



BONNES PRATIQUES DE GESTION ET DE PARTAGE DES DONNÉES DE RECHERCHE : PRÉALABLE À LA RÉDACTION DU DMP

17^{ème} Ecole Qualité Inter-organismes - QuaRES

11 SEPTEMBRE 2019

CNRS – Inist/DVDR/Service Formation-DoRANum

SOMMAIRE

- Définition, cycle de vie de données, acteurs et contributeurs, DoRANum
- Principes FAIR
- Plan de Gestion des Données de recherche
- Aspects juridiques et éthiques, modalités de protection
- Création, collecte et description des données de recherche
- Métadonnées
- Identifiants pérennes
- 3 étapes de sauvegarde tout au long du projet
- Stockage
- Partage, diffusion des données, dépôt dans un entrepôt
- Archivage pérenne
- Réutilisation et valorisation des données



DÉFINITION
CYCLE DE VIE DES
DONNÉES
ACTEURS ET
CONTRIBUTEURS
DORANUM



DÉFINITION ET DIVERSITÉ DES DONNÉES DE RECHERCHE

Les données de la recherche sont « **l'ensemble des informations, spécimens et matériaux produits, recueillis et documentés par les chercheurs, et qui sont collectées et exploitées à des fins de recherche et de preuve par les chercheurs et leurs équipes** »

Définition des archivistes de la Section AURORE de l'AAF

Les données de recherche peuvent être :

- **produites**, lors de campagnes de recherche (observations, mesures...)
- **collectées** : données déjà existantes (corpus, archives...)



Plusieurs définitions des données de la recherche existent.

Les « données de la recherche » sont, pour les archivistes de la Section AURORE de l'AAF, « l'ensemble des informations, spécimens et matériaux produits, recueillis et documentés par les chercheurs, et qui sont collectées et exploitées à des fins de recherche et de preuve par les chercheurs et leurs équipes. »

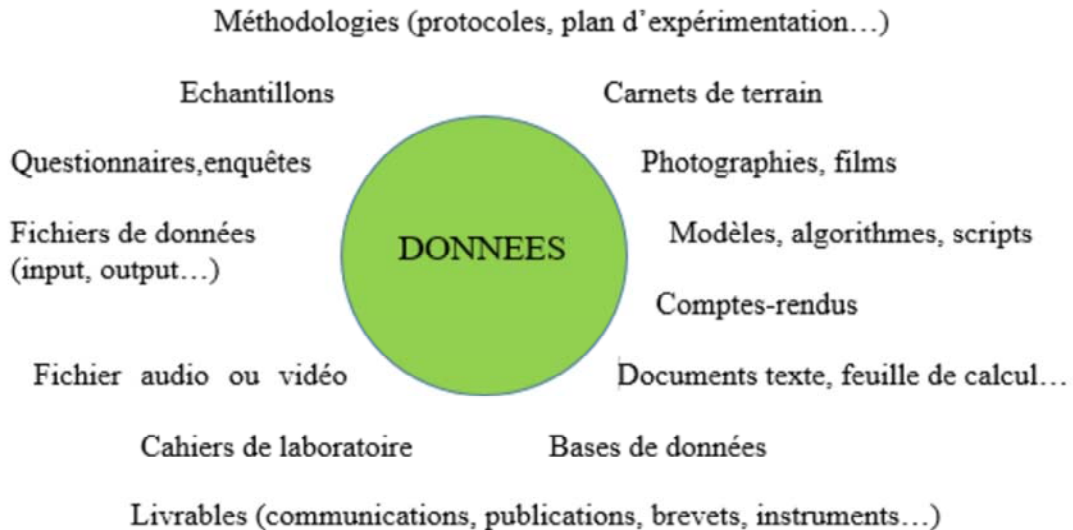
Source :

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

Wikipedia – Données de la recherche -

https://fr.wikipedia.org/wiki/Donn%C3%A9es_de_la_recherche

DÉFINITION ET DIVERSITÉ DES DONNÉES DE RECHERCHE



Selon leur contexte de création (capture ou production), leur exploitation, leur analyse et les traitements qu'elles subissent, les données de recherche peuvent être

- de différente **nature** : brutes, dérivées, formatées, nettoyées, primaires, secondaires, traitées....
- contenues dans divers **supports** : carnets de laboratoire, documents électroniques, logiciels, papier, programmes informatiques...
- de tous **types** : archives, audio, vidéo, bases de données, codes sources, géospatiales, images, photographies, langages de programmation, matérielles et physiques, modèles, visualisations, 3D, numériques, textuelles, numérisations, scans, qualitatives, quantitatives, statistiques...

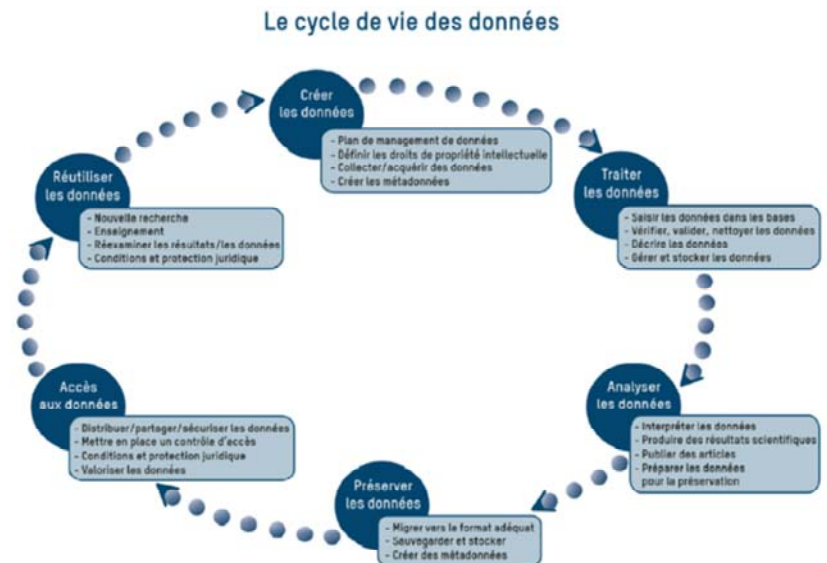
Sources :

