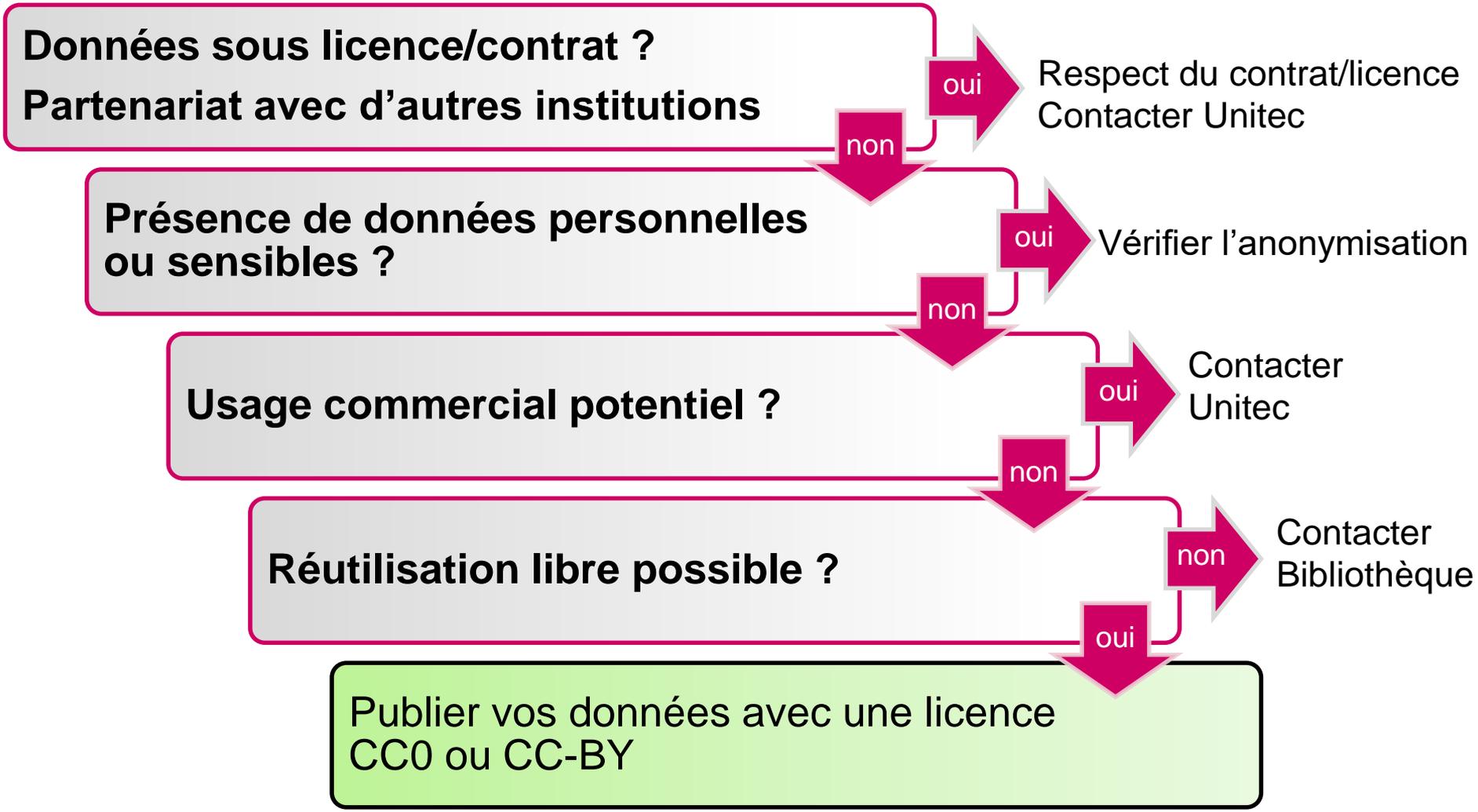
Les éditions spécifiques de certains textes analysés dans le cadre du projet sont **soumises à des droits d'auteur, et ne pourront être diffusés.**

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 sa licence pour partager



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.

