

CÉCILE ARÈNES
CHARGÉE DE MISSION DONNÉES DE LA
RECHERCHE ET HUMANITÉS NUMÉRIQUES

2022 – 2023

DONNEES DE LA RECHERCHE ET PLANS DE GESTION DES DONNEES



Sommaire

- 1. LES DONNÉES DE LA RECHERCHE**
- 2. POURQUOI OUVRIR LES DONNÉES ?**
- 3. LES DATA MANAGEMENT PLANS**
- 4. RÉDIGER LE DMP ET LES ÉLÉMENTS IMPORTANTS**
- 5. RÉDIGER ET AUTO-ÉVALUER SON DMP**
- 6. PUBLIER UN DATAPAPER**
- 7. NOUS CONTACTER**

1

LES DONNÉES DE LA RECHERCHE ↴

Le plan national pour la science ouverte – 2

[Plan national pour la science ouverte 2, 2021-2024.](#)

Axe 2 : structurer, partager et ouvrir les données de la recherche :

- Mettre en œuvre l'**obligation de diffusion des données de recherche financées sur fonds publics**
- Créer **Recherche Data Gouv**, la plateforme nationale fédérée des données de recherche
- Promouvoir l'adoption d'une **politique de données** sur l'ensemble du cycle des données de la recherche, pour les rendre faciles à trouver, accessibles, intéropérables et réutilisables (**FAIR**)

Axe 3 : ouvrir et promouvoir les codes sources produits par la recherche

- **Valoriser et soutenir la diffusion sous licence libre des codes sources** issus de recherches financées sur fonds publics
- **Mettre en valeur la production des codes sources** de l'enseignement supérieur, de la recherche et de l'innovation
- Définir et promouvoir une **politique en matière de logiciels libres**



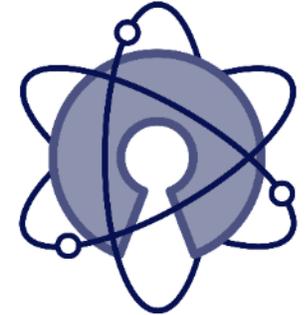
La feuille de route du MESRI pour les données et codes

[Politique des données, des algorithmes et des codes sources, 2021-2024.](#)

- **Structurer, ouvrir et partager les données de recherche - Action 6**
- **Suivre l'ouverture des données et des codes de la recherche Baromètre de la science ouverte - - Action 7**
- **Accompagner les chercheurs dans la gestion des données et le « FAIR by design » - Action 9**
- **Collecter, préserver et partager les codes sources – Software Heritage - Action 11**
- **Accélérer les demandes d'accès des chercheurs aux données publiques - Actions 18 & 9**
- **Favoriser l'accès des chercheurs aux données d'intérêt général- Action 20**
- **Mettre en place des identifiants uniques et pérennes - Actions 22 à 27**
- **Accompagner l'adoption de licences libres - Action 45**



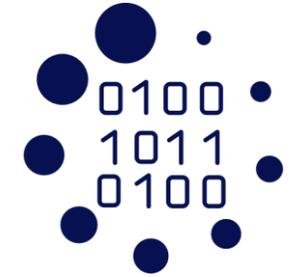
La science ouverte A Sorbonne Université



- Une [Charte de Sorbonne Université pour le libre accès aux publications](#)
- Une Déclaration sur le [droit des données de la recherche](#), signée en janvier 2020
- Une [Politique d'ouverture des données de la recherche](#) au sein de l'Alliance Sorbonne Université
 - inciter leurs équipes et personnels de recherche à ouvrir les données issues de la recherche autant que possible ;
 - promouvoir l'élaboration de plans de gestion des données dès la conception des protocoles de recherche ;
 - valoriser l'engagement des personnels de recherche qui partagent leurs données et œuvrent pour la science ouverte ;
 - mettre en place des programmes de formation et mobiliser les moyens nécessaires afin de créer un environnement favorisant l'ouverture des données.

Les données de la recherche

Qu'est-ce que c'est ?



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.

Les données de la recherche

Qu'est-ce que c'est ?



« Research data means data in the form of **facts, observations, images, computer program results, recordings, measurements or experiences** on which an **argument, theory, test or hypothesis, or another research output is based.**

Data may be **numerical, descriptive, visual or tactile.** It may be raw, cleaned or processed, and may be held in any format or media. »

DANS, « What Is Research Data », 2017.

https://www.ands.org.au/_data/assets/pdf_file/0006/731823/Whatis-research-data.pdf



Les données de la recherche

Cerner le périmètre

Dans le cadre du plan de gestion des données, on ne prendra pas en compte ces productions :

- Analyses préliminaires et projets de documents scientifique
- Programmes de travaux futurs
- Examens par les pairs
- Communications personnelles avec des collègues
- Objets matériels
- Publications scientifiques
- Supports de formations
- Données administratives

Attention, elles constituent des archives et certaines sont à conserver de façon pérenne.

(voir : [Référentiel de gestion des archives de la recherche](#))

2

POURQUOI OUVRIR LES DONNÉES ? ↴

Cadre juridique de l'ouverture des données

Ouverture des données de la recherche.
Guide d'analyse du cadre juridique en
France. 2017.
https://www.ouvrirlascience.fr/wp-content/uploads/2018/11/Guide_Juridique_V2.pdf

Arènes, C., Maurel, L., & Rennes, S. (2022).
*Guide d'application de la Loi pour une
République numérique pour les données de
la recherche* [Report]. Comité pour la science
ouverte. <https://doi.org/10.52949/31>;
<https://hal-lara.archives-ouvertes.fr/hal-03968218>

Ouverture des données de recherche

Guide d'analyse
du cadre juridique en France


Contenu sous licence ouverte

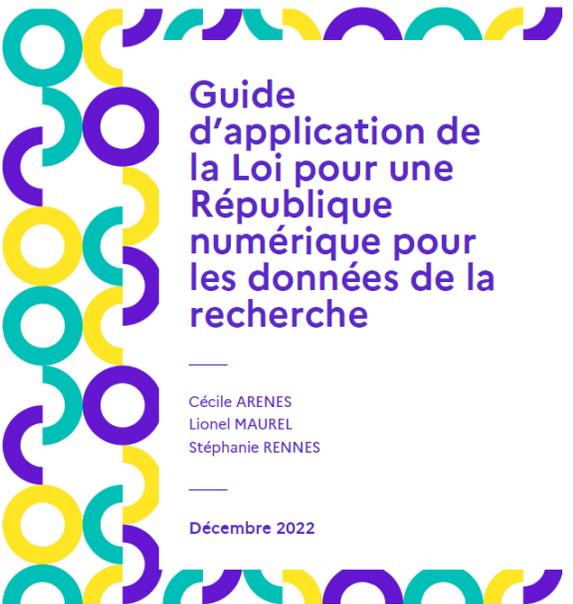
Le présent guide est issu des réflexions d'un groupe de travail inter-organismes animé par l'INRAE. Il ne prétend pas à l'exhaustivité et est fourni uniquement à titre d'information. Il ne saurait en tout état de cause se substituer aux pratiques d'établissements, au respect des dispositions législatives ou réglementaires et au respect de la jurisprudence applicable en la matière. Ce guide peut évoluer.
Membres du groupe de travail : BÉCARD Nicolas (INRAE), CASSEYS-RENAUD Céline (DTU), CHASSANG Gaëlle (Inserm),
Membre de la Plateforme Générale Société, DANFANT Martin, FLETY-CATIN Laurence (Inrae), GARDON Nathalie
(co-animatrice, INRA), MARTIN Caroline (AgroParisTech), MARTELLETTI Andrea (Scalable INRA, M2 droit et informatique),
MENDOZA-CAMINADE Alexandra (DTU), MORERETTE Nathalie (co-animatrice, INRA), NERAC Claire (Oréal),
avec la participation d'Inra? (Benjamin JEAN, Laure KASSAM).



Avec le soutien du Comité pour la science ouverte

V2 - Décembre 2017

 Ouvrir
la science!

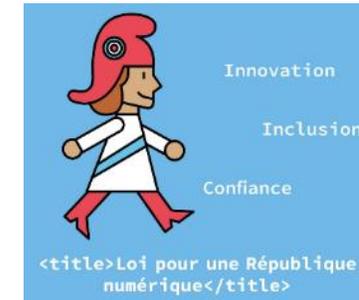


Guide
d'application de
la Loi pour une
République
numérique pour
les données de la
recherche

Cécile ARENS
Lionel MAUREL
Stéphanie RENNES

Décembre 2022

Loi pour une République numérique



Open research data : article 30

« II.- Dès lors que les **données** issues d'une activité de recherche **financée** au moins pour **moitié** par des dotations de l'Etat, des collectivités territoriales, des établissements publics, des subventions d'agences de financement nationales ou par des fonds de l'Union européenne **ne sont pas protégées par un droit spécifique** ou une réglementation particulière et qu'elles ont été rendues publiques par le chercheur, l'établissement ou l'organisme de recherche, **leur réutilisation est libre.**

« III.- L'**éditeur** d'un écrit scientifique mentionné au I **ne peut limiter la réutilisation** des données de la recherche rendues publiques dans le cadre de sa publication.

Image : https://www.economie.gouv.fr/files/files/PDF/DP_LoiNumerique.pdf

Le décret intégrité scientifique du 3 décembre 2021

Article 2 :

Les établissements [...] **promeuvent la diffusion** des publications en accès ouvert et la **mise à disposition des méthodes et protocoles, des données et des codes sources associés aux résultats de la recherche afin d'en garantir la traçabilité et la reproductibilité**. Ils incitent à la **publication des résultats de recherche dits négatifs**



BIBLIOTHÈQUE
UNIVERSITAIRE

Article 5 :

Les **résultats bruts** des travaux scientifiques mentionnés au troisième alinéa de l'article L. 211-2 du code de la recherche sont constitués des **données produites au cours du processus de recherche, ou à défaut des données traitées et scientifiquement validées, ainsi que des codes sources utilisés dans le traitement de ces données**.

Article 6 :

Les établissements publics et fondations reconnues d'utilité publique mentionnés au troisième alinéa de l'article L. 211-2 du code de la recherche **définissent une politique de conservation, de communication et de réutilisation des résultats bruts des travaux scientifiques menés en son sein**. A cet effet, ils veillent à la **mise en œuvre par leur personnel de plans de gestion de données et contribuent aux infrastructures qui permettent la conservation, la communication et la réutilisation des données et des codes sources**.

Principe d'ouverture des données

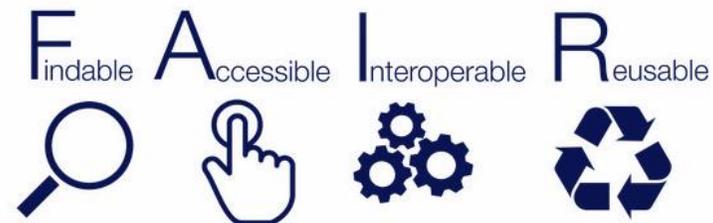


« As **open** as possible, as **closed** as necessary. »

- Un principe d'**ouverture par défaut** ;
- Des **limitations** en fonction du type de données.

Mise en œuvre : production de données FAIR

Principes



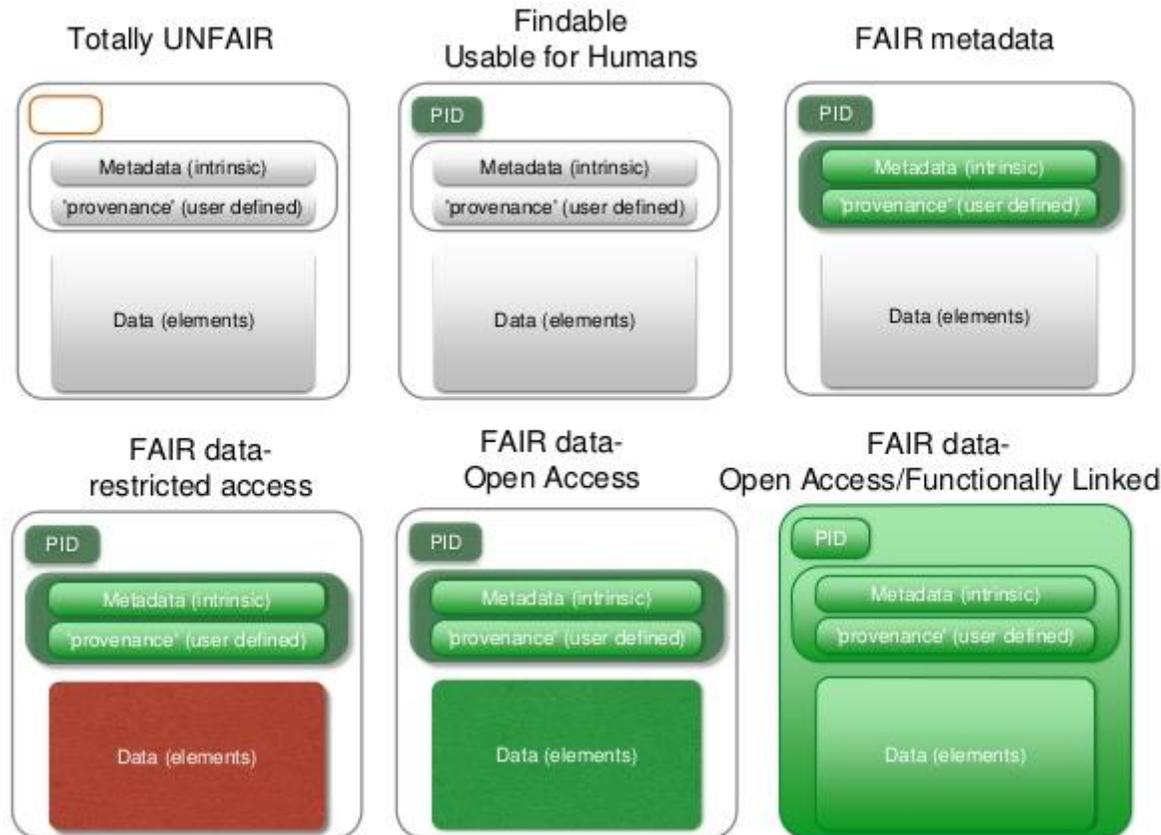
Détail des principes FAIR :
<https://www.force11.org/group/fairgroup/fairprinciples>

Logo FAIR: SangyaPundir. CC: BY-SA 4.0.