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

UNIL – Université de Lausanne – Nature, structure et types des données de recherche - <https://uniris.unil.ch/researchdata/sujet/comprendre-gestion-donnees-recherche/donnees-de-recherche-definitions/nature-structure-types/>

CYCLE DE VIE DES DONNÉES DE RECHERCHE

Ensemble des étapes de gestion, de conservation et de diffusion des données de recherche, associées aux activités de recherche



D'après "The research data lifecycle" <http://data-archive.ac.uk/create-anode/life-cycle>

Istee - DP2VIST

Bonnes pratiques de gestion et de partage des données de recherche

16.09.19

P 6

Créer les données :

- Plan de Gestion des Données
- Définir les droits de propriété intellectuelle
- Collecter / acquérir des données
- Créer les métadonnées

Traiter les données :

- Saisir les données dans les bases
- Vérifier, valider, nettoyer les données
- Décrire les données
- Gérer et stocker les données (tout au long du projet)

Analyser les données :

- Interpréter les données
- Produire des résultats scientifiques
- Publier des articles
- Préparer les données pour la conservation

Conservier les données :

- Migrer vers le format adéquat
- Sauvegarder et stocker (pour la conservation)
- Créer des métadonnées complémentaires
- Prévoir un archivage à long terme si besoin

Donner accès aux données :

- Distribuer / partager / sécuriser les données
- Mettre en place un contrôle d'accès
- Conditions et protection juridique (licences)
- Valoriser les données

Réutiliser les données :

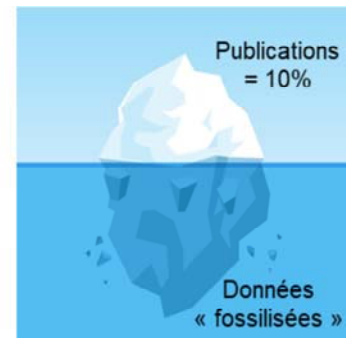
- Nouvelle recherche
- Enseignement
- Réexaminer les résultats / les données
- Conditions et protection juridique

Source :

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

POURQUOI GÉRER ET PARTAGER SES DONNÉES

- Gain de temps et d'argent, diminution des risques
- Croisements fertiles issus d'ensembles hétérogènes
- Gisements en pleine expansion
- Exhumation de données « fossilisées »
- Eviter la perte de données uniques
- Reproductibilité et réutilisabilité



"Designed by wstudio / Freepik"

- Gain de temps évident pour les chercheurs, diminution des risques
- Croisements fertiles issus d'ensembles hétérogènes : analyse des données
- Gisements en pleine expansion
- Exhumation de données « fossilisées » : les publications permettent d'accéder à environ 10 % des données, le reste demeurant disponible mais non utilisé sur les disques durs d'ordinateurs
- Eviter la perte de données uniques, riches en informations...
- Reproductibilité et réutilisabilité

Source :

Durand-Barthez Manuel. Les données de la Recherche. 17 avril 2018.

ACTEURS ET CONTRIBUTEURS



Chercheur : coordinateur du DMP, responsable des données : description des données, découpage des jeux de données...

Ingénieur-projet : coordonne les actions autour du DMP, agrément, éligibilité des coûts

Informaticien : interlocuteur pour le stockage et la sécurisation des données, les aspects infrastructure et les coûts associés

Spécialiste de l'IST : propose des standards, des métadonnées, conseille sur les entrepôts, réalise des alignements avec des référentiels existants...

Archiviste : aide le chercheur à sélectionner les données pour la conservation, à définir les durées et les solutions techniques

Juriste : conseille sur la propriété intellectuelle des données

Editeur : impose parfois le choix d'un entrepôt

DORANUM



(Données de la Recherche :
Apprentissage numérique)

- Dispositif de formation à distance
- Ressources d'autoformation sur la gestion et le partage des données de la recherche
- 9 thématiques
- 3 niveaux de formation

A screenshot of the DORA Num website. At the top, there is a navigation bar with links for 'ACCUEIL', 'RESSOURCES', 'ACTUALITES', 'A PROPOS', and 'CONTACT'. Below this is a header section with the title 'DONNÉES DE LA RECHERCHE APPRENTISSAGE NUMÉRIQUE' and a subtitle 'DES RESSOURCES POUR ACCOMPAGNER LA COMMUNAUTÉ SCIENTIFIQUE DANS LA GESTION ET LE PARTAGE DE LEURS DONNÉES'. A search bar is visible with the text 'Recherche avancée: 00000000000000000000' and a 'RECHERCHER' button. Below the search bar, there is a section titled 'VOUS SOUHAITEZ RÉUTILISER LES RESSOURCES ?' with a sub-heading 'Toutes les ressources ont été traduites en français. Vous pouvez les utiliser à votre convenance. Vous souhaitez plus de langues ? Contactez-nous !'. The main content area is titled 'THÉMATIQUES ABORDÉES' and 'PARCOUREZ LES RESSOURCES CLASSÉES PAR THÉMATIQUES'. It features a grid of nine thematic cards, each with an icon and a question: 1. 'ENJEUX & BÉNÉFICES' (gear icon) 'Pourquoi partager ses données ? Quelles sont les bonnes pratiques ?'; 2. 'ASPECTS JURIDIQUES, ÉTHIQUES, INTÉGRITÉ SCIENTIFIQUE' (scales icon) 'Qu'est-ce que le droit de la recherche ? Quelles pratiques de recherche respecter ?'; 3. 'PLAN DE GESTION DE DONNÉES' (document icon) 'Pourquoi et comment rédiger un plan de gestion des données ?'; 4. 'MÉTADONNÉES' (ID icon) 'Comment gérer les données ?'; 5. 'IDENTIFIANTS PÉRENNES' (ID icon) 'Comment associer durablement des données à son auteur ?'; 6. 'DÉPÔT & ENTREPÔTS' (warehouse icon) 'Comment et où déposer ses données ?'; 7. 'STOCKAGE & ARCHIVAGE' (storage icon) 'Quelles données stocker et à long terme ? Comment ?'; 8. 'DATA PAPERS & DATA JOURNALS' (document icon) 'Comment publier ses données comme un article scientifique ?'; 9. 'ACCÈS & VISUALISATION' (eye icon) 'Qu'est-ce que l'accès et comment accéder et visualiser ses données ?'; 10. 'ACCÈS & VISUALISATION' (eye icon) 'Qu'est-ce que l'accès et comment accéder et visualiser ses données ?'. At the bottom of the screenshot, there is a footer with the text 'Bonnes pratiques de gestion et de partage des données de recherche 16.09.19 P 9'.