Bonnes pratiques de sauvegarde

→ Toujours stocker **trois copies** des données



1. L'original
2. Une **copie** sur un dispositif externe sur place
3. Une **copie** sur un dispositif externe dans un **autre lieu**

Le **NAS académique**, espace de stockage partagé, couvre les points 2. et 3.

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

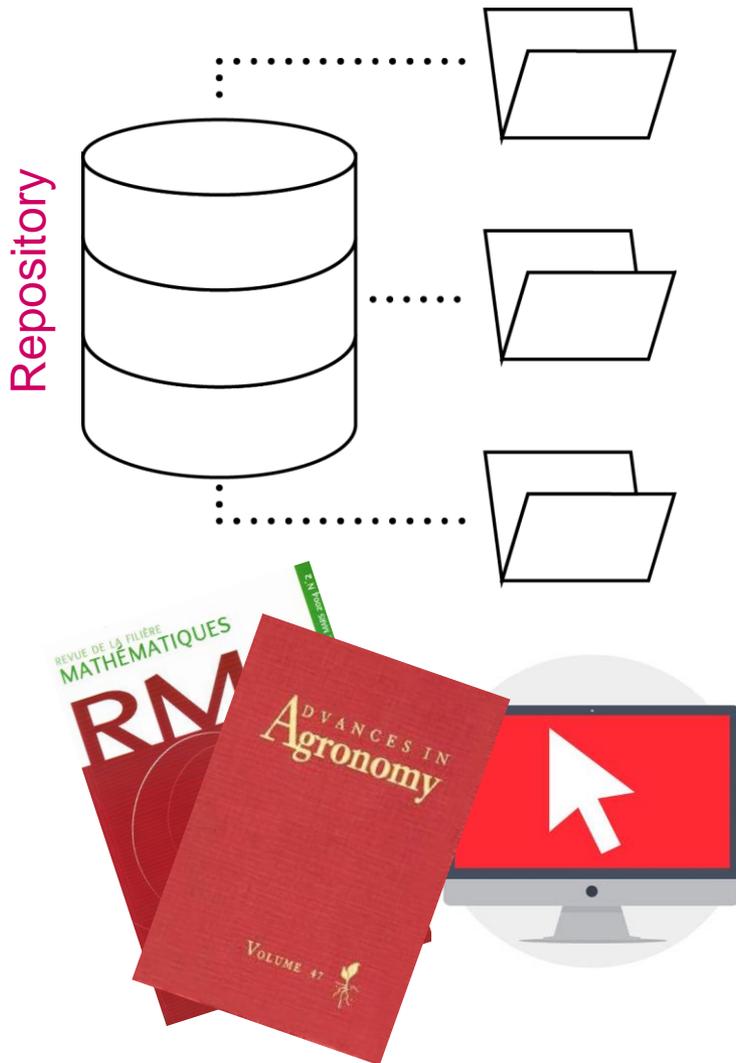
4.1 De quelle manière et où seront partagées les données ?

Archivage et publication :