Image : [ANDS](#). CC: BY 4,0,



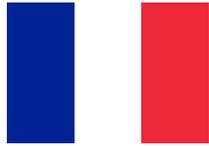
Data as increasingly FAIR Digital Objects



Mons B, Neylon C, Velterop J, Dumontier M,
 da Silva Santos LOB, Wilkinson MD.
 Cloudy, increasingly FAIR; revisiting the
 FAIR Data guiding principles for the
 European Open Science Cloud.
 Information Services & Use.
 2017;37(1):49-56.
<https://doi.org/10.3233/ISU-170824>

Ouverture des données et DMP

Exigences des financeurs



ANR :

- **DMP obligatoire** pour tout projet
- Pas d'obligation d'ouverture des données



Europe, Horizon Europe :

- **DMP obligatoire** pour tout projet
- Gestion des données **FAIR**
- **Ouverture** des données si leur régime juridique le permet



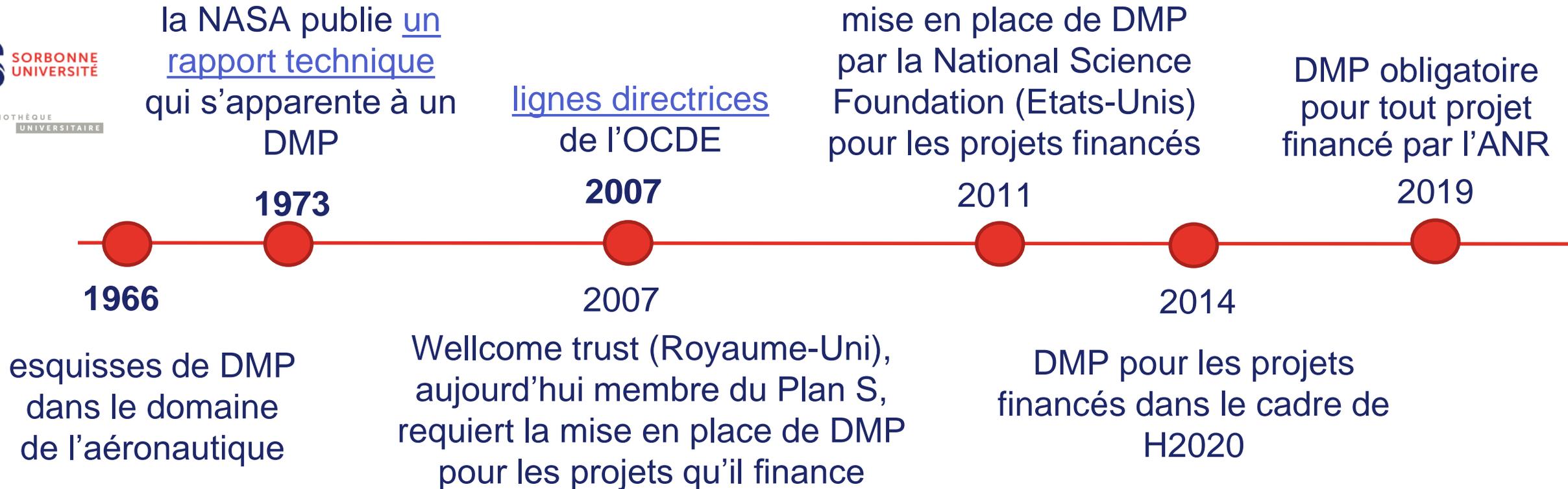
Europe, H2020 :

- Si pilote, **DMP obligatoire** pour tout projet
- Si pilote, **Ouverture** des données si leur régime juridique le permet

3

LES DATA MANAGEMENT PLANS [↓](#)

Le DMP en quelques dates



DATAACC. « Gestion des données : une nouvelle exigence, de nouvelles compétences », 2020. <https://www.dataacc.org/bonnes-pratiques/adopter-un-plan-de-gestion-des-donnees/gestion-des-donnees-une-nouvelle-exigence-de-nouvelles-competences/>.

Chronologie inspirée de : Smale, Nicholas, et al. « The History, Advocacy and Efficacy of Data Management Plans ». BioRxiv, octobre 2018. www.biorxiv.org, [doi: 10.1101/443499](https://doi.org/10.1101/443499).

Qu'est-ce qu'un plan de gestion des données (DMP) ?

Un **document synthétique** qui aide à organiser et anticiper toutes les étapes du cycle de vie de la donnée. Il explique pour chaque jeu de données **comment seront gérées les données d'un projet**, depuis leur création ou collecte jusqu'à leur partage et leur archivage.

Source : INIST-CNRS, doranum.fr

N. B. : Un DMP peut devenir un *Data Paper* : un article décrivant un jeu de données publié dans un *Data Journal*

Le DMP en bref

- Comment la gestion et le partage des données sont-ils financés, en particulier à long terme ?

Ressources

- En quoi consiste le projet ?
- Qui sont les partenaires ?
- Quelle est la politique de gestion des données ?
- Qui est responsable de la gestion des données ?

Responsabilités dans le projet

- Quelles données seront produites/utilisées au cours du projet ? (type, format, volume et accroissement...)?
- Comment seront-elles produites ou transformées ?

Collecte des données

- Comment les données seront-elles identifiées, décrites ?
- Quels standards de métadonnées utilisera-t-on ?
- Comment seront générées les métadonnées ?

Documentation des données

- Qui pourra accéder aux données ?
- Les données seront-elles publiées ?
- Comment ?
- Dans quel délai ?
- Sous quelle licence ?

Accès et partage des données

- Comment, où, par qui, seront stockées, sauvegardées et sécurisées les données ?

Sauvegarde des données

- Qui sera propriétaire des données produites ?
- Des données externes seront-elles utilisées ?

Propriété intellectuelle

- Des données sensibles seront-elles produites ou utilisées ?
- Comment sera assurée leur anonymisation ?

Ethique

- Quel est le plan d'archivage et de préservation à long terme ?

Archivage et préservation des données

Modèles de DMP par financeur

Quelques exemples

- Des modèles de DMP existent en fonction des financeurs :
 - Programmes **européens** : Horizon Europe, H2020, ERC
 - Financeurs **français** : ANR, INCa
 - **Etablissements** : Institut Pasteur
- Le DMP est un **document évolutif**, demandé, en général, à **trois** moments du projet.
- **Evaluation** du DMP par les financeurs :
 - **ANR** : livrable, non évalué
 - **Europe** : livrable, évalué

Les principaux champs des modèles de DMP



DMP : **Horizon Europe**

- 1. Data summary
- 2. FAIR data
 - 2.1. Making data *findable*, including provisions for metadata
 - 2.2. Making data openly *accessible*
 - 2.3. Making data *interoperable*
 - 2.4. Increase data *re-use* (through clarifying licences)
- 3. Other research outputs
- 4. Allocation of resources
- 5. Data security
- 6. Ethics
- 7. Other issues

https://ec.europa.eu/info/funding-tenders/opportunities/docs/2021-2027/horizon/guidance/programme-guide_horizon_en.pdf



DMP : **ANR**

Modèle
de Plan de gestion
des données (PGD)

Informations générales

1. Description des données et collecte ou réutilisation des données
2. Documentation et qualité des données
3. Stockage et sauvegarde pendant le processus de recherche
4. Exigences légales et éthiques, codes de conduite
5. Partage des données et conservation à long terme
6. Responsabilités et ressources en matière de gestion de données

https://anr.fr/fileadmin/documents/2019/ANR_Modele_de_DMP_francais_DMPOPIDoR_2019_07_24_mis_en_page_2.pdf

Un comparatif des deux modèles

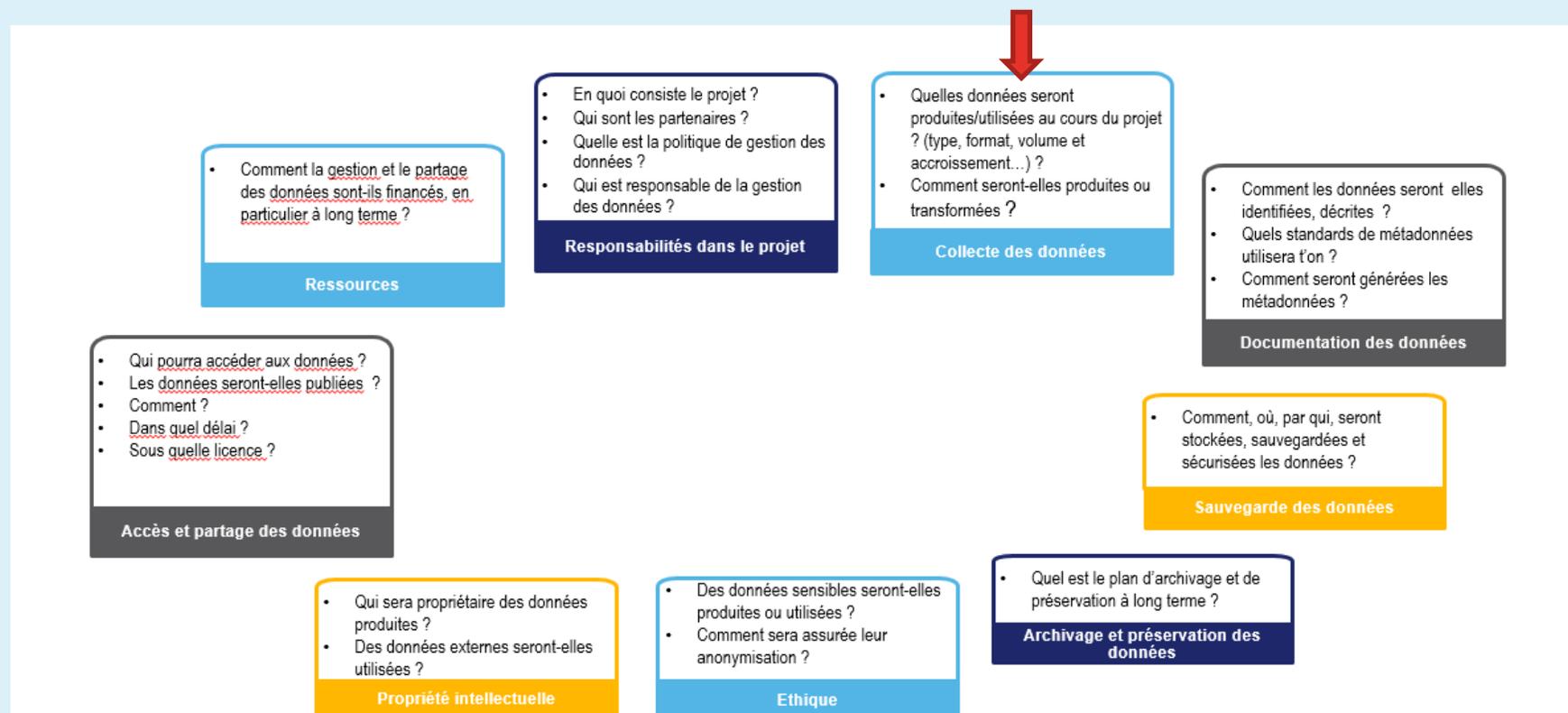
A retrouver sur [DropSU](#)