PRINCIPES FAIR



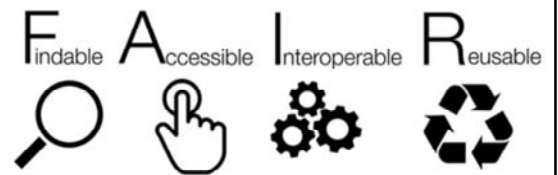
PRINCIPES FAIR

4 principes à respecter pour garantir une utilisation optimale des données de recherche et des métadonnées associées, à la fois par les hommes et par les machines.

- **F** (Findable) = Facile à trouver
- **A** (Accessible) = Accessible
- **I** (Interoperable) = Interopérable
- **R** (Reusable) = Réutilisable

Principes admis par les différentes communautés scientifiques au niveau international, ainsi que par les financeurs (ex : Commission européenne, ANR, etc.)

« aussi ouvert que possible,
aussi fermé que nécessaire »



Applicable tout au long du cycle de vie des données

MISE EN ŒUVRE DES PRINCIPES FAIR

- Sauvegarder et stocker ses données de manière sécurisée tout au long du projet → **FAIR**
- Bien organiser et nommer ses fichiers → **FAIR**
- Renseigner les métadonnées associées aux données → **FAIR**
 - de façon détaillée
 - en utilisant de préférence des standards existants (DublinCore étendu ; IPTC ; etc.)
- Utiliser des vocabulaires contrôlés disciplinaires (ontologies, lexiques, thesaurus...) → **FAIR**
- Utiliser des identifiants uniques et pérennes (ex : DOI) → **FAIR**
- Déposer ses données dans un entrepôt de données destiné au partage → **FAIR**
- Déposer ses données dans une plateforme d'archivage pérenne si vous souhaitez permettre une réutilisation à long terme (au-delà de 10 ans) → **FAIR**
- Utiliser des formats ouverts et non propriétaires → **AIR**
- Appliquer des licences de réutilisation → **AR**
- Communiquer ses codes sources → **AIR**



Principes FAIR :

Facile à trouver : facilite la découverte des données par les humains et les systèmes informatiques

- Données et métadonnées identifiées par un identifiant global unique et pérenne (ex. DOI)
- Métadonnées riches pour décrire les données (standards de métadonnées disciplinaires)
- Données et métadonnées enregistrées et indexées dans un dispositif permettant de les rechercher (ex. portail de données)
- Métadonnées spécifiant l'identifiant de la donnée.

Accessible : stockage durable des données et des métadonnées, accès et/ou téléchargement facilités, en spécifiant les conditions d'accès et d'utilisation

- Données et métadonnées accessibles par leur identifiant via un protocole de communication standardisé
- Protocole ouvert, libre, pouvant être implémenté de manière universelle (privilégier le dépôt dans un entrepôt certifié proposant un accès ouvert)

Interopérable : téléchargeable, utilisable, intelligible et combinable avec d'autres données, par des humains et des machines

- Données et métadonnées utilisant un langage formel, accessible, partagé et largement applicable pour la représentation des connaissances
- Données et métadonnées utilisant des vocabulaires disciplinaires (ontologies et vocabulaires contrôlés standards)

- Données et métadonnées incluant des liens vers d'autres (méta)données (versions antérieures ou plus récentes, données complémentaires, etc.) et vers des publications (articles citant les données, data papers).

Réutilisable : caractéristiques rendant les données réutilisables pour de futures recherches ou d'autres finalités (enseignement, innovation, reproduction/transparence de la science)

- Données et métadonnées : ayant des attributs multiples et pertinents ; mises à disposition selon une licence explicite et accessible ; associées à leur provenance ; correspondant aux standards des communautés indiquées.

Source :

INRA, Institut National pour la Recherche Agronomique. IST-Données de la Recherche. Produire des données FAIR. [En ligne]. 9 août 2018. Disponible sur :

<https://www6.inra.fr/datapartage/Produire-des-donnees-FAIR>

QU'APPORTENT LES PRINCIPES FAIR ?

Pour le chercheur

- Clarification de l'environnement de travail
- Balisage du cycle de vie des données par de bonnes pratiques
- Gain de temps dans la gestion d'un projet de recherche
- Meilleure visibilité du chercheur et de ses travaux de recherche
- Meilleure accessibilité à ses données
- Favorise l'intégrité scientifique
- Favorise les collaborations

Pour la communauté scientifique

- Accès facile à des données publiques, réutilisables
- Accès à des corpus utiles pour d'autres domaines
- Interopérabilité des données
- Gain de temps et d'argent : ne pas recréer des données déjà existantes
- Reproductibilité de la recherche



Ressources :

- DoRANum - <https://dorum.fr/>
- 5 ★ DATA RATING TOOL - <http://oznome.csiro.au/5star/#page-top>



PLAN DE GESTION DES DONNÉES DE RECHERCHE (PGD OU DMP)



PLAN DE GESTION DES DONNÉES IMPORTANCE

- **Echelle mondiale** : DMP de plus en plus recommandé ou exigé
- **Echelle européenne** : obligation pour Horizon 2020
- **En France** : obligation pour l'ANR
- **Echelle des organismes** : recommandations, politique d'établissements...
- **Echelle disciplinaire** : modèles de DMP par domaine



Le DMP est un phénomène mondial incontournable. Il est de plus en plus recommandé ou exigé, partout dans le monde.