→ *Quel dépôt pour le partage ?*

→ *Comment les utilisateurs vont trouver mes données ?*

Archivage et diffusion

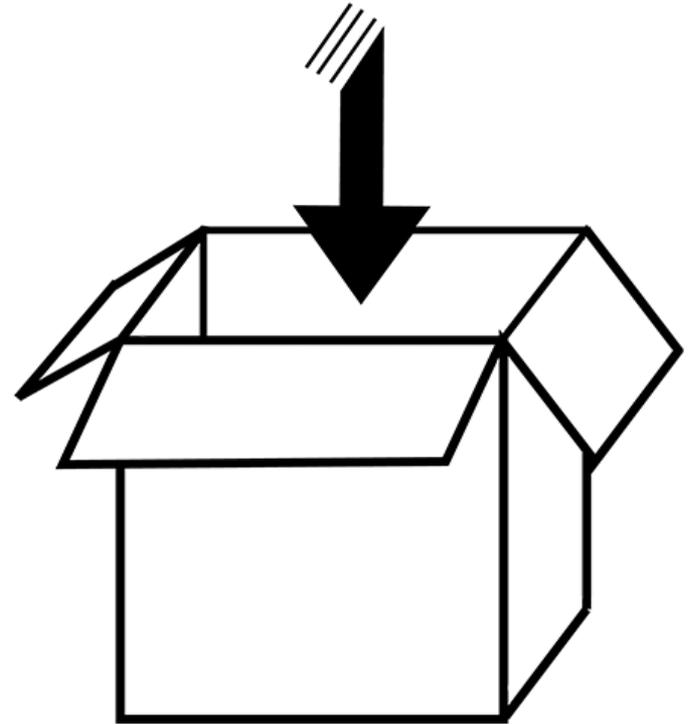


- ✓ générique (multi-disciplinaire)
- ✓ disciplinaire
- ✓ institutionnel

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

Trouver un repository

- ✓ **Yareta**
- ✓ Zenodo
- ✓ OSF
- ✓ Dryad
- ✓ ...



Pour tous les autres dépôts, consultez : re3data.org
REGISTRY OF RESEARCH DATA REPOSITORIES

- ✓ Créé dans le cadre d'un projet de loi cantonal
- ✓ Repository pour les **chercheurs genevois**
- ✓ Serveurs (original et copie) en **Suisse** 
- ✓ Répond aux **principes FAIR &** attribue des **DOI**
- ✓ **Gratuit** jusqu'à 50 Go
(au-delà, 100 CHF /Tb/année de préservation)

<https://yareta.unige.ch>

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 ?

Individual research subjects' data cannot legally nor ethically be made available to non authorised people (HRA, *cf.* §2.1). Only the sponsor, the investigation team, reviewers, auditors and inspection authorities are entitled to access such data.

No personal data or data that may easily identify subjects will be provided, with respect to the Swiss law on human research (Federal Act on Research involving Human Beings (HRA)) and its applicable ordinance 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

→ *[checkbox]*

- 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]*

Mon repository est-il FAIR ?

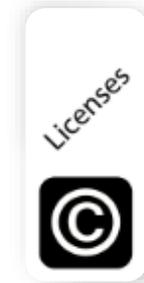


Findable

Accessible

Interoperable

Reusable



Au moins une icône bleue est présente

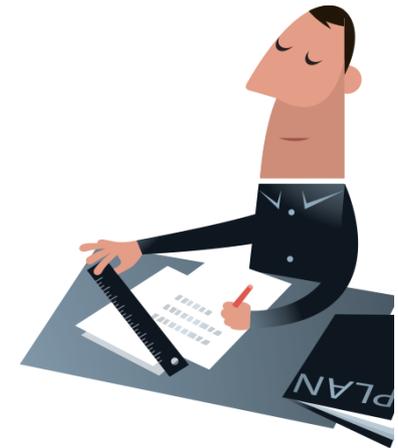
Présence d'une icône cadenas orange

Présence d'une icône licence

Coûts de gestion

Le FNS permet de demander au maximum **10'000.- CHF en plus** pour :

- **Préparation et validation** des données
- **Chargement** (upload) des données sur un repository FAIR non-commercial



re3data.org

Repository details

Zenodo

General **Institutions** Terms Standards

Institution name **European Organization for Nuclear Research**

...

Type of institution **non-profit**



Déblocage des fonds subordonné à la soumission d'un DMP « plausible »

DMP **modifiable** durant toute la durée de vie du subside (doit évoluer au cours du projet)

- ✓ Obligation de **mettre à jour le DMP à la fin du subside** pour être examiné en même temps que le rapport scientifique final
- ✓ Version finale du DMP mise à disposition dans la **base de données P3** du FNS

DONNÉES DE RECHERCHE



Planifier

Collecter & Organiser

Stocker

Préserver

Partager

Services & Formations

À propos

Accueil

Bienvenue sur le site Researchdata qui regroupe toutes les informations sur les questions de gestion des données de recherche.