A	B	C	D	E	F	G	H
Horizon Europe			Concordance		ANR		
Will rich metadata be provided to allow discovery? What metadata will be created? What disciplinary or general standards will be followed? In case metadata standards do not exist in your discipline, please outline what type of metadata will be created and how. F2	HE 11 métadonnées	entrepôt	HE16, HE17, ANR59		format	ANR 11	<ul style="list-style-type: none"> • Privilégier les formats standards et ouverts car ils facilitent le partage et la réutilisation à long terme des données (plusieurs catalogues fournissent des listes de ces "formats préférés"). • Donner des détails sur les volumes (qui peuvent être exprimés en espace de stockage requis (octets), et/ou en quantités d'objets, de fichiers, de lignes, et colonnes).
Will search keywords be provided in the metadata to optimize the possibility for discovery and then potential re-use?	HE 12 métadonnées	éthique	ANR46, ANR47, HE53		volume	ANR 12	
Will metadata be offered in such a way that it can be harvested and indexed? F3	HE 13 métadonnées	format	HE3, ANR9, ANR10, ANR11			ANR 13	2. DOCUMENTATION ET QUALITE DES DONNEES
2.2. Making data openly accessible	HE 14	identifiants	HE10, HE18, ANR63, ANR65			ANR 14	2a. Quelles métadonnées et quelle documentation (par exemple méthodologie de collecte et mode d'organisation des données) accompagneront les données ? F4
Repository:	HE 15	licences	HE36, ANR42		métadonnées	ANR 15	<ul style="list-style-type: none"> • Indiquer quelles métadonnées seront fournies pour aider à la recherche et à l'identification des données.
Will the data be deposited in a trusted repository?	HE 16 entrepôt	logiciels/codes	HE42, HE43, ANR3, ANR60		métadonnées	ANR 16	<ul style="list-style-type: none"> • Indiquer quels standards de métadonnées seront utilisés (par exemple DDI, TEI, EML, MARC, CMDI).
Have you explored appropriate arrangements with the identified repository where your data will be deposited?	HE 17 entrepôt	métadonnées	HE11, HE12, HE13, HE27, ANR15, ANR16, ANR17		métadonnées	ANR 17	<ul style="list-style-type: none"> • Utiliser les standards de métadonnées des communautés scientifiques lorsque ceux-ci existent.
Does the repository ensure that the data is assigned an identifier? Will the repository resolve the identifier to a digital object? A1	HE 18 identifiants	méthodo	HE56, ANR71		nommage	ANR 18	<ul style="list-style-type: none"> • Indiquer comment les données seront organisées au cours du projet, en mentionnant par exemple les conventions de nommage, le contrôle de version et les structures des dossiers. Des données bien classées et gérées de façon cohérente seront plus faciles à retrouver, à comprendre et à réutiliser. • Penser à la documentation qui serait nécessaire pour permettre une réutilisation des données. Il peut s'agir notamment de l'information sur la méthodologie utilisée pour collecter les données, sur les procédures et méthodes d'analyse utilisées, sur la définition des variables, des unités de mesure, etc.
Data:	HE 19	nommage	ANR18		documentation	ANR 19	
Will all data be made openly available? If certain datasets cannot be shared (or need to be shared under restricted access conditions), explain why, clearly separating legal and contractual reasons from intentional restrictions. Note that in multi-beneficiary projects it is also possible for specific beneficiaries to keep their data closed if opening their data goes against their legitimate interests or other constraints as per the Grant Agreement.	HE 20 partage	objectif	HE4		documentation	ANR 20	<ul style="list-style-type: none"> • Tenir compte de la façon dont ces informations seront obtenues et enregistrées par exemple dans une base de données avec des liens vers chacun des fichiers, dans un fichier texte de type « lisez-moi », dans les en-têtes de fichiers, dans un livre de référence (« code book ») ou dans les cahiers de laboratoire.
If an embargo is applied to give time to publish or seek protection of the intellectual property (e.g. patents), specify why and how long this will apply, bearing in mind that research data should be made available as soon as possible.	HE 21 embargo	partage	HE20, HE51, ANR50, ANR51		qualité	ANR 21	2b. Quelles mesures de contrôle de la qualité des données seront mises en œuvre ? I2, R1, R1.2
Will the data be accessible through a free and standardized access protocol? A1.1	HE 22	propriété	ANR41, ANR44		qualité	ANR 22	<ul style="list-style-type: none"> • Expliquer comment la qualité et la conformité de la collecte des données seront contrôlées et documentées. Il s'agit là de préciser les processus comme la calibration, la répétition des échantillons ou des mesures, la capture standardisée des données, la validation de saisie des données, la revue par les pairs, ou la représentation basée sur des vocabulaires contrôlés.
If there are restrictions on use, how will access be provided to the data, both during and after the end of the project? A1.2	HE 23 restrictions	pseudonymisation	ANR37			ANR 23	3. STOCKAGE ET SAUVEGARDE PENDANT LE PROCESSUS DE RECHERCHE
How will the identity of the person accessing the data be ascertained?	HE 24 accès	qualité	HE39, HE40, ANR21, ANR22		stockage/sauvegarde	ANR 24	3a. Comment les données et les métadonnées seront-elles stockées et sauvegardées tout au long du processus de recherche ?
Is there a need for a data access committee (e.g. to evaluate/approve access requests to personal/sensitive data)?	HE 25 accès	responsabilité	HE47, ANR67, ANR68, ANR69, ANR70		stockage/sauvegarde	ANR 25	<ul style="list-style-type: none"> • Décrire l'endroit où les données seront stockées et sauvegardées au cours du processus de recherche et la fréquence à laquelle la sauvegarde sera effectuée. Il est recommandé de stocker les données dans au moins deux lieux distincts. • Privilégier l'utilisation de systèmes de stockage robustes, avec sauvegarde automatique, tels que ceux fournis par les services informatiques de l'institution d'origine. Le stockage des données sur des ordinateurs portables, des disques durs externes, ou des périphériques de stockage tels que des clés USB n'est pas recommandé.
Metadata:	HE 26	restrictions	ANR4, ANR45, ANR53, ANR56, HE23		stockage/sauvegarde	ANR 26	

4

RÉDIGER LE DMP, ÉLÉMENTS ESSENTIELS ↴

DESCRIPTION DES DONNEES COLLECTEES OU RÉUTILISEES





Description et collecte

Exemple – 1



BIBLIOTHÈQUE
UNIVERSITAIRE

Cavalier, Jean-François. "LipInTB Plan de Gestion de Données," 2019. <https://dmp.opidor.fr/plans/4624/export.pdf>



1. Description des données et collecte ou réutilisation de données existantes

Les données recueillies dans le cadre de l'étude proposée comprendront la synthèse de composés chimiques (protocoles et caractérisation) et leur utilisation en tant que sonde moléculaire. Chaque composé synthétisé sera purifié et entièrement caractérisé pour confirmer leur structure et leur pureté chimique à l'aide de méthodes spectroscopiques et chromatographiques. Les données spectroscopiques typiques comprennent la RMN 1H, la RMN 13C, la spectrométrie IR et la spectrométrie de masse ; les techniques chromatographiques comprennent la chromatographie en phase liquide couplée à la spectrométrie de masse (HPLC-MS).

Les composés seront ensuite utilisés en tant que sonde moléculaire pour étudier le métabolisme des lipides chez *M. tb*. Les données préexistantes et déjà publiées seront également utilisées pour la validation des protéines cibles de nos composés. Une fois les protéines validées, elles seront cristallisées et leurs structures 3-D respectives seront déterminées par la méthode de diffraction au rayons-X.

Les données seront collectées par chaque scientifique impliqué dans le projet à l'aide de logiciel adapté à chaque instrument. Les protocoles expérimentaux sont conservés sous forme papier (dans les cahiers de laboratoire) et version électroniques (.docx, .pdf). Les données RMN seront stockées sous forme électronique (fichiers FID bruts et format PDF) et en laboratoire sous forme de copies papier des spectres. Les fichiers FID bruts collectés sur les appareils Bruker peuvent être traités par n'importe quel logiciel de RMN, et seront disponibles sur demande. Les fichiers PDF contiendront à la fois des images des spectres originaux et des tableaux numériques détaillant le déplacement chimique, les constantes de couplage, les valeurs de masse et autres interprétations numériques des données.

Les données biochimiques générées seront : des données expérimentales (formats csv, xlsx), des données textuelles (format pdf, docx, pptx), des données de protéomique, des données cristallographiques, des images (microscopie confocale et électronique - format png). Les données protéomiques seront de plusieurs types, les fichiers bruts en sortie des spectromètres de masse (.raw), les données traitées (.txt, .xml, .mztab) et les résultats (.txt, .xlsx, .docx et images). Le volume des données protéomiques qui seront générés pendant ce projet est estimé à plusieurs Go.

Les données de diffraction au rayons X seront stockées sous forme électronique (format .cbf et .h5) à la fois par les sites de rayonnement synchrotron (Soleil et ESRF), et au sein du laboratoire du partenaires du projet. Le volume des données de diffraction qui seront générées pendant ce projet est estimé à plusieurs To. Selon les directives H2020 les sites de rayonnement synchrotron doivent donner accès sur demande aux données de diffraction trois ans après leur acquisition. Les fichiers avec les coordonnées atomique des structure 3-D des protéines seront ensuite déposés à la Protein Data Bank (<https://www.rcsb.org/>) au format PDB.

Dans la mesure du possible des formats standards et ouverts seront privilégiés à des fins de partage et de réutilisation.





Description et collecte

Exemple – 2



BIBLIOTHÈQUE
UNIVERSITAIRE

Lumley, Emily. "CompBioMed D3.1_Data Management Plan_v1.0," 2020, 12. https://www.compbiomed.eu/wp-content/uploads/2017/03/D1.3_DataManagementPlan_CBK_v1.3.pdf



5 Data Summary

Provide a summary of the data addressing the following issues:

- State the purpose of the data collection/generation
- Explain the relation to the objectives of the project
- Specify the types and formats of data generated/collected
- Specify if existing data is being re-used (if any)
- Specify the origin of the data
- State the expected size of the data (if known)
- Outline the data utility: to whom will it be useful

CompBioMed2 is the second phase of the Computational Biomedicine Centre of Excellence (CoE), CompBioMed, an outward facing CoE comprising members from academia, industry and the healthcare sector. Its core purpose is to facilitate the uptake and exploitation of HPC-based Computational Biomedicine simulation approaches to a greater number of therapeutic areas.

Understanding the complex outputs of such simulations requires the convergence of High Performance Computing (HPC) and high-performance data analytics (HPDA). CompBioMed2 seeks to combine these approaches with the large, heterogeneous datasets from medical records and from the experimental laboratory to underpin clinical decision support systems. All data collected, used and generated by the project is done in support of this objective

The project's three core research strands focus on the areas of cardiovascular, musculoskeletal and molecular modelling. Each of these scientific communities and associated software packages employ different data formats. Major types of data comprise:

- **Imaging data** which stem mostly from clinical trials, experiments, or visualize results of the simulations. They may serve as initial data for data analytics and machine learning tasks. Typical file formats used for these data are JPG, PNG, DICOM, MP4, and MOV.
- **Other clinical and experimental data** which serve as reference for simulations. Here a broad variety of file formats are used from formatted/unformatted plain text, PDF and DOCX files, tabulated data formats like CSV and XLSX, as well as (raw) binary data.
- **Musculoskeletal data** which record the motion of joints, bones, muscles, etc. They are the key output of musculoskeletal simulations and use file formats such as C3D and XMDF.
- **Cardiovascular data** resulting from heart and blood flow simulations in the project are mostly recorded in HDF5 and VTK file formats.
- **Molecular modelling data:** Structures of complex biomolecules, assemblies thereof, smaller molecules and molecular dynamics trajectories serve as initial conditions or are output of simulations conducted in the project and may be targets of HPA approaches. Most frequently used file formats are PDB, PSF, XTC, TRR. A variety of tools is available that are able to read and convert these different formats.





Description et collecte

Exemple – 3



BIBLIOTHÈQUE
UNIVERSITAIRE

Aventurier, Pascal. « Bridge Research through Interoperable Data Governance and Environments : data management plan », 2020.

<https://dmp.opidor.fr/plans/5954/export.pdf>



WP3 IT TOOLS : Recommandations pratiques et techniques pour améliorer la FAIRness des entrepôts

Les documents suivants seront rédigés à partir de guides et documents techniques. Des tests seront effectués sur les applications "use case" et les entrepôts de données des trois partenaires.

WP3-IT Tools	Rapport de FAIRness des entrepôts IRD, CIRAD et INRA	Création
WP3-IT Tools	Recommandations pratiques pour améliorer la FAIRness des entrepôts	Création

WP3-IT Tools	Rapport de FAIRness des entrepôts IRD, CIRAD et INRA	Textes (Latex + ODT + PDF) 100Mo max
WP3-IT Tools	Recommandations pratiques pour améliorer la FAIRness des entrepôts	Textes (Latex + ODT + PDF) 100 Mo max

WP3 autocomplétion : Widget d'autocomplétion pour améliorer les métadonnées

Développement d'un widget d'autocomplétion pour améliorer les métadonnées, en langage *Javascript*.

WP3-IT Tools	Widget d'autocomplétion pour améliorer les métadonnées	Création de code Javascript
--------------	--	-----------------------------

Création du code par le projet pour un volume estimé de moins de 2 Go.



Documenter la collecte et la réutilisation

- Un guide :

Arnould, Pierre-Yves, et Marie-Christine Jacquemot-Perbal. 2017 « Guide de bonnes pratiques: gestion et valorisation des données de recherche ». OTELO.

<https://ordar.otelo.univ-lorraine.fr/record?id=10.24396/ORAR-1>



GUIDE DE BONNES PRATIQUES

Gestion et valorisation des données de recherche

ARNOUD Pierre-Yves (OTeLo), JACQUEMOT-PERBAL Marie-Christine (Inist-CNRS)
V1.1
01/02/2016

Relecteurs : AUCLERC Apolline (OTeLo – LSE), BEGUIRISTAIN Thierry (OTeLo-LIEC),
LEGUÉDOIS Sophie (OTeLo – LSE), MONTAGERS-PELLETIER Emmanuelle (OTeLo – LIEC),
RIPAMONTI-CHENOT Elodie-Denise (OTeLo – LSE)



Cette œuvre est mise à disposition selon les termes de la
[Licence Creative Commons Attribution - Pas d'Utilisation Commerciale - Partage dans les Mêmes Conditions 4.0 International](https://creativecommons.org/licenses/by-nc-sa/4.0/).

Conventions de nommage

Exemples



A naming convention for photographs (WP 3 and task 2.2) has being established. All levels of information are separated by “_” (underscore symbol). It develops as follows:

- a 3-letter code defines the name of the site, followed by a 2-letters code of the site locality. In case of same names, the second site will use the 4th letter of its name in place of the 3rd. This to avoid exact same place-names. (e.g. MAR-CR for Martinšćica in Cres)
- year of acquisition of the photo

then, for site photos :

- o sector (S01 to S99)
- o increment 0001 to 9999

or, for finds and site facts:

- o IND (burial) or PO (small object) or LAP (lapidary) or MON (coin)
- o Increment 0001 to 9999

Ex.: MAR_2019_S03_0001 (photo du secteur 3)
 MAR_2019_S03_0002 (mosaïque LIDAR ou photogrammétrie)
 MAR_2019_PO_0001 (objet 1 de l'année 2019)
 MAR_2019_IND_0002 (2e sépulture de l'année 2019)



Over the course of data collection, a clear versioning system will be used consisting of file naming conventions (ProjectDescription-Date yyyy-mm-dd), together with standard headers listing creation dates and version numbers. Only final versions will be deposited as part of the archived datasets.



Marchand, Julie. « MONACORALE - MONAsteriorum CORpus Adriaticorum et Locorum Ecclesiasticorum v1(june2021) : Plan de gestion de données », 2021. <https://dmp.opidor.fr/plans/9691/export.pdf>.

Doran, Michelle. “D1.2 ORDP: KPLEX - Open Research Data Pilot – 2018-01-31.” Research Report. Trinity College Dublin, March 2018.

<https://hal.archives-ouvertes.fr/hal-01842371>.