- Echelle européenne :
 - Exigence de la Commission européenne (Modèles Horizon 2020; ERC)
 - Déploiements d'outils, infrastructures etc. d'ampleur européenne en lien avec la gestion et le partage des données de la recherche (l'entrepôt Zenodo, l'infrastructure OpenAIRE ...)
- Echelle nationale :
 - Financier ANR : DMP obligatoire depuis 2019. Modèle ANR
 - Etat français : politique nationale avec le Plan national pour la science
- Echelle des organismes :
 - Mise en place de « trame » de DMP institutionnelles (CIRAD, INRA, Institut Pasteur, Irstea, Universités...)
 - Politique d'établissements (INRA...)
 - A minima, recommandations intentionnelles (intégrées dans DMP OPIDoR)
- Echelle disciplinaire : création finalisée ou en cours de DMP par domaine disciplinaire (en astronomie, en archéologie...)

PLAN DE GESTION DES DONNÉES OUTIL DE GESTION DE PROJET

Document synthétique et évolutif qui



- facilite une bonne organisation de ses données tout au long du projet



- décrit la façon dont les données seront obtenues, traitées, organisées, stockées, sécurisées, préservées, partagées...



- garantit des données fiables et bien gérées, compréhensibles, disponibles et préservées sur le long terme

Ressources :

- [DoRanum - Plan de gestion de données](#)
- [Nathalie Reymonet, Magalie Moysan, Aurore Cartier, Renaud Délémontez - Réaliser un plan de gestion de données FAIR](#)



- Commencer à rédiger le DMP dès le début du projet (avec les éléments déjà connus ou prévus). Il peut être demandé dès la soumission du projet
- Compléter le DMP au fur et à mesure du projet
- Prévoir 3 versions au minimum :
 - o Au début du projet
 - o Au milieu du projet
 - o A la fin du projet.
- Désigner nominativement la ou les personne(s) responsable(s) de la gestion des données pour toutes les étapes du projet et au sein du partenariat s'il y a lieu :
 - o saisie des données,
 - o production des métadonnées,
 - o contrôle de la qualité des données,
 - o stockage, partage et archivage des données,
 - o mise à jour du DMP.
- Mutualiser les compétences des équipes autour du DMP
- Evaluer les ressources nécessaires (budget, temps alloué, personnels) permettant la mise en œuvre des actions décrites dans le DMP :
 - o temps nécessaire à la préparation des données pour le stockage, le partage et l'archivage des données
 - o coûts de matériel, rémunération des personnels
 - o frais de stockage (serveurs dédiés, traitement, maintenance, sécurité, accès...), partage (site web, publication...) et d'archivage des données

Ressources :

UK Data Service – Data management costing tool and checklist : <https://data-archive.ac.uk/media/247429/costingtool.pdf>

PLAN DE GESTION DES DONNÉES MODÈLE ANR

- Modèle composé de **6 grandes thématiques** illustrant les bonnes pratiques de gestion et de partage :



Modèle de DMP pour l'ANR

Bonnes pratiques de gestion et de partage des données de recherche

16.09.19

P 17

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

- a. Comment de nouvelles données seront-elles recueillies ou produites et/ou comment des données préexistantes seront-elles réutilisées ?
- b. Quelles données (types, formats et volumes par ex.) seront collectées ou produites ?

2. Documentation et qualité des données

- a. Quelles métadonnées et quelle documentation (par exemple méthodologie de collecte et mode d'organisation des données) accompagneront les données ?
- b. Quelles mesures de contrôle de la qualité des données seront mises en œuvre ?

3. Stockage et sauvegarde pendant le processus de recherche

- a. Comment les données et métadonnées seront-elles stockées et sauvegardées tout au long du processus de recherche ?
- b. Comment la sécurité des données et la protection des données sensibles seront-elles assurées tout au long du processus de recherche ?

4. Exigences légales et éthiques, codes de conduite

- a. Si des données à caractère personnel sont traitées, comment le respect des dispositions de la législation sur les données à caractère personnel et sur la sécurité des données sera-t-il assuré ?
- b. Comment les autres questions juridiques, comme la titularité ou les droits de propriété intellectuelle sur les données, seront-elles abordées ? Quelle est la législation applicable en la matière ?
- c. Comment les éventuelles questions éthiques seront-elles prises en compte, les codes déontologiques respectés ?

5. Partage des données et conservation à long terme

- a. Comment et quand les données seront-elles partagées ? Y-a-t-il des restrictions au partage des données ou des raisons de définir un embargo ?
- b. Comment les données à conserver seront-elles sélectionnées et où seront-elles préservées sur le long terme (par ex. un entrepôt de données ou une archive) ?
- c. Quelles méthodes ou quels outils logiciels seront nécessaires pour accéder et utiliser les données ?
- d. Comment l'application d'un identifiant unique et pérenne (comme le DOI) sera réalisée pour chaque jeu de données ?

6. Responsabilités et ressources en matière de gestion des données

- a. Qui (par exemple rôle, position et institution de rattachement) sera responsable de la gestion des données ?
- b. Quelles seront les ressources (budget et temps alloués) dédiées à la gestion des données permettant de s'assurer que les données soient FAIR (Facile à trouver, Accessible, Interopérable, Réutilisable) ?