GESTION DES DONNÉES DE RECHERCHE

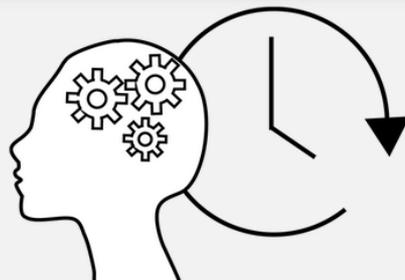
PLANIFIER |

COLLECTER & ORGANISER |

STOCKER |

PRÉSERVER |

PARTAGER



✓ À PROPOS DU DMP

✓ REMPLIR LE DMP DU FNS

✓ H2o2o - OPEN RESEARCH DATA PILOTE

Ecosystème Data UNIGE

<https://datascience.unige.ch/ecosysteme-data-unige>

UNIVERSITÉ DE GENÈVE

UNIVERSITÉ FACULTÉS ETUDIANTS COLLABORATEURS SERVICES ALUMNI

Portail | Q | FR | EN |

CENTRE DE COMPÉTENCES EN SCIENCE DES DONNÉES

Accueil Le CCSD Réseau d'experts Recherche Formation Cité Ecosystème Data Actualités

ECOSYÈME DATA

Une multitude d'initiatives et de services liés aux données sont à la disposition de la communauté UNIGE. En travaillant ensemble, ces initiatives constituent l'écosystème de données de l'UNIGE.

COMMUNAUTÉS DE CHERCHEURS ET DE CHERCHEUSES

CENTRE DE COMPÉTENCES EN SCIENCE DES DONNÉES

Centre de compétences en science des données (CCSD)

Constitue autour d'un réseau de chercheurs et chercheuses, le Centre de Compétences en Science des Données (CCSD) de l'Université de Genève a pour mission de fédérer les compétences et initiatives de l'UNIGE en matière de science des données, dans le but, par la transversalité, de favoriser l'émergence de recherches innovantes et d'accompagner une transformation critique et informée de la culture scientifique à l'ère des données massives. Pour ce faire, il organise de nombreuses activités de rencontre scientifique comme des séminaires, des cliniques, des journées d'études, etc.

[Vers le site du CCSD >>](#)

DataFORUM

Echange et entraide sur les données de recherche
Research data discussion and assistance

DataFORUM

Ouvert à toute la communauté des hautes écoles genevoises, Data Forum est un espace d'échange et d'entraide sur les questions pouvant apparaître aux différentes étapes du cycle de vie des données de recherche (gestion et analyse).

[Vers le DataFORUM >>](#)

Groupe rdf-Linked-data

Le groupe *rdf-linked-data* a pour objectif de mettre en relation les personnes et équipes de l'UNIGE qui ont un intérêt pour les techniques du Web sémantique: graphes de données et de connaissances (RDF); ontologies; langages de requête et d'inférence; stockage et publication de données liées; architectures logicielles. Les échanges d'idées et d'expériences ont lieu lors de présentations sur un thème précis ou directement à travers la liste de distribution.

[Vers le site du groupe *rdf-linked-data* >>](#)

[Inscrivez-vous à la liste de diffusion >>](#)

RN CH Swiss Reproducibility Network

Swiss reproducibility network

Le Swiss Reproducibility Network (SwissRN) est un consortium dirigé par des pairs qui vise à promouvoir et à garantir des pratiques de recherche rigoureuses en Suisse par la mise en place d'activités de formation, la conception et l'évaluation des efforts d'amélioration de la recherche, et la diffusion des meilleures pratiques, tout en travaillant avec les parties prenantes pour assurer la coordination des efforts dans l'ensemble du secteur (e.g., le Fonds national suisse de la recherche scientifique, Académies suisses des sciences).

[Vers le site du SwissRN >>](#)

Une multitude d'initiatives et de services liés aux données sont à la disposition de la communauté UNIGE.

Merci pour votre attention

Pour plus d'informations sur les DMP et les données de recherche:



researchdata-info@unige.ch

<http://www.unige.ch/researchdata>

Illustrations : CC BY Jørgen Stamp, <http://digitalbevaring.dk>



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>.