Nommer les fichiers

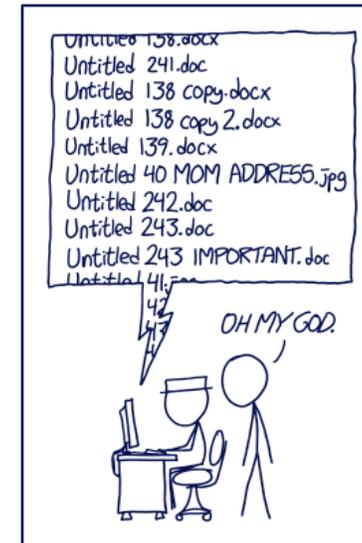


Classement thématique

- par sujet :
sujet_type_date_version.extension
reunion_CR_20200227_V01.docx
- par typologie de document :
type_sujet_date_version.extension
CR_reunion_20200227_V01.docx

Classement chronologique

- par sujet :
date_sujet_type_version.extension
20200227_reunion_CR_V0.1.docx
- par type de document :
date_type_sujet_version.extension
20200227_CR_reunion_V0.1.docx



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

<https://xkcd.com/1459/>



Un exemple d'organisation de jeux de données de projet

Colomb, Julien, Thorsten Arendt, Deepti Mittal, et Keisuke Sehara. 2020. « Folder Structure Template for Research Repositories ». <https://doi.org/10.5281/zenodo.4410128>

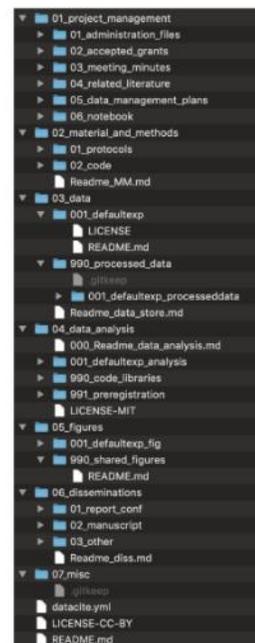
Towards a Standardized Research Folder Structure – GenR. 2021. <https://genr.eu/wp/towards-a-standardized-research-folder-structure/>

Experiment level:

add sub-folders for each experiment

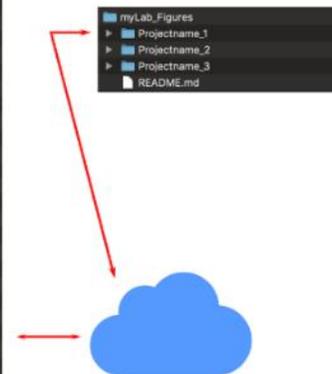


Project level



Laboratory level:

mirror sub-folders in other structures



- template_par
 - ._LICENSE-CC-BY
 - 01_project_management
 - ._05_data_management_plans
 - ._06_notebook
 - 05_data_management_plans
 - ._DMP_main.txt
 - 06_notebook
 - ._gitkeep
 - 04_data_analysis
 - ._LICENSE-MIT
- template_par
 - 01_project_management
 - 01_administration_files
 - ._gitkeep
 - 02_accepted_grants
 - ._gitkeep
 - 03_meeting_minutes
 - ._gitkeep
 - 04_related_literature
 - ._gitkeep
 - 05_data_management_plans
 - DMP_main.txt
 - 06_notebook
 - ._gitkeep
 - 02_material_and_methods
 - 01_protocols
 - ._gitkeep
 - 02_code
 - ._gitkeep
 - Readme_MM.md
 - 03_data
 - 001_defaultexp
 - LICENSE
 - README.md
 - 990_processed_data
 - ._gitkeep
 - 001_defaultexp_processeddata
 - ._gitkeep
 - Readme_data_store.md
 - 04_data_analysis
 - 000_Readme_data_analysis.md
 - 001_defaultexp_analysis
 - ._gitkeep
 - 990_code_libraries
 - ._gitkeep
 - 991_preregistration
 - ._gitkeep
 - LICENSE-MIT
 - 05_figures
 - 001_defaultexp_fig
 - ._gitkeep
 - 990_shared_figures
 - Readme_sharedfig.md
 - 06_disseminations
 - 01_report_conf
 - Readme_conf.md



Formats de fichiers

Formats ouverts, pour quoi faire ?

► Pour que d'autres utilisateurs puissent lire mes documents sans entrave

Des utilisateurs s'échangent un rapport.



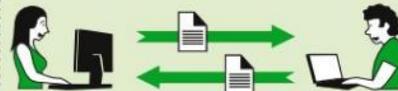
FORMAT FERMÉ, LOGICIELS IDENTIQUES

Alice utilise le logiciel « Carcera⁽¹⁾ ». Elle enregistre son rapport dans un format fermé (ne permettant pas l'interopérabilité), puis l'envoie à Bob qui possède le même logiciel. Celui-ci peut lire le document, le modifier et le renvoyer à Alice.



FORMATS FERMÉS, LOGICIELS DIFFÉRENTS

Le lendemain, Alice envoie son rapport à Albert. Il ne possède pas le même logiciel et celui-ci refuse de lire le document. Albert n'a pas d'autre solution que d'acquiescer le logiciel Carcera utilisé par Alice, en espérant qu'il soit adapté à son ordinateur.



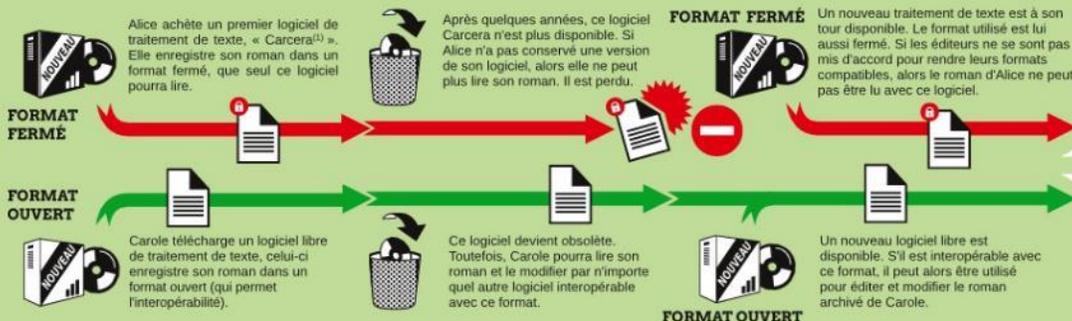
FORMATS OUVERTS, LOGICIELS DIFFÉRENTS

Carole, une autre utilisatrice, choisit d'enregistrer son rapport dans un format ouvert (permettant l'interopérabilité) et l'envoie à David. David peut lire le document, le modifier et l'enregistrer à son tour, soit en utilisant le même logiciel, soit en utilisant un autre logiciel interopérable.

Pour que vos documents puissent être lus facilement par d'autres personnes, sans avoir besoin de vous soucier du logiciel qu'elles utilisent, choisissez des formats ouverts.

► Pour assurer la pérennité de mes documents dans le temps

Alice et Carole utilisent des logiciels de traitement de texte pour rédiger, enregistrer et conserver un roman.



La disponibilité et la pérennité de vos documents, enregistrés dans un format fermé, dépendent des décisions des éditeurs de logiciels.

Dans votre intérêt, choisissez des logiciels qui enregistrent vos documents dans des formats ouverts. Ils ne sont pas dépendants d'un logiciel en particulier.

April Pour en savoir plus, rendez-vous sur : www.april.org. Document réalisé par l'April avec des logiciels libres. Graphisme : Antoine Bardelli. Licence Art Libre 1.3 ou ultérieure / Creative Commons Paternité, Partage à l'Identique 2.0 ou ultérieure / GFDL 1.3 ou ultérieure. (1) Nom d'un logiciel propriétaire fictive pour les besoins de la démonstration.

Privilégiez les formats ouverts :

Format de fichier déconseillé

- Excel (.xls, .xlsx)
- Word (.doc, .docx)
- Powerpoint (.ppt, .pptx)
- Photoshop (.psd)
- Quicktime (.mov)

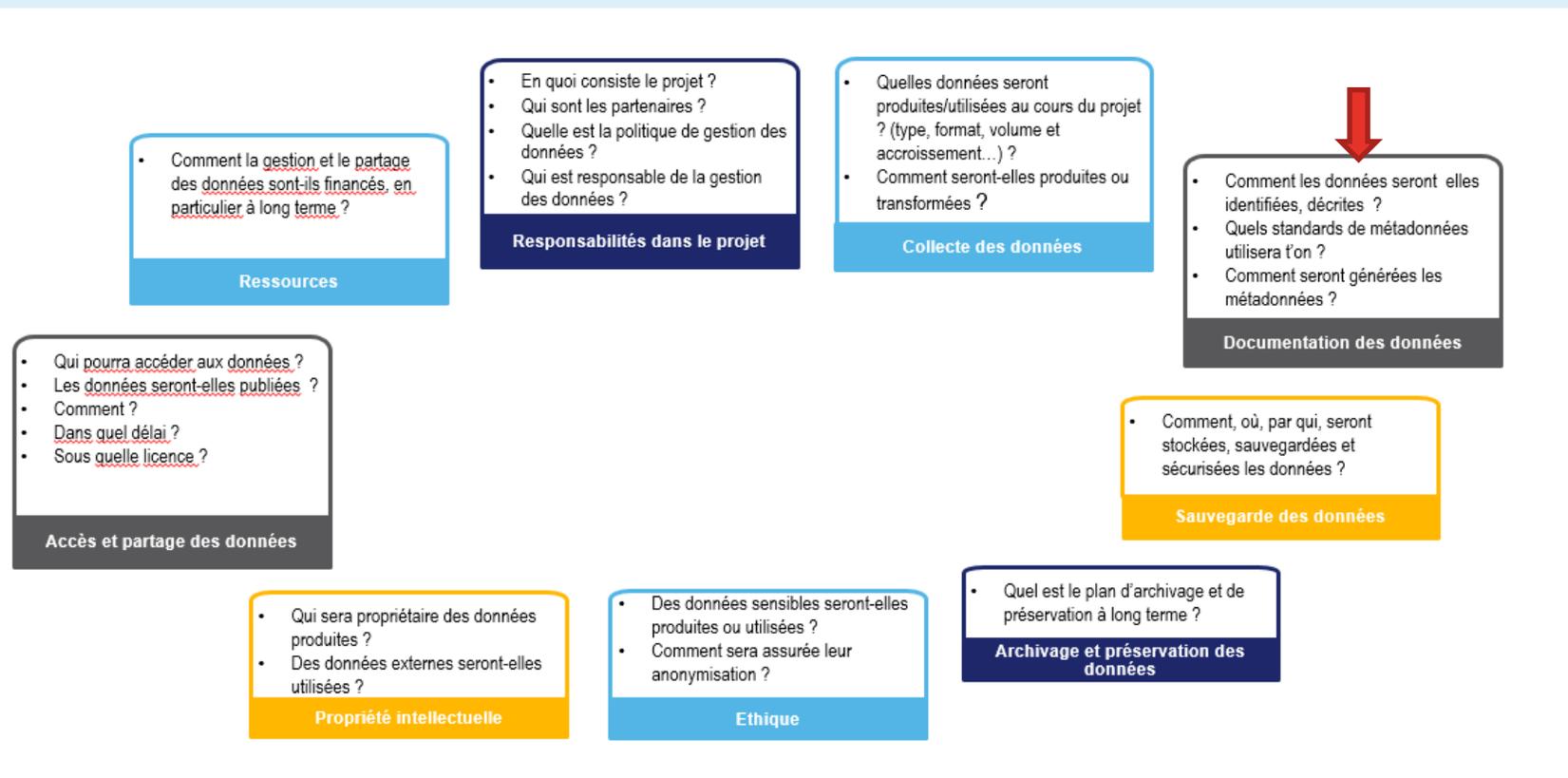
Format de fichier à privilégier

- Comma Separated Values (.csv)
- Texte ascii (.txt) ou PDF/A si formatage
- PDF/A (.pdf)
- TIFF (.tif, .tiff)
- MPEG-4 (.mp4)

Si un **format propriétaire** est utilisé par toute la communauté disciplinaire, conservez-le et précisez-le dans le DMP.

Rivet, Alain, Marie-Laure Bachellerie, Auriane Denis-Meyere, et Delphine Tisserand. 2018. « Traçabilité des activités de recherche et gestion des connaissances : Guide pratique de mise en place ». http://qualite-en-recherche.cnrs.fr/IMG/pdf/guide_tracabilite_activites_recherche_gestion_connaissances.pdf

DOCUMENTATION ET QUALITÉ





Documentation et qualité

Exemple – 1



BIBLIOTHÈQUE
UNIVERSITAIRE

2a. Quelles métadonnées et quelle documentation (par exemple méthodologie de collecte et mode d'organisation des données) accompagneront les données ?



Chaque fichier de données sera accompagné au minimum d'un fichier texte (.txt) de type " *Read_me.txt*" indiquant les métadonnées telles que la liste des noms (header) des variables enregistrées, avec pour chaque variable (header), le type d'appareillage utilisé pour la prise de mesure, les conditions dans lesquelles les données ont été collectées, l'unité de la variable mesurée ou toutes autres informations permettant une réutilisation des données. La langue utilisée sera l'anglais pour une plus grande réutilisation potentielle des données collectées. Le standard de métadonnées envisagé dans le cadre du projet est le standard "*Ecological Metadata Language*" ([EML](#)) avec une implémentation possible sous le logiciel [R](#) grâce au [package EML](#).

La convention de nommage des fichiers ou dossiers de données sera celle recommandée sur le site [DoRANum](#) en utilisant si possible la date, au format "AAAAMMJJ", de la dernière version du fichier de données, sans utiliser d'accents ou de caractères spéciaux mais le symbole "_" ou "*underscore*" en tant que séparateur de mots dans le nom du fichier et enfin en affichant l'élément important en premier dans le nom (p.ex. "20200218_DMP_IMPRINT.docx"). Dans le cas de documents pouvant être amenés à évoluer, nous préciserons la version dans le nom du fichier (p.ex. "20200218_DMP_IMPRINT_v01.docx").

Une documentation détaillée des données utilisées ainsi que des analyses effectuées à partir des données sera également produite au format [RMarkdown](#) (.rmd) à partir du logiciel libre [RStudio](#). A partir du code brut (.rmd), plusieurs formats de sorties sont possibles (.html, .pdf, .docx) pour documenter les données en terme de : (i) méthodologie utilisée pour acquérir ou analyser la donnée ; (ii) nature des variables utilisées ; (iii) d'unités de mesure utilisées ; ou bien (iv) de distribution de la donnée. Des sorties graphiques permettant une meilleure visualisation des données et de leurs distributions peuvent être intégrées directement dans le document final, avec ou sans le code brute utilisé pour générer ces sorties graphiques.