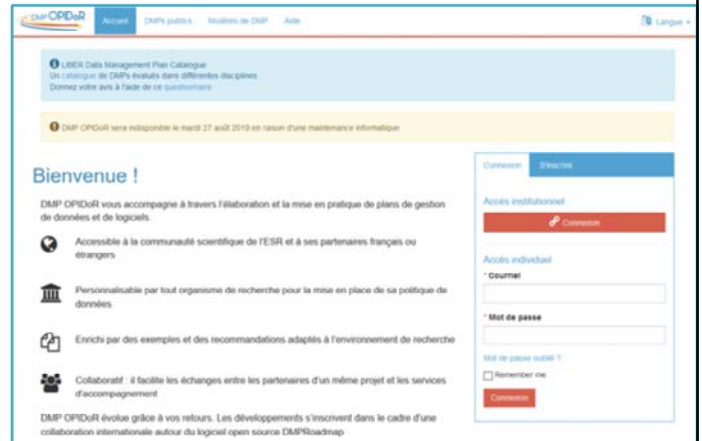
Source :

Modèle de DMP pour l'ANR : basé sur les recommandations de Science Europe (Science Europe – Guide pratique pour une harmonisation internationale de la gestion des données de recherche – juillet 2019 - <https://www.ouvrirlascience.fr/science-europe-guide-pratique-pour-une-harmonisation-internationale-de-la-gestion-des-donnees-de-recherche/>)

PLAN DE GESTION DES DONNÉES DMP OPIDOR

- Outil collaboratif d'aide à la rédaction en ligne de DMP
- Comprend :
 - Des modèles de DMP
 - Des exemples de DMP publics proposés en libre accès par leur rédacteur
 - Une aide dans laquelle on peut trouver des ressources utiles et des exemples de DMP

DMP OPIDoR



- Outil d'aide à la création en ligne de DMP permettant d'anticiper la gestion des données dans le cadre de projets de recherche
- Accessible à la communauté scientifique de l'ESR et à ses partenaires français ou étrangers
 - Personnalisable par tout organisme de recherche pour la mise en place de sa politique de données
 - Enrichi par des exemples et des recommandations adaptés à l'environnement de recherche
 - Collaboratif : il facilite les échanges entre les partenaires d'un même projet et les services d'accompagnement

EXEMPLES DE DMP

- Exemple de DMP public :
 - G2WAS – Grape Genes for Water Scarcity - https://dmp.opidor.fr/plan_export/2486.pdf
- Exemple de DMP associé à une thèse :
 - Data Management Plan for PhD Thesis "Climatic Limitation of Alien Weeds in New Zealand: Enhancing Species Distribution Models with Field Data" - <https://riojournal.com/article/8664/>
- Exemples de DMP disciplinaires :
 - Future TDM - <https://zenodo.org/record/1240420#.XW5xaHvgrx8>
 - BEHAVE Working Data-Management-Plan - <https://zenodo.org/record/1243717#.XW95IXvgrx8>
 - IPER-MAN Innovative PERmeable Materials for Airfoil Noise reduction Data Management Plan - <https://zenodo.org/record/1243999#.XW96DHvgrx8>



Les DMPs publics ne sont pas évalués. Ce sont les auteurs qui les mettent à disposition.



**ASPECTS JURIDIQUES
ET ETHIQUES
MODALITÉS DE
PROTECTION**



PROPRIÉTÉ INTELLECTUELLE DES DONNÉES DE RECHERCHE

- Règle générale : attribution de la propriété intellectuelle des données à l'établissement de tutelle des producteurs de données
- Particularités devant être précisées dans l'accord de consortium (entre partenaires) :
 - Cas de collaborations entre secteurs public et privé (déploiement industriel, commercialisation...)
 - S'il n'y a pas de consortium, la propriété des données et les responsabilités doivent être précisées dans le DMP
- Accompagnement par un juriste recommandé

Ressources :

- [La diffusion numérique des données en SHS – Guide des bonnes pratiques éthiques et juridiques](#)
- [Qui a les droits, quelles obligations ?](#)



Il ne s'agit pas du même droit que les publications (droit d'auteur principalement). Les données relèvent d'un régime lié au droit des bases de données. Dans ce cas, le droit de propriété appartient légalement au « producteur » de la base de données, compris au sens de la personne qui réalise l'investissement financier et matériel nécessaire à la constitution de la base. Il s'agira donc en général de l'établissement de tutelle des chercheurs qui sera considéré comme le titulaire effectif du droit de propriété.

Mais si ce droit existe formellement, il ne peut plus être opposé aux droits des ré-utilisateurs des données (principe d'ouverture des données). En effet, la loi pour une République numérique a explicitement « neutralisé » le droit des bases de données des administrations pour faire primer le principe de libre réutilisation. Il en résulte que les données produites par les chercheurs sont bien comprises dans le principe d'ouverture par défaut.

DROITS ET OBLIGATIONS DU CHERCHEUR

- Dès le début du projet, au moment de la collecte et de la production des données, le chercheur doit être vigilant concernant ses droits et obligations

Ex :

- Dans le cas d'une interview ou de prises de son ou de vue, il doit recueillir le consentement écrit des personnes concernées
 - Dans le cas de consultation de données d'archives, quels sont les droits afférents
 - Dans le cas de collecte d'objets archéologiques, quels sont les droits liés au pays de collecte.
- Cette étape est cruciale car elle détermine la latitude dont il disposera ensuite pour publier, diffuser et communiquer ses données et les résultats de ses recherches

Ressource :

- [Fiches pratiques sur le Règlement Général pour la Protection des Données](#)



OBLIGATIONS/RECOMMANDATIONS DE DIFFUSION

Principe « Aussi ouvert que possible, aussi fermé que nécessaire »

- **Obligations** : suivant le financeur du projet, il peut être obligatoire de diffuser ses données et de rédiger un Plan de Gestion des Données (PGD)
 - ANR : élaboration d'un PGD pour tous les projets financés, dans les 6 mois qui suivent le démarrage du projet
 - Horizon 2020 : rendre librement accessibles les articles scientifiques et les données (dont celles liées aux publications) et rédaction d'un PGD
- **Recommandations** : certaines institutions peuvent émettre des recommandations précises en matière de diffusion et de partage des données produites. Elles peuvent proposer à leurs chercheurs un modèle précis pour la rédaction du PGD