IMPRINT / Jonathan Lenoir. 2020.
<https://dmp.opidor.fr/plans/5082/export.pdf>



Documentation et qualité

Exemple – 2



BIBLIOTHÈQUE
UNIVERSITAIRE



2b. Quelles mesures de contrôle de la qualité des données seront mises en œuvre ?

Les Bonnes Pratiques de Laboratoire seront suivies pour le contrôle et la qualité des données.

Au niveau des données de spectrométrie de masse, la qualité et la conformité des données seront contrôlées par les divers contrôles qualités et les calibrations des appareils réalisés régulièrement. De plus des répliques biologiques et techniques seront réalisés pour des études de protéomique quantitative.

Au niveau de la synthèse de nouvelles molécules, les spectromètres RMN sont calibrés et révisés par le Laboratoire de Mesure Physique de l'IBMM.

Concernant les mesures de diffraction aux rayons X, les lignes de lumières sont calibrées au quotidien par les responsables de ligne. Le traitement des données à la volée, fournissant un grand nombre d'indicateurs de qualité, permet également de contrôler la qualité des données.



LipInTB / Jean-François Cavalier. 2021.
<https://dmp.opidor.fr/plans/4624/export.pdf>

Métadonnées



- Un standard générique, le **dublin core** : 15 champs
 - associé à des **standards disciplinaires** : **DDI** (SHS), **EML** (écologie), **DwC** (Darwin Core, biodiversité)
 - souvent transparent lors du dépôt, champs à remplir sur les **formulaire**s des entrepôts.
- Voir : DCC. **Disciplinary metadata**.
<http://www.dcc.ac.uk/resources/metadata-standards>
- Un outil pour **générer des métadonnées** : https://doranum.fr/wp-content/uploads/datacite_metadata_generator_4.0.html

Dublin core : 15 champs à remplir

1. Titre (métadonnée)
2. Créateur (métadonnée)
3. Sujet (métadonnée) ou mots clé
4. Description (métadonnée)
5. Éditeur
6. Contributeur
7. Date (métadonnée)
8. Type
9. Format
10. Identifiant de la ressource
11. Source
12. Langue (métadonnée)
13. Relation (métadonnée)
14. Couverture (métadonnée)
15. Gestion de droits (métadonnée)

Métadonnées

Codes et logiciels



Projet CodeMeta

- **Fichier de métadonnées en json**
- **Un générateur en ligne**

CodeMeta generator

Most fields are optional. Mandatory fields will be highlighted when generating Codemeta.

The software itself	Discoverability and citation
Name <input type="text" value="My Software"/> <small>the software title</small>	Unique identifier <input type="text" value="10.151.xxxxx"/> <small>such as ISBNs, GTIN codes, UUIDs etc. http://schema.org/identifier</small>
Description <input type="text" value="My Software computes ephemerides and orbit propagation. It has been developed from early '80."/>	Application category <input type="text" value="Astronomy"/>
Creation date <input type="text" value="YYYY-MM-DD"/>	Keywords <input type="text" value="ephemerides, orbit, astronomy"/>
First release date <input type="text" value="YYYY-MM-DD"/>	Funding <input type="text" value="PRA_2018_73"/> <small>grant funding software development</small>
License <input type="text"/> <small>from SPDX licence list</small>	Funder <input type="text" value="Università di Pisa"/> <small>organization funding software development</small>
	Authors and contributors can be added below

Mots-clés, vocabulaires



- Indispensable pour **décrire** le jeu de données et augmenter sa **visibilité** dans les moteurs de recherche
- Exemples : **Agrovoc** (agriculture), **CIDOC CRM** (patrimoine culturel), **AsCoPain-T** (agriculture), **SNOMED** (biomédical), etc.
- Rechercher des **vocabulaires** :
 - BARTOC <https://bartoc.org/>
 - FAIRsharing : <https://fairsharing.org/standards/>

Keyword

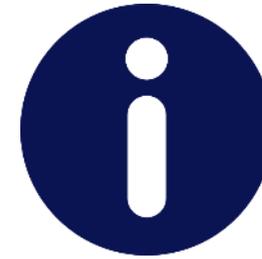
Term

Vocabulary



Vocabulary URL

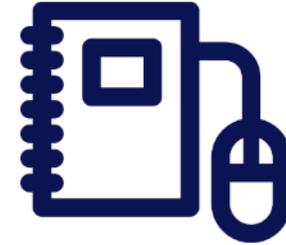
Le fichier Readme



- Documente le dépôt
- Présente les **règles de nommage** et d'**organisation** du jeu de données,
- Décrit le **contenu**,
- Précise les **logiciels** ou codes informatiques nécessaires pour l'utilisation

- Trames de fichiers Readme :
 - 4TU. **Guidelines for creating a README file.**
https://researchdata.4tu.nl/fileadmin/user_upload/Documenten/Guidelines_for_creating_a_README_file.pdf
 - Doranum. **Gabarit « Readme ».** https://doranum.fr/wp-content/uploads/gabarit_readme.txt

Le cahier de labo



- *Est mentionné par certains modèles de DMP, mais n'est pas à partager*

A conserver !

- Le cahier de labo est mentionné dans la plupart des modèles de DMP, dont celui de l'ANR
- AAF. 2012. « Référentiel de gestion des archives de la recherche » <https://www.archivistes.org/Referentiel-de-gestion-des-archives-de-la-recherche>

Le contrôle qualité

- Garantir la traçabilité
- Un guide :

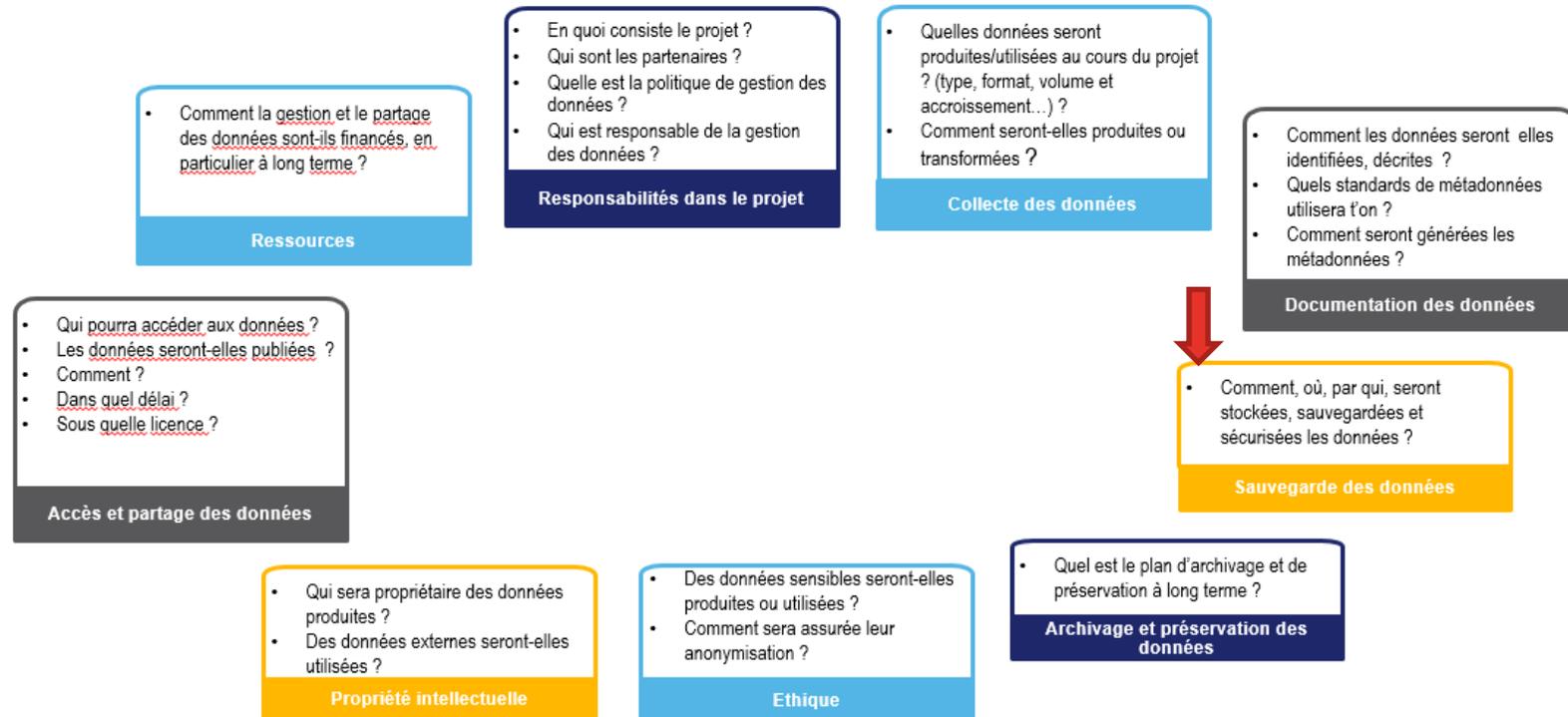
Rivet, Alain, Marie-Laure Bachèlerie, Auriane Denis-Meyere, et Delphine Tisserand. 2018.

« Traçabilité des activités de recherche et gestion des connaissances : Guide pratique de mise en place ».

<https://qualite-en-recherche.cnrs.fr/gt/tracabilite-des-activites-de-recherche/>



STOCKAGE & SAUVEGARDE SÉCURITÉ



Sauvegarde Exemples



“

Une sauvegarde automatique est effectuée 2 fois par jour à midi et à minuit avec une rétention de 99 jours. Un certificat est installé sur le serveur , le certificat active le cadenas et le protocole « https » dans les navigateurs, afin d'assurer une connexion sécurisée entre le serveur web et le navigateur.

”

Brau, Frédéric. « Base de données d'images en sciences de la vie d'Université Côte d'Azur : data management plan », 2020. http://unice.fr/plateformes/mica/ressources/base-de-donnees-omero-1/Base_de_donnes_dimages_en_sciences_de_la_vie_dUniversit_Cte_dAzur.pdf

“

Pour la durée du projet ANR les données générées seront hébergés sur les postes de travail chiffrés des membres du projet LipInTB, et pour les données de protéomique et de cristallographie sur des serveurs dédiés et sécurisés. Des sauvegardes sur un serveur propre à chaque unité seront effectuées à intervalles réguliers. Les données de diffraction aux rayons X seront également sauvegardées par les sites de rayonnement synchrotron.

”

Cavalier, Jean-François. “LipInTB Plan de Gestion de Données,” 2019. <https://dmp.opidor.fr/plans/4624/export.pdf>.

Sécurité des données

Exemple : accès aux données



Gestion des groupes et comptes

La base de données comporte :

-d'une part des groupes privés nominatifs par équipe de recherche auxquels sont associés des comptes individuels nominatifs. Chaque équipe aura donc un groupe contenant les comptes des membres de l'équipe et un groupe de collaboration pouvant contenir les comptes des membres de l'équipe et des comptes de collaborateurs extérieurs au périmètre (***) de la base. A leur demande 1 à 2 groupes de collaborations supplémentaires pourront être créés. A l'intérieur de ces groupes les différents comptes inclus auront accès à l'ensemble des images et données du groupe. Pour accéder aux comptes une authentification individuelle est nécessaire.

Les référents sont regroupés dans un groupe dédié. Ils créent/suppriment les groupes, les comptes utilisateurs.

Les personnels des services d'imagerie de la plateforme *** posséderont des comptes et seront ajoutés aux comptes collaborations des équipes en fonction de leur soutien à ces équipes. Il n'y aura pas de groupe dédié à ces services.

- d'autre part un groupe public sans restriction d'accès avec des restrictions d'écriture. Ce groupe pourra accueillir des images libres de droit ou sous licence Creative Commons, sans faire office de plateforme de publication.

Gestion des mots de passe

Les référents de chaque institut doivent choisir un mot de passe « robuste », c'est-à-dire composé d'au moins dix caractères alphanumériques, d'une majuscule et un ou plusieurs caractères spéciaux. Si besoin, l'utilisateur devra modifier le mot de passe lors de la première connexion avec un niveau de sécurité au moins équivalent.

La base de données *** présente actuellement une méthode d'authentification avec un niveau de sécurité faible. En effet, l'utilisateur a la possibilité de choisir un mot de passe sans aucune restriction. Dans ce contexte, afin de renforcer la sécurité de la base de données, nous envisageons de nous tourner, si c'est possible, vers une fédération d'identité s'appuyant sur les technologies SSO (« Single Sign-on » désignant les technologies qui permettent aux utilisateurs de mutualiser la phase d'authentification entre plusieurs services). Ceci aussi pour faciliter le travail des utilisateurs sur plusieurs systèmes, tout en réduisant la charge administrative liée à la gestion de l'accès à ces systèmes.

Temporalité des comptes

- Les comptes des utilisateurs contractuels ayant quitté les équipes (doctorants, stagiaires, post-doctorants) seront désactivés.

Préalablement les images seront détruites ou réaffectées au chef d'équipe du groupe concerné et resteront accessibles pour ce groupe.

Cette mise à jour manuelle des comptes sera effectuée deux fois par an, en septembre et janvier. Les administrateurs de la base enverront un message aux référents qui feront remonter les départs d'utilisateurs de leurs laboratoire. En cas de changement de propriétaire des images un formulaire en ligne est disponible pour spécifier le compte de départ, d'arrivée et les projets concernés. Ceci pour permettre à l'administrateur d'effectuer l'ensemble des changements de propriétaires répertoriées en une seule fois.

Les comptes des utilisateurs supprimés seront dénommés "A_utiliser_numero", désactivés et affectés au groupe "User desactives"

- Dans le futur, une procédure de curation automatique des comptes est envisagée suivi d'une vérification annuelle. Les référents seront informés de la liste des comptes inusités et valideront le changement de propriétaire des images, voire leur suppression.