Ex : Cirad, Inra

Ressources :

- [Ouverture des données de recherche. Guide d'analyse juridique en France](#)
- [La diffusion numérique des données en SHS – Guide des bonnes pratiques éthiques et juridiques](#)
- [Qui a les droits, quelles obligations ?](#)



COMMUNICABILITÉ DES DONNÉES



La communicabilité des données peut être conditionnée par

- la nature ou le type des données
- leur origine
- leur utilisation

Elle peut être empêchée temporairement ou définitivement

COMMUNICABILITÉ DES DONNÉES OBLIGATIONS OU RESTRICTIONS

- **Communication obligatoire** pour certaines disciplines
Ex : Données géographiques ou environnementales
- **Communication interdite** par principe
Ex : secrets professionnels, secrets défense, sécurité de l'établissement...
- **Communication sous conditions**
Ex : Données protégées par le droit d'auteur ou par un contrat, données personnelles, statistiques...



Communication obligatoire pour certaines disciplines :

- Données géographiques
- Données environnementales

Communication interdite par principe :

- Données présentant des risques pour la protection du secret de la défense nationale
- Données présentant des risques pour la sûreté de l'État, la sécurité publique, la sécurité de l'établissement
- Secrets professionnels

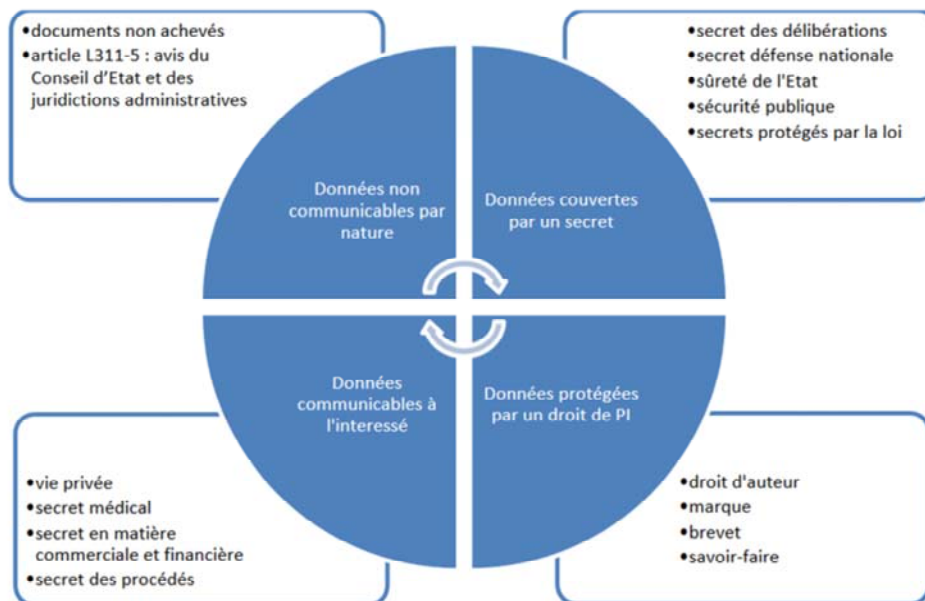
Communication sous conditions :

- Données présentant des risques pour la protection du potentiel scientifique et technique de la nation
- Cas des zones à régime restrictif (ZRR)
- Données protégées par le droit d'auteur et autres droits de propriété intellectuelle
- Données relatives aux personnes, à la vie privée
- Données statistiques
- Données liées à un contrat avec un tiers non soumis à une obligation de service public

Source :

Inra - Pôle Données de la Recherche IST – Partager / Publier – 8 août 2018 - <https://www6.inra.fr/datapartage/Partager-Publier>

RÉCAPITULATIF DES RESTRICTIONS



Bonnes pratiques de gestion et de partage des données de recherche

16.09.19

P 26

PI = Propriété Intellectuelle

Source :

DIST-CNRS. Le travail de la science et le numérique : données, plateformes, Publications. Une analyse systémique de la Loi numérique. 24 janvier 2017.

http://www.cnrs.fr/dist/z-outils/documents/20170203_analyse%20syst%C3%A9mique_vf.pdf

Libertés et restrictions :

La loi, oui mais... S'il y a des orchidées sauvages dans telle ou telle zone, peut-on diffuser l'information relative sans nuire à l'espèce protégée par la loi ?

Au contraire, le fait de les pointer du doigt ne va-t-il pas attirer l'attention sur la nécessité de les protéger ?

Seul le spécialiste peut juger de l'opportunité d'une mise à disposition ou d'une rétention.

MODALITÉS DE PROTECTION

▪ Accès



- Dispositif d'accès contrôlé : mot de passe...
- Accès limité dans le temps par un embargo
 - ✓ Pour disposer du temps nécessaire au dépôt de brevets...
 - ✓ En fonction de la discipline
 - ✓ Peut être déterminé par les éditeurs
- Accès limité à certaines personnes
Ex : limitation aux membres du consortium ou à une communauté scientifique



- Cryptage des données sensibles pour éviter des intrusions malveillantes



- Licences pour éviter une mauvaise utilisation des données par autrui

LICENCES DE DIFFUSION

Attribuer une licence de diffusion permet d'afficher clairement les modalités de réutilisation

| Recommandations Besoins de protection | Licence adaptée |
|--|---|
| Pays | France : Licence ouverte (Etalab) |
| Données dans le domaine public | Creative Commons CC0 |
| Auteur cité a minima | Creative Commons CC By |
| Protection plus importante | Licence CC combinée avec 3 éléments : <ul style="list-style-type: none">- NC : pas d'utilisation commerciale- SA : partage dans les mêmes conditions- ND : pas de modifications |
| Protection très restrictive | CC By-NC-ND |
| Logiciels | GNU GPL, CeCILL-B, etc. |
| Bases de données | ODbL |

16.09.19

P 28



Ressources :

Licences de réutilisation : <https://www.data.gouv.fr/fr/licences>

Licentia by Inria (<http://licentia.inria.fr/>) :

Permet de choisir quelle licence attribuer à ses données en utilisant quelques critères (permissions / obligations / interdictions).