Stocker pendant le projet



- **Conserver plusieurs copies de ses données sur différents lieux de stockage. Règle de la sauvegarde 3-2-1 : 3 copies sur 2 supports différents, avec au moins 1 une copie à distance**
- **Privilégier des systèmes de stockage robustes, avec sauvegarde automatique, tels que ceux fournis par les services informatiques de l'institution d'origine**
- **Gérer les accès aux données : accès sécurisé pour les partenaires, vérification des accès en fin de projet ou en cas de départ d'une personne**
- **Chiffrer les données si elles ont un caractère sensible**
- **Utiliser un cloud sécurisé pour les accès distants**
 - **DropSU**
 - **ODS MyCore au CNRS et à l'INSERM**

Stocker pendant le projet



Unité de service mutualisée : **SACADO** Service d'Aide au Calcul et à l'Analyse de Données

Conseil et ingénierie

Architecture d'équipement de calcul / d'analyse de données ; analyse de performance, optimisation et développement ; déploiement d'application ; archivage et gestion de données

Accès à la plateforme technologique MeSU

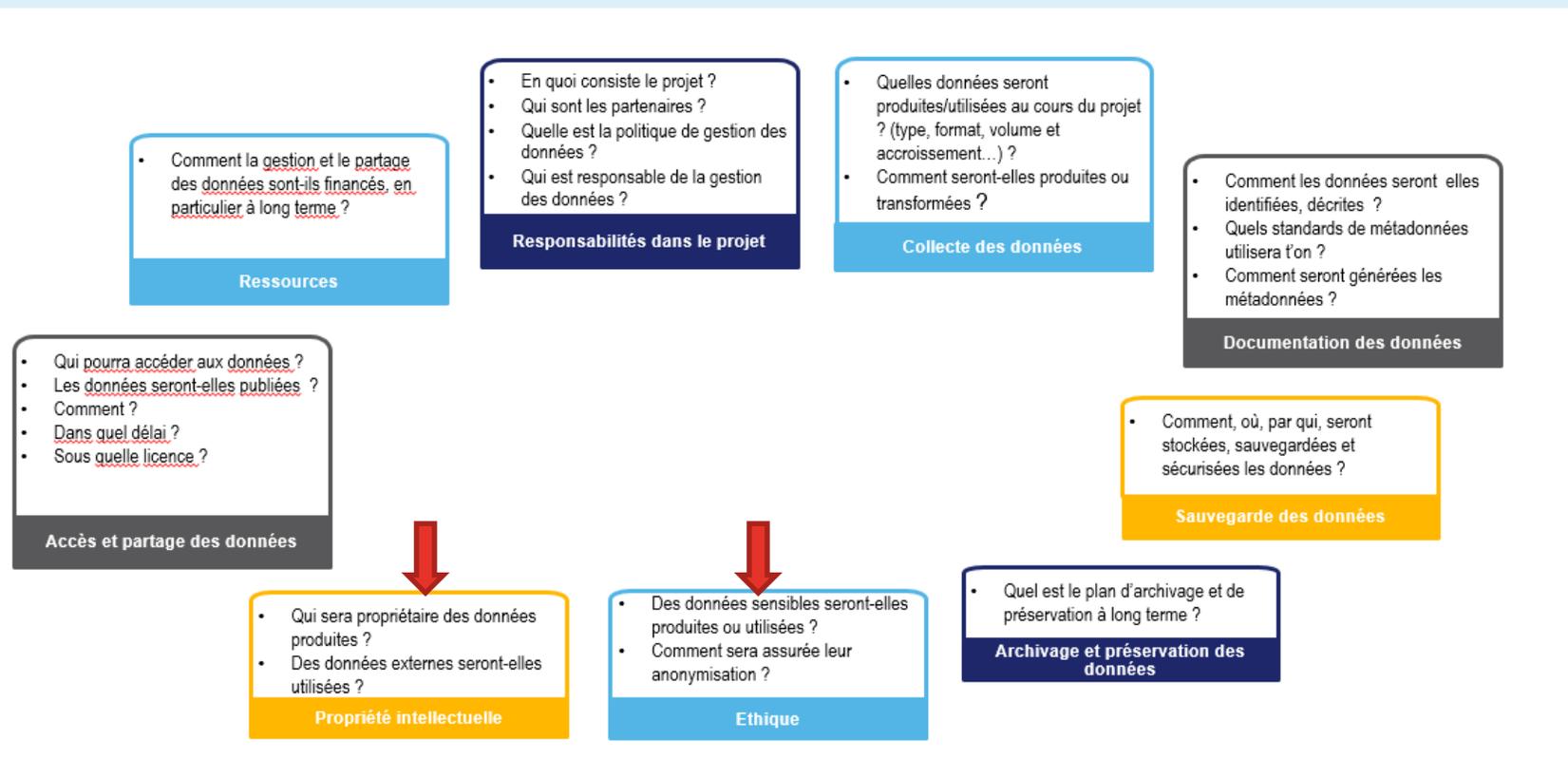
Création de compte utilisateur ; compilation/installation d'application ; planification d'une campagne de calcul ; hébergement d'une application scientifique ; accueil pour une session de visualisation

Accompagnement technique

Portage ; compilation d'application ; assistance à la soumission de travaux

Contact : **Nicolas Benoit** nicolas.benoit@sorbonne-universite.fr

ETHIQUE, RGPD, PROPRIÉTÉ DES DONNÉES





Ouvrir les données

Cadre français

Open data : principe de communication et de réutilisation libre et gratuite

- En **Europe** : Directive PSI
- En **France** :
 - Lois CADA, Valter et loi pour une République numérique
 - Code des relations entre le public et l'administration

En pratique, tout dépend du type de données :

- **Communication obligatoire**
 - Directive INSPIRE : géographie
 - Convention d'Aarhus : environnement

Source :
Saby, Mathieu. Gérer et diffuser ses données en
SHS. <https://osf.io/zj9as/> (CC: BY-NC)



Ouvrir les données

Restrictions

- **Communication interdite des données portant atteinte :**
 - A la défense et la politique étrangère de la France, aux délibérations du gouvernement, au pouvoir exécutif
 - A la sûreté de l'État, la sécurité publique, à la sécurité des personnes , à la sécurité des biens de l'établissement et de ses systèmes d'informations
 - Au secret professionnel (médical, etc.)
 - Au secret industriel et commercial
- **Communication possible sous conditions de données :**
 - Protégées par le droit d'auteur ou autres droits de propriété intellectuelle
 - Relatives à des personnes privées (données personnelles, vie privée)
 - Soumises au secret statistique
 - Liées à un contrat avec un tiers non soumis à une obligation de service public
 - Présentant des risques pour la protection du potentiel scientifique et technique de la nation (« unité protégée » ou « Zone à régime restrictif »)

Ethique Exemple



5. Ethical aspects

All ethical aspects connected to data collection and data sharing are overseen by an external ethics expert, namely the Chair of the Trinity College Dublin Faculty of Arts, Humanities and Social Sciences Ethics Committee (as nominated by Faculty Dean Darryl Jones).

The Ethics Committee of the Faculty of Arts Humanities and Social Science within Trinity College Dublin is comprised of a Chair (nominated annually by the Dean of Faculty of Arts, Humanities and Social Sciences) and 5 Faculty Representatives – normally Directors of Research – with one of the members always to be from the School of Law.

The members of this Committee are tasked: to read and review relevant applications, with a view to identifying potential ethical issues and thereby granting or refusing consent; to detail any concerns regarding the application from an ethical perspective; to review any notification of adverse effect; to review any monitoring reports submitted to the Committee; to review any end of project reports submitted to the Committee; and to keep up to date with legislative / best practice guidelines in the Ethics arena. More information about this Committee is available at:

<https://ahss.tcd.ie/Faculty%20Ethics%20Committee/FacultyEthicsCommittee.php>

A copy of the Trinity College Dublin, Faculty of Arts, Humanities and Social Science Research Ethics Committee decision can be found in Annex I.



Le comité d'éthique de la recherche de Sorbonne Université

<https://cer.sorbonne-universite.fr/>

Le CER est compétent pour émettre un **avis éthique consultatif sur des protocoles de recherche non interventionnelle impliquant des sujets humains**, dès lors qu'ils ne relèvent pas des compétences d'un Comité de Protection des Personnes (CPP) agréé par le Ministère de la Santé.

Le CER de Sorbonne Université émet un **avis au plan éthique** sur les protocoles de recherche qui lui sont soumis au regard, notamment :

- de la **protection des personnes**,
- de la **balance bénéfice/risque** y compris au plan psychologique,
- dans les **règles d'ordre public en la matière**, notamment la non-discrimination, sauf si la nécessité de la recherche l'impose objectivement.

Les protocoles doivent être soumis au CER **avant** le début du projet.



Intégrité, déontologie

- Chartes de déontologie ou d'intégrité de l'établissement de la discipline
- En France, « Charte nationale de déontologie des métiers de la recherche ». 2015. https://cache.media.enseignementsup-recherche.gouv.fr/file/Enseignement_superieur/47/6/charte_nationale_deontologie_metiers_de_la_R_915476.pdf

Signée par le CNRS, l'INSERM et la CPU, en cours à Sorbonne Université

- En Europe, ALLEA. 2017. « The European Code of Conduct for Research Integrity ». 2017. <https://allea.org/wp-content/uploads/2017/05/ALLEA-European-Code-of-Conduct-for-Research-Integrity-2017.pdf>

Données personnelles

Exemple



2.2 Confidentiality and archiving

Based upon these national regulation and the European legislation (EU Directive 95/46/EC and the forthcoming GDPR, see note 1), but also in the push towards open data and the guidelines generated by the EU-funded project OPENAire, the EU commission has generated the Open Data Pilot in which InSPIRES project is included. Being part of this pilot implies designing a protocol for assuring that all the data generated by the project are open and sharable but also protected by the different data protection regulations. Following this insight, InSPIRES data management protocol will consist, in general terms, of two phases:

- (i) Confidentiality that will assure ethical and legal treatment of data. In this phase, all researchers must obtain from the subjects of study their informed consent prior performing the research itself, being it an interview, questionnaire, focal group, etc. A model of this informed consent can be checked in Annex I. Once the research is performed, researchers must anonymize the outputs, i.e. in interview transcripts should never be mentioned any personal data that might led to the identification of the interviewee.
- (ii) archiving that will assure that the data, once confidential, can be obtained by anyone willing to use them for secondary or contrasting research. InSPIRES will archive all the data generated in Zenodo, an open and public repository created by OPENAire and owned by the European Commission. Whenever possible, in addition to Zenodo, consortium members might upload project data in local and public repositories belonging to their institutions.



Obligations liées au RGPD



Université Paris Lumières. 2019. « Fiches pratiques sur le Règlement Général pour la Protection des Données ».

<https://www.u-plum.fr/wp-content/uploads/2019/09/Guide-RGPD-2019-web.pdf>

Sciences humaines

InSHS. 2019. « Les sciences humaines et sociales et la protection des données à caractère personnel dans le contexte de la science ouverte : guide pour la recherche ».

https://inshs.cnrs.fr/sites/institut_inshs/files/pdf/guide-rgpd.pdf

Santé

CNIL. 2018. « Recherche médicale : quel est le cadre légal ? »

2018. <https://www.cnil.fr/en/node/24981>

Travailler avec son DPD

A Sorbonne Université, documentation sur l'intranet :

- Un [guide](#),
- Des exemples de fiches de registre d'activité par faculté.

Contact de la DPD :
dpd@sorbonne-universite.fr

UMR :
- Le DPD a été choisi par la/le DU, soit au CNRS/INSERM, soit à l'université.

Fiche de registre d'activité

MISE EN ŒUVRE D'UN TRAITEMENT PORTANT SUR DES DONNÉES PERSONNELLES	
Nom du traitement	_____
Date de création Date de révision de la fiche	_____
Service chargé de la mise en œuvre du traitement et nom du responsable	_____
Coordonnées du sous-traitant ou du service informatique chargé de la sécurité des données	_____
Population concernée (étudiants, personnels...)	_____
Finalité principale: décrire les objectifs poursuivis et l'usage pour le service	_____
Énumérer chacune des tâches automatisées par le traitement: collecte, transfert, anonymisation, traitement comparatif, statistiques...	<ul style="list-style-type: none">● _____● _____● _____● _____● _____
Information des personnes concernées ou recueil du consentement éclairé (conserver les justificatifs)	<input type="checkbox"/> par mention sur le questionnaire de collecte <input type="checkbox"/> par affichage <input type="checkbox"/> par la remise d'un document <input type="checkbox"/> par une mention sur le site internet <input type="checkbox"/> par envoi de courrier <input type="checkbox"/> par intranet <input type="checkbox"/> Autres: _____
Échanges de données	• Existe-t-il un échange de données avec d'autres fichiers? OUI <input type="checkbox"/> NON <input type="checkbox"/> Si oui, le(s)quel(s): _____ • Existe-t-il des transferts d'informations hors de l'Union européenne? OUI <input type="checkbox"/> NON <input type="checkbox"/>

Envoyer à dpd@sorbonne-universite.fr

Anonymisation des données



Deux exemples d'outil :

- **ARX**
<https://arx.deidentifier.org/>
- **Amnesia**
<https://amnesia.openaire.eu/>



PERSONAL DATA		ANONYMOUS DATA
Fully identifiable data	Pseudonymous data	Data which cannot be traced back to an individual person. Data collected from human subjects, either directly or when re-used for secondary research, is rarely ever anonymous.
Any data from which a person could immediately be identified with little to no effort (e.g. names, faces, unique characteristics).	Data from which a person cannot be immediately identified, however it remains possible to identify a person from the data.	
EXAMPLE Patient number 90210 City Leeuwarden Date of birth 27-4-1967 Income 7.861 Job Judge Car DeLorean License Plate SN-09-HN	Study subject 47110009 Region Friesland Year of birth 1967 Income 7.500-10.000 Job Legal Car DeLorean	Country Netherlands Age 51-60 Income 5.000 – 15.000 Job Legal Car sportwagen

Ouvrir les données : propriété intellectuelle



- Les données sont-elles protégées par le droit d'auteur ?
 - **Oui** pour les « œuvres » même sans caractère artistique, si elles présentent un degré minimal d'« originalité »: textes, discours, photos, vidéos, musique, cartes, sculptures...
 - **Oui** pour le code informatique
 - **Non** pour des informations purement factuelles (mesures, comptages...)
 - **Non** pour les œuvres tombées dans le domaine public (pour être précis il reste un droit moral, inaliénable)

Le fait que la collecte des données ait demandé un **investissement** (humain, matériel, financier) n'est pas **créateur de droit d'auteur**.