- En vert : les licences compatibles avec TOUS les critères
- En orange : les licences pour lesquelles il manque certains critères
- En rouge : les licences ne correspondant à aucun des critères

Il permet également de savoir si une licence est compatible avec ses besoins, de visualiser et télécharger une licence, de la convertir en RDF.

License Selector : <http://ufal.github.io/public-license-selector/>

Choose an open source license : <https://choosealicense.com/>

ETHIQUE DES DONNÉES DE RECHERCHE



- Dans le cas de données devant respecter des règles d'éthique particulières : préciser les normes, chartes, déclarations, codes, politiques auxquels on se réfère
- En cas de recours à un comité d'éthique : expliquer le processus de recrutement et d'évaluation

Ressources :

- Politique en matière d'éthique et d'intégrité scientifique (ANR)
- La diffusion numérique des données en SHS – Guide des bonnes pratiques éthiques et juridiques
- Questions éthique et droit en SHS



Des normes, chartes, déclarations, codes et politiques en éthique et en intégrité scientifique encadrent déjà les pratiques pour l'ensemble de nombreux acteurs internationaux de la recherche.

L'ANR encourage les équipes de recherche à intégrer dans leur démarche une réflexion sur les enjeux éthiques qui pourraient être soulevés par les objectifs, la méthodologie ou les résultats attendus de leur projet de recherche.

Responsabilités éthiques :

Les équipes qui réalisent les projets de recherche soutenus par l'ANR doivent :

- rechercher systématiquement l'originalité et l'innovation
- respecter les dispositions en vigueur concernant la recherche sur l'être humain, l'expérimentation animale et le respect de l'environnement
- prendre toutes les mesures raisonnables pour estimer les risques et les dangers qui pourraient survenir dans le cadre de la recherche
- prendre toutes les précautions nécessaires pour protéger la santé et la sécurité de ceux qui prennent part à la recherche, tant ceux qui la réalisent que ceux qui y participent comme sujets
- respecter les droits et règles (dont coutumières), concernant l'accès aux informations (collections, enquête,...) et aux ressources en vigueur dans les pays et collectivités notamment outre-mer, accueillants, la recherche pouvant être exécutée dans ou avec des pays tiers
- s'assurer que la collaboration au sein du partenariat est équitable. Les partenaires doivent s'impliquer en toute liberté et sans pression
- s'assurer que la nature confidentielle des informations recueillies et le droit à la protection des renseignements personnels soient garantis
- prendre des mesures appropriées pour que les données soient conservées ou détruites conformément aux législations et normes en vigueur.

Source :

ANR - Politique en matière d'éthique et d'intégrité scientifique - <https://anr.fr/fileadmin/documents/2014/Politique-ethique-integrite-scientifique-aout-2014.pdf>



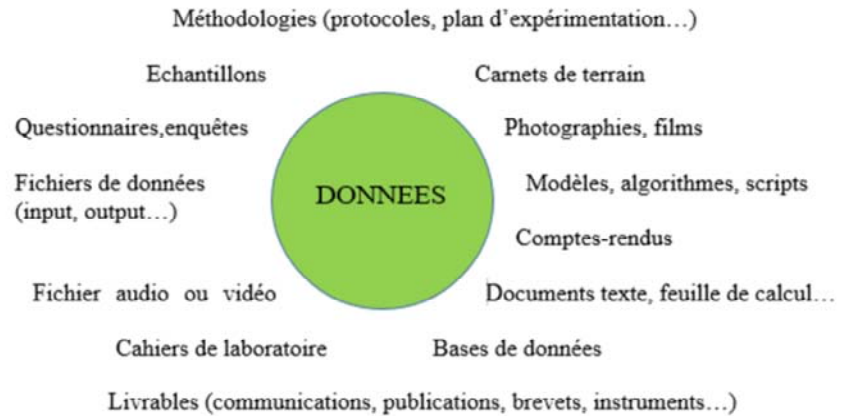
CRÉATION, COLLECTE ET DESCRIPTION DES DONNÉES DE RECHERCHE



RAPPEL : DIVERSITÉ DES DONNÉES DE RECHERCHE

Les données de recherche peuvent être

- **produites**, lors de campagnes de recherche (observations, mesures...)
- **collectées** : données déjà existantes (corpus, archives...)



Selon leur contexte de création (capture ou production), leur exploitation, leur analyse et les traitements qu'elles subissent, les données de recherche peuvent être

- de différente **nature** : brutes, dérivées, formatées, nettoyées, primaires, secondaires, traitées....
- contenues dans divers **supports** : carnets de laboratoire, documents électroniques, logiciels, papier, programmes informatiques...
- de tous **types** : archives, audio, vidéo, bases de données, codes sources, géospatiales, images, photographies, langages de programmation, matérielles et physiques, modèles, visualisations, 3D, numériques, textuelles, numérisations, scans, qualitatives, quantitatives, statistiques...