- Autres droits de propriété intellectuelle :
 - Marques, dessins, modèles
 - Brevets, éléments brevetés
 - Droits des bases de données

Interlocuteur à l'université : Bureau de la propriété intellectuelle

Sources :

Saby, Mathieu. Gérer et diffuser ses données en SHS. <https://osf.io/zi9as/> (CC: BY-NC)

Dantant, Martin. Droit d'auteur des chercheurs, Logiciels, Bases de Données et Archives Ouvertes. http://isidora.cnrs.fr/IMG/pdf/2014-07-07_-

[Droit d'auteur des chercheurs Logiciels Bases de Données et Archives Ouvertes - Grenoble ssc.pdf](#)

Propriété intellectuelle

A qui appartiennent les résultats ?



A votre employeur, dans la plupart des cas :

- Ce qui est produit dans le cadre de nos mission est considéré comme un document administratif (au sens de la loi Valter).
- *Attention !* Pour les doctorants : si la thèse est cofinancée ou réalisée en collaboration avec un partenaire privé => il faut se reporter au contrat.
- Cas particulier : je suis **chercheur ou enseignant-chercheur** : mes œuvres (écrits, cartes, photographies, plans, etc.) qui sont originales et donc soumises au droit d'auteur m'appartiennent (exception – loi DADVSI 2006-961). **Mais le reste appartient bien à mon employeur.**

Licences



Plusieurs types :

- Creative commons
(international)

- Licence ouverte
(France)

CREATIVE COMMONS LICENSES		COPY & PUBLISH	ATTRIBUTION REQUIRED	COMMERCIAL USE	MODIFY & ADAPT	CHANGE LICENSE
	PUBLIC DOMAIN	✓	✗	✓	✓	✓
	CC BY	✓	✓	✓	✓	✓
	CC BY-SA	✓	✓	✓	✓	✗
	CC BY-ND	✓	✓	✓	✗	✗
	CC BY-NC	✓	✓	✗	✓	✓
	CC BY-NC-SA	✓	✓	✗	✓	✗
	CC BY-NC-ND	✓	✓	✗	✗	✗

Legend:

- ✓ You can redistribute (copy, publish, display, communicate, etc.)
- ✗ You have to attribute the original work
- ✓ You can use the work commercially
- ✓ You can modify and adapt the original work
- ✓ You can choose license type for your adaptations of the work.



LICENCE OUVERTE
OPEN LICENCE

Licences : logiciels et bases de données



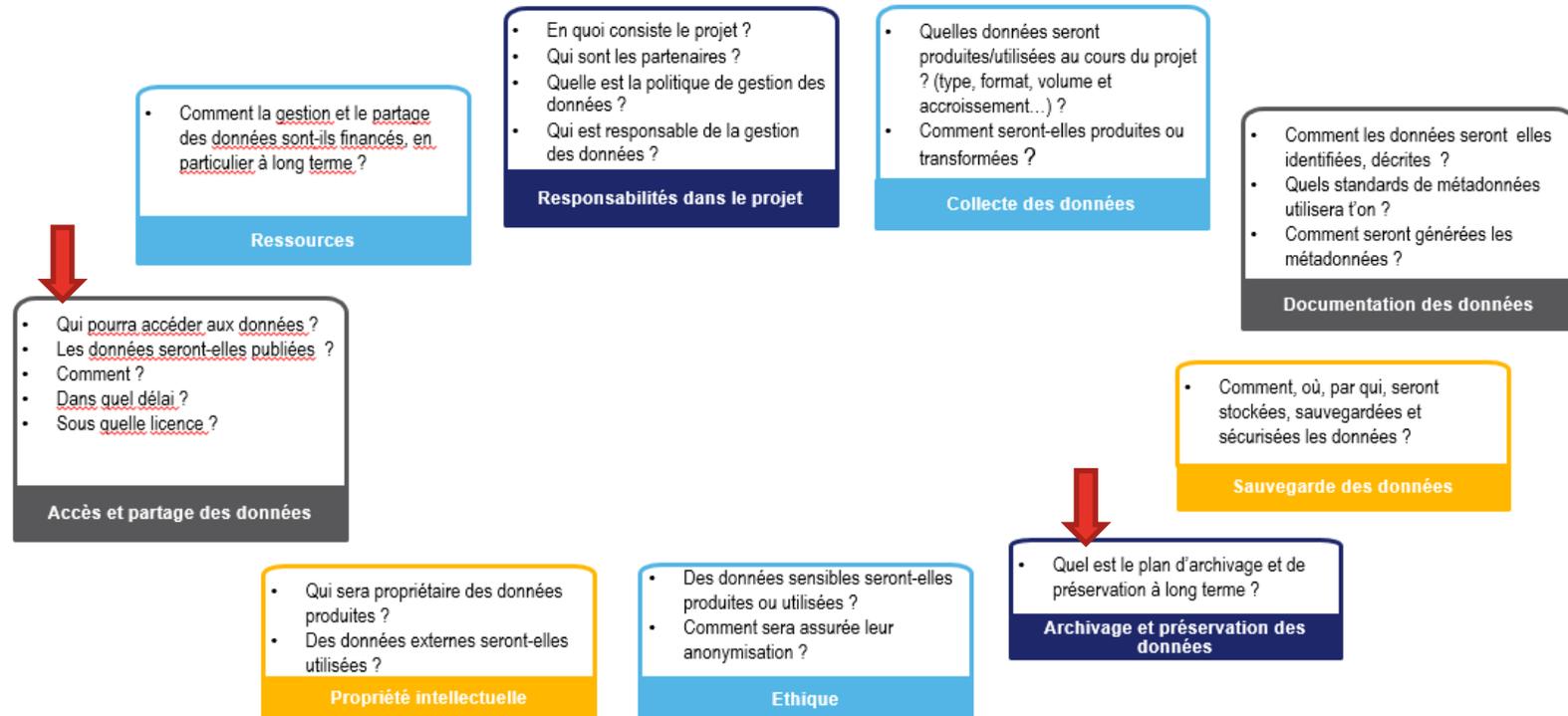
- Bases de données : Licences de l'Open Knowledge Foundation, PDDL, ODC-By, ODC-ODbL
- Logiciels : licences GNU



ODbL

Un outil pour trouver une licence publique :
<https://ufal.github.io/public-license-selector/>

PARTAGE & CONSERVATION



Partage Exemple



5. Partage des données et conservation à long terme

Lorsqu'elle seront finalisées et au plus tard à la fin du projet, les données seront diffusées sous licence CC-BY dans l'entrepôt de données DataVerse DataSuds, qui garantit une pérennisation des données sur le long terme. Un embargo ne se justifie pas. De même les codes seront distribués sous licence CeCill-C compatible GPLv3.

Les données seront déposées dans l'entrepôt de données DataVerse DataSuds. Les codes informatiques dans *GitHub* et dans Software Heritage ou éventuellement dans DataVerse DataSuds. Les documents de type analyse ou guidelines seront déposés dans les archives ouvertes de chaque partenaires, leur code source (documentation technique) archivée dans Dataverse DataSuds et sur Overleaf.

Il n'y a pas de méthode spécifique. Les vocabulaires qui seront exploités par le code sont propres à chaque organisme et décrits dans un langage standardisé (SKOS). Tout utilisateur pourra utiliser ses propres vocabulaires.

Les développements seront codés dans des langages très utilisés (*JavaScript*) et indépendants d'outils propriétaires.

Des identifiants *DOI* seront attribués par l'entrepôt DataVerse DataSuds, ou directement par l'IRD si nécessaire pour des documents spécifiques.

Pour le code informatique : identifiant attribué par *Software Heritage* et lien *GitHub*.



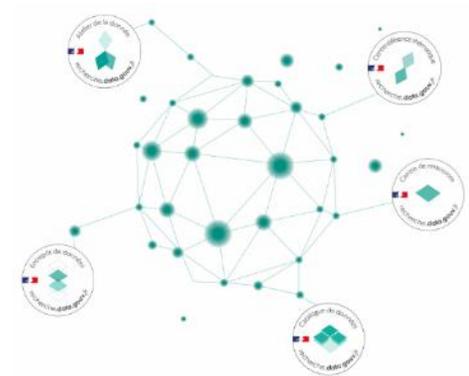
Aventurier, Pascal. « Bridge Research through Interoperable Data Governance and Environments : data management plan », 2020. <https://dmp.opidor.fr/plans/5954/export.pdf>

Des exemples d'entrepôts de données

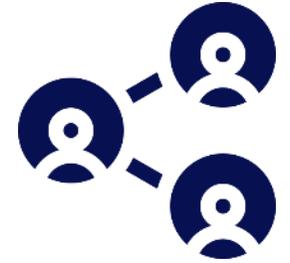


Un entrepôt de données ?

- Archives en ligne permettant le dépôt, la description, la conservation, la recherche et la diffusion de jeux de données



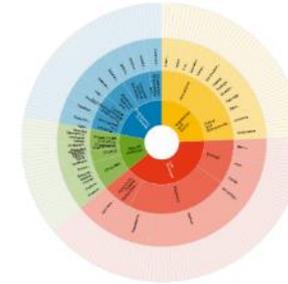
Où et comment partager ses données ?



- Entrepôt pluridisciplinaire (UE / CERN) : [Zenodo](#)
- Entrepôts privés : [Figshare](#), [GigaDB](#), [Mendeley data](#)
- Entrepôts disciplinaires : [DRYAD](#), [NCBI](#), [Nakala](#), [Gbif](#), [Seano](#)
- Entrepôts d'établissements : [Edinburgh Datashare](#), [INRAE Datapartage](#)

Rechercher des entrepôts de données

- Trouver des entrepôts par disciplines



- Entrepôts français :
CAT OPIDoR



- Rechercher des jeux de données

[Search DataCite](#)

[OpenAIRE](#)



Recherche Data Gouv, l'entrepôt national français



RÉPUBLIQUE
FRANÇAISE

*Liberté
Égalité
Fraternité*

recherche.data.gouv.fr

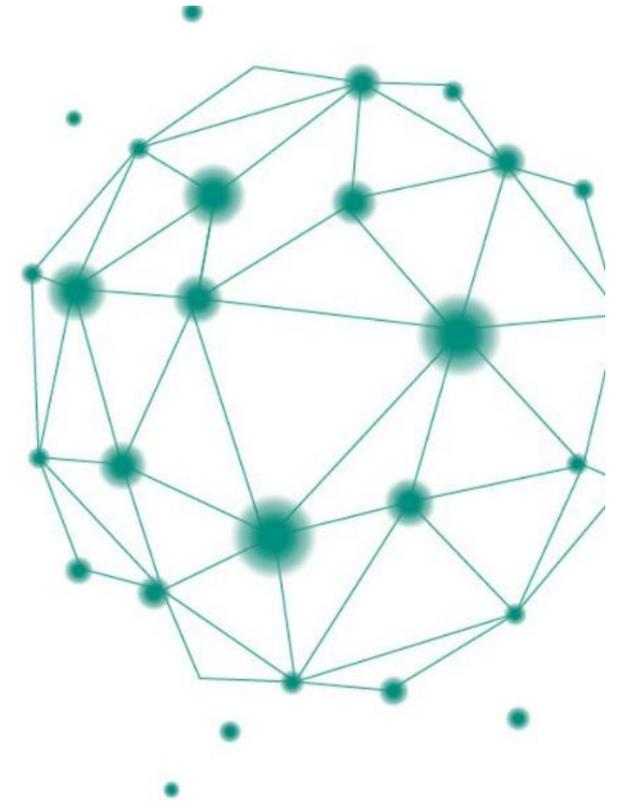
- **Entrepôt généraliste**, pour les disciplines qui n'ont pas encore d'entrepôt disciplinaire ou pour les projets transdisciplinaires
- Entrepôt dédié aux **données associées aux publications**

Site : <https://recherche.data.gouv.fr/fr>

Entrepôt : <https://entrepot.recherche.data.gouv.fr/>

Espace Sorbonne Université :

<https://entrepot.recherche.data.gouv.fr/dataverse/sorbonne-univ>



Partager logiciels et codes sources

Les forges logicielles” [GitHub](#), [GitLab.com](#), [BitBucket](#), etc.

- **Ne remplacent pas un archivage pérenne !**



Archive ouverte et pérenne, [Software Heritage](#)

- **Garantit une conservation et un accès pérennes**



Software Heritage

Mission ▾

Archive ▾

Communauté ▾

Grants

Soutien ▾

A propos ▾

News ▾

Faire un don



```
throttled = !list_empty(&cfs_b->throttled_cfs_b);
cfs_b->nr_periods += overrun;
```

```
/*
 * idle depends on !throttled (for the case of a large backlog) and if
 * we're going inactive then everything else can be deferred
 */
```

```
if (cfs_b->idle && !throttled)
    __refill_cfs_bandwidth_runtime(cfs_b);
```

```
if (!throttled) {
    /* mark as potentially idle for the upcoming period */
    cfs_b->idle = 1;
```

Les logiciels sont notre patrimoine

Archivage Exemple

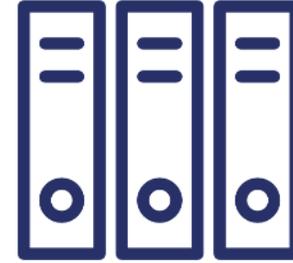


Durée de conservation des données : Cette base de donnée ayant vocation à être utilisée pour gérer des projets, elle ne constitue pas une solution de stockage pérenne et les images seront conservées pour une durée maximale de 5 ans. Les utilisateurs seront avertis par courriel pour supprimer les projets ou images dépassant la durée maximale prévue. Une prolongation pourra être accordée exceptionnellement pour une durée maximale de 1 an. Dans l'esprit FAIR les utilisateurs de la base de données seront invités à déposer les images ayant conduit à valorisation dans la base de donnée "Image Data Resource" (IDR) sous license Creative Commons (CC-BY).



Brau, Frédéric. « Base de données d'images en sciences de la vie d'Université Côte d'Azur : data management plan », 2020. [http://unice.fr/plateformes/mica/ressources/base-de-donnees-omero-1/Base de donnees dimages en sciences de la vie dUniversit Cte dAzur.pdf](http://unice.fr/plateformes/mica/ressources/base-de-donnees-omero-1/Base%20de%20donnees%20dimages%20en%20sciences%20de%20la%20vie%20dUniversit%C3%A9%20Cte%20dAzur.pdf).

Archiver les données



Dès le début du projet :

- Adopter des conventions pour nommer les dossiers et les fichiers
- Anticiper les éventuelles opérations d'anonymisation des données



Pendant le projet :

- Déterminer ce qui sera conservé ou non
- Faire le point sur la durée de conservation



Après le projet :

- Procéder au tri des jeux produits
- Prévoir l'archivage pérenne (intégrité et lisibilité des fichiers sur le long terme)

**Service des archives
et du recueil des actes de
Sorbonne Université :**
Océane Valencia
[sara-archives@sorbonne-
universite.fr](mailto:sara-archives@sorbonne-universite.fr)

Un guide :
Association des Archivistes de
France. « Référentiel de gestion
des archives de la recherche ».
[https://www.archivistes.org/Refer-
entiel-de-gestion-des-archives-de-
la-recherche](https://www.archivistes.org/Referentiel-de-gestion-des-archives-de-la-recherche)

Les identifiants pérennes



Le plus utilisé :

- **DOI** : identifiant pérenne, qui permet de **retrouver l'emplacement d'un document en ligne si son URL a changé.**

Automatiquement généré lors du dépôt sur un entrepôt de données

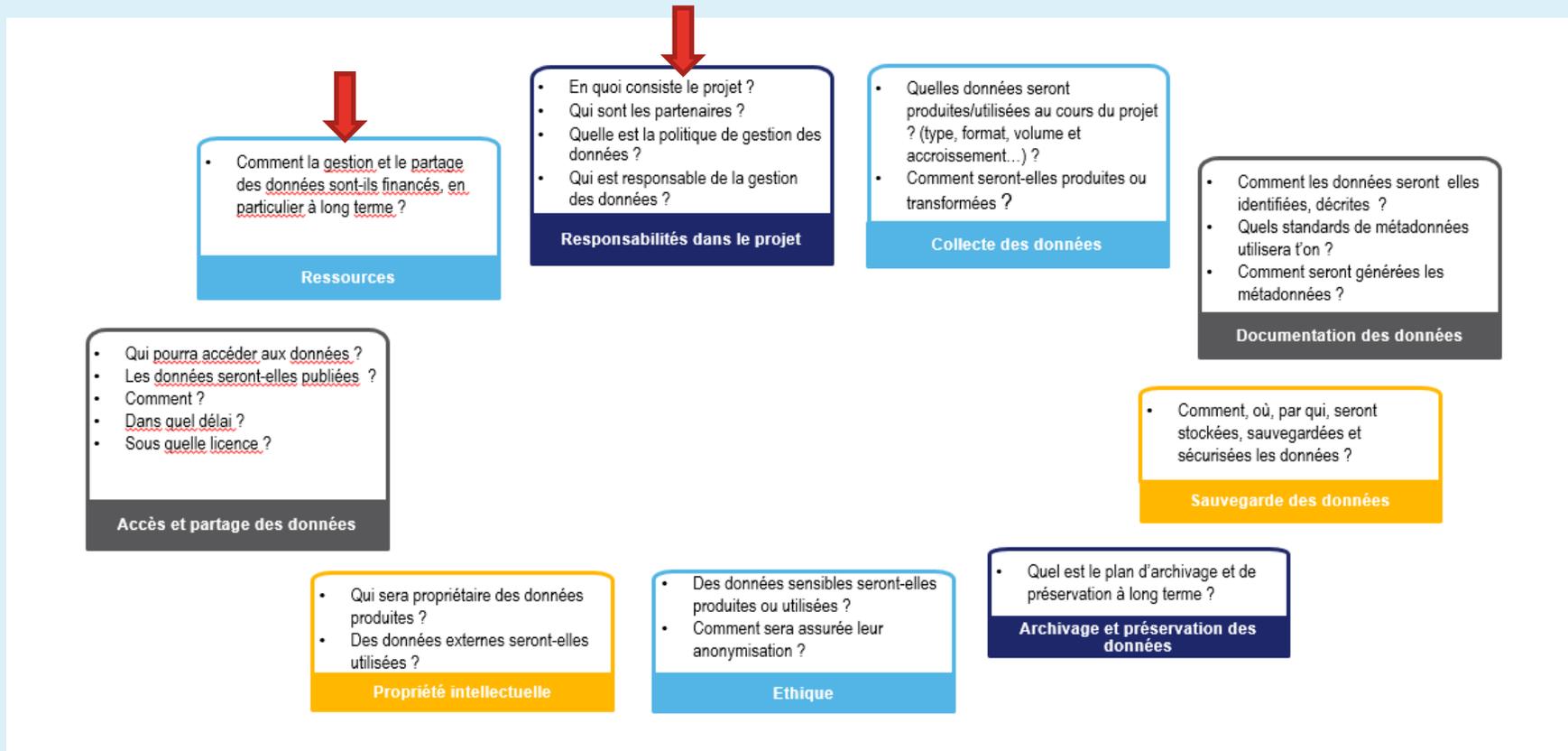
Un service de fourniture de DOI à l'INIST, le service [PID OPIDoR](#) :

- demander l'**ouverture d'un compte** auprès de DataCite ;
- **attribuer** des DOI à partir de votre compte ;
- bénéficier de services de contrôle et de traitement en nombre des DOI.

Autres identifiants pérennes :

- **Handle system** : permet d'attribuer des identifiants pérennes à des objets numériques
- **ARK** : identifiant pérenne pour des objets numériques et physiques, très utilisé par les bibliothèques.

RESPONSABILITÉS & RESSOURCES



La responsabilité du DMP



Le responsable du DMP :

- pas forcément la responsable du projet
- parfois une équipe de rédacteurs (répartis par WP ou type de données) et un coordinateur responsable

Le DMP est un livrable, qui évolue au fil du projet :

- V1 attendue 6 mois après la contractualisation,
- V2 à mi-projet,
- *Version supplémentaire à chaque étape importante (facultative),*
- V3 en fin de projet.

Coûts



Description des coûts de stockage et d'archivage, des RH mises sur la gestion des données

Outils et checklists pour chiffrer les coûts de stockage et d'archivage :

- **RDM costs. OpenAIRE.** <https://www.openaire.eu/how-to-comply-to-h2020-mandates-rdm-costs>
- **UK Data Service - Data management costing tool and checklist**
<https://www.ukdataservice.ac.uk/media/622368/costingtool.pdf>
- **EPFL Library Cost Calculator for Data Management**
<https://costcalc.epfl.ch/>
- **Utrecht University Cost of data management**
<https://www.uu.nl/en/research/research-data-management/guides/costs-of-data-management>

5

RÉDIGER ET AUTO-ÉVALUER SON DMP ↴

Rédiger le DMP

Un outil, DMP OPIDoR : <https://dmp.opidor.fr/>

- Un outil **d'aide à la rédaction du DMP**, conçu pour travailler en équipe
- Les modèles de DMP publiés par différentes institutions (H2020, ERC, ANR) sont **pré-remplis**
- On peut gérer ses différents DMP dans son espace personnel
- On peut télécharger le DMP en PDF, docx, HTML

Une [guide](#) sur l'utilisation du modèle structuré de DMP OPIDoR



Exemples de DMP

Modèles ANR, H2020 et ERC



Des exemples de DMP rendus publics par leurs auteurs :

- **Sur DMP OPIDoR : https://dmp.opidor.fr/public_plans**
- **DMP Online : https://dmponline.dcc.ac.uk/public_plans**
- **Autres : <https://www.zotero.org/groups/2468537/bsu-data/library>**

Auto-évaluer un DMP rédigé

Deux grilles



Proposition de grille de relecture de modèles de PGD rédigés selon le modèle ANR / Justine Ancelin-Fabre, URFIST de Paris.

<https://dropsu.sorbonne-universite.fr/s/cwRCK6PKzWxiad3>

Grille d'évaluation de plans de gestion des données de la recherche ouvertes : [modèle H2020] / Justine Ancelin-Fabre, URFIST de Paris.

<https://dropsu.sorbonne-universite.fr/s/QjLAnpm6pdXznea>

6

PUBLIER UN DATA PAPER ↴

Publier un data paper

- Article décrivant un jeu de données, notamment les méthodes de recueil
- Détaille les potentiels de réutilisation de données
- Le DMP peut servir de trame
- Peut être publié dans une revue académique classique, plus souvent publié dans un data journal
- Les articles font l'objet d'un peer-reviewing

SCIENTIFIC DATA 

Data Descriptor | OPEN | Published: 10 April 2018

Long-term dataset on aquatic responses to concurrent climate change and recovery from acidification

Taylor H. Leach , Luke A. Winslow, Frank W. Acker, Jay A. Bloomfield, Charles W. Boylen, Paul A. Bukaveckas, Donald F. Charles, Robert A. Daniels, Charles T. Driscoll, Lawrence W. Eichler, Jeremy L. Farrell, Clara S. Funk, Christine A. Goodrich, Toby M. Michelena, Sandra A. Nierzwicki-Bauer, Karen M. Roy, William H. Shaw, James W. Sutherland, Mark W. Swinton, David A. Winkler & Kevin C. Rose

Scientific Data 5, Article number: 180059 (2018) | [Download Citation](#) ↓

Abstract

Concurrent regional and global environmental changes are affecting freshwater ecosystems. Decadal-scale data on lake ecosystems that can describe processes affected by these changes are important as multiple stressors often interact to alter the trajectory of key ecological phenomena in complex ways. Due to the practical challenges associated with long-term data collections, the majority of existing long-term data sets focus on only a small number of lakes or few response variables. Here we present physical, chemical, and biological data from 28 lakes in the Adirondack Mountains of northern New York State. These data span the period from 1994–2012 and harmonize multiple open and as-yet unpublished data sources. The dataset creation is reproducible and transparent; R code and all original files used to create the dataset are provided in an appendix. This dataset will be useful for examining ecological change in lakes undergoing multiple stressors.

Metadata summary

Design Type(s)	observation design • time series design • data integration objective
Measurement Type(s)	geographic location • Physical Phenomenon or Property • Inorganic Chemistry • planktonic material • weather
Technology Type(s)	digital curation • data acquisition system • Analytical Chemistry • light microscopy
Factor Type(s)	
Sample Characteristic(s)	Adirondack Park • freshwater lake biome

[Download metadata file](#) ↓

Leach TH, Winslow LA, Acker FW, Bloomfield JA, Boylen CW, Bukaveckas PA, et al. Long-term dataset on aquatic responses to concurrent climate change and recovery from acidification. Scientific Data [Internet]. 2018 [cité 28 mars 2019];5. Disponible sur: <https://www.nature.com/articles/sdata201859>

Publier un data paper

Quelques data journals

Quelques exemples de data journals :

- [Biomedical Data Journal](#) (Biomédical)
- [Journal of Open Health Data](#) (Données de santé)
- [Gigascience](#) (Sciences de la vie)
- [Scientific Data](#) (Multidisciplinaire)
- [Data in Brief](#) (Multidisciplinaire)

Susini, Vanessa, Laura Caponi, Veronica Lucia Rossi, Antonio Sanesi, Nadia Romiti, Aldo Paolicchi, et Maria Franzini. « Data about Performances of Whole and Monovalent Half-Fragments Antibodies in Immunosorbent Assays ». *Data in Brief* 35 (avril 2021): 106778. <https://doi.org/10.1016/j.dib.2021.106778>.

Data Article

Data about performances of whole and monovalent half-fragments antibodies in immunosorbent assays

Vanessa Susini ^a, Laura Caponi ^a, Veronica Lucia Rossi ^b, Antonio Sanesi ^b, Nadia Romiti ^a, Aldo Paolicchi ^a, Maria Franzini ^a

[Show more](#) ▾

[+](#) Add to Mendeley [🔗](#) Share [🗣️](#) Cite

<https://doi.org/10.1016/j.dib.2021.106778>

[Get rights and content](#)

Under a Creative Commons license

[open access](#)

Refers to Vanessa Susini, Laura Caponi, Veronica Lucia Rossi, Antonio Sanesi, Nadia Romiti, Aldo Paolicchi, Maria Franzini

[Sensitivity and reproducibility enhancement in enzyme immunosorbent assays based on half fr...](#)

Analytical Biochemistry, Volume 616, 1 March 2021, Pages 114090

[Download PDF](#)

Abstract

The data here presented are related to the research article entitled “Sensitivity and reproducibility enhancement in enzyme immunosorbent assays based on half fragment antibodies” [1] aimed to compare the performance in ELISA of whole antibodies and their corresponding monovalent half-fragments obtained by reduction. Half-fragment antibodies represent an interesting method to orient antibodies in high-sensitive immunoassays taking advantage of the free sulfhydryl groups of the hinge region [2], [3], [4] that allow their oriented binding on maleimide functionalized microplates. Data here presented describe the contribution of both chemical reduction and orientation on the antigen binding capacity of whole and half-fragments antibodies. For this purpose, monoclonal anti-

7

NOUS CONTACTER ↴

Services autour de la science ouverte à la BSU

Cellule Données de la recherche & Humanités numériques :
data-bsu@sorbonne-universite.fr

Département Publications et Open Access :
hal@sorbonne-universite.fr



Service des Archives et du Recueil des Actes :
sara-archives@sorbonne-universite.fr

Vos questions

Par chat ou de vive voix !



BIBLIOTHÈQUE
UNIVERSITAIRE



MERCI !

BIBLIOTHÈQUE DE SORBONNE UNIVERSITÉ

CÉCILE ARÈNES

CHARGÉE DE MISSION DONNÉES DE LA RECHERCHE &
HUMANITÉS NUMÉRIQUES

CECILE.ARENES@SORBONNE-UNIVERSITE.FR



BIBLIOTHÈQUE
UNIVERSITAIRE



Sauf logos et mention contraire, cette présentation est mise à disposition selon les termes de la Licence Creative Commons Attribution 2.0 France.