Sources :

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

UNIL – Université de Lausanne – Nature, structure et types des données de recherche - <https://uniris.unil.ch/researchdata/sujet/comprendre-gestion-donnees-recherche/donnees-de-recherche-definitions/nature-structure-types/>

PRÉPARATION ET DOCUMENTATION DES DONNÉES

Il est impératif de **bien préparer et documenter ses données** afin d'optimiser le stockage, le partage, l'archivage et la réutilisation

Dans le DMP, il faudra indiquer de manière précise quelles méthodes sont utilisées pour recueillir ou produire les données :



Attention aux **données sensibles, personnelles ou confidentielles** : prendre les précautions nécessaires afin de respecter les règles juridiques et éthiques en vigueur

Dans le cas de données collectées :

- leur provenance (corpus, archives...),
- sur quels critères elles ont été sélectionnées
- les conditions de réutilisations préexistantes de ces données

Dans le cas de données produites (observations, mesures, etc.) :

- le contexte de création,
- les méthodes utilisées,
- les protocoles suivis ou établis,
- les contrôles qualité mis en place

bonnes pratiques de gestion et de partage des données de recherche 10/09/19 p. 32



Qualité des données :

L'ouverture des données - mais aussi des logiciels, ontologies et métadonnées qui en permettent l'exploitation - impliquent une nouvelle responsabilité : celle d'être particulièrement soucieux de la qualité des informations et des données offertes ainsi que de la clarté de la documentation qui les accompagne. Pour permettre à d'autres de répliquer ou de réutiliser des données, il est nécessaire de vérifier le caractère intègre et interopérable des données, l'identification de leurs sources, leurs dates de recueil ou de traitement, ainsi que l'examen détaillé des différentes étapes de la constitution de dépôts de données : collecte, classification, standardisation, mise à disposition, réutilisation, conservation, destruction.

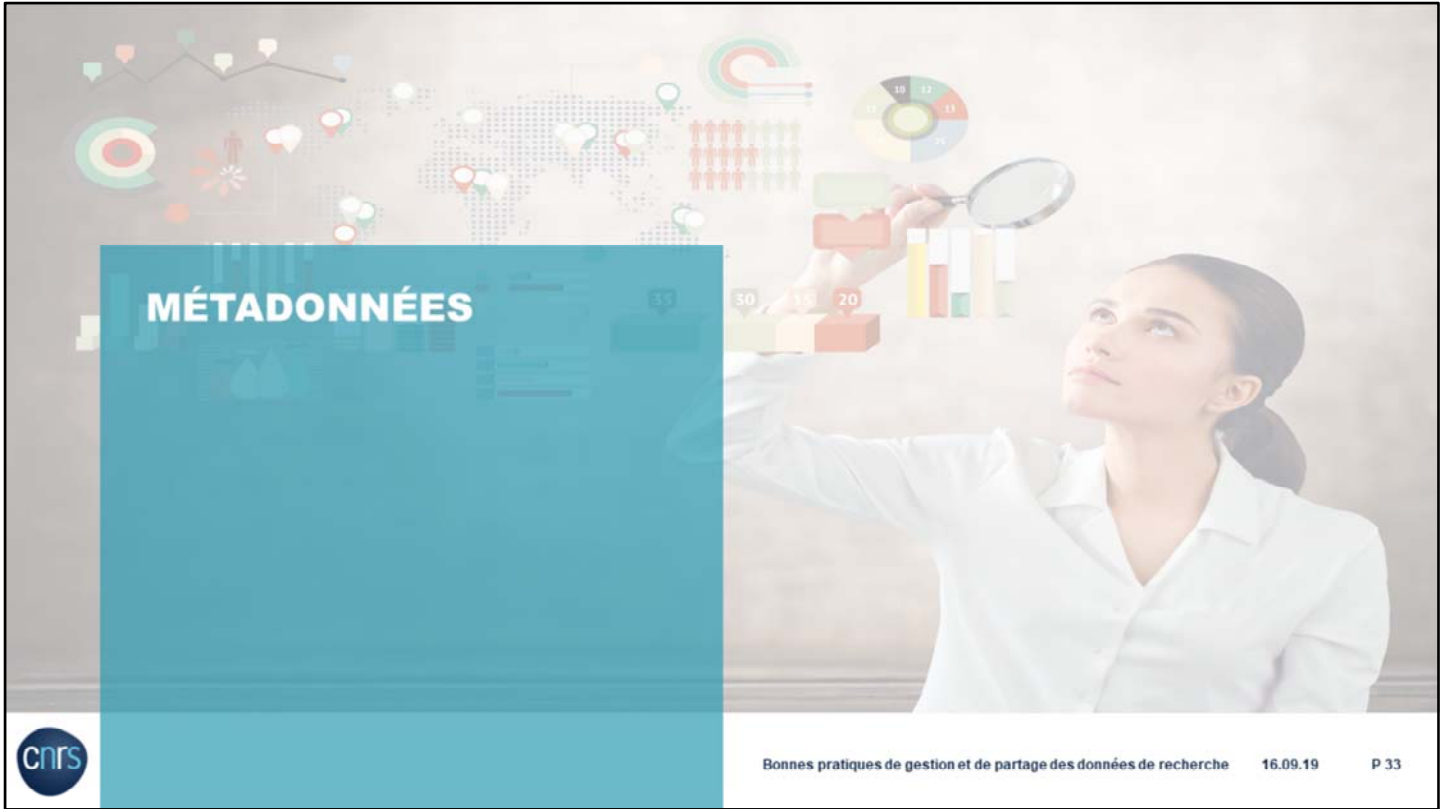
La charte nationale de déontologie des métiers de la recherche souligne ces points essentiels :

- la description détaillée du protocole de recherche dans le cadre des cahiers de laboratoire, ou de tout autre support, doit permettre la traçabilité des travaux expérimentaux
- tous les résultats bruts (qui appartiennent à l'institution) ainsi que l'analyse des résultats doivent être conservés de façon à permettre leur vérification.

Sources :

DIST-CNRS, Direction de l'Information Scientifique et Technique. Livre blanc - Une science ouverte dans une république numérique. 21 mars 2016. <http://www.cnrs.fr/dist/z-outils/documents/2016%2003%2024%20Livre%20blanc%20Open%20Science.pdf>

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



MÉTADONNÉES



MÉTADONNÉES DÉFINITION

- Les métadonnées permettent de décrire plus précisément les données
Ex : boîte de conserve (étiquette = métadonnées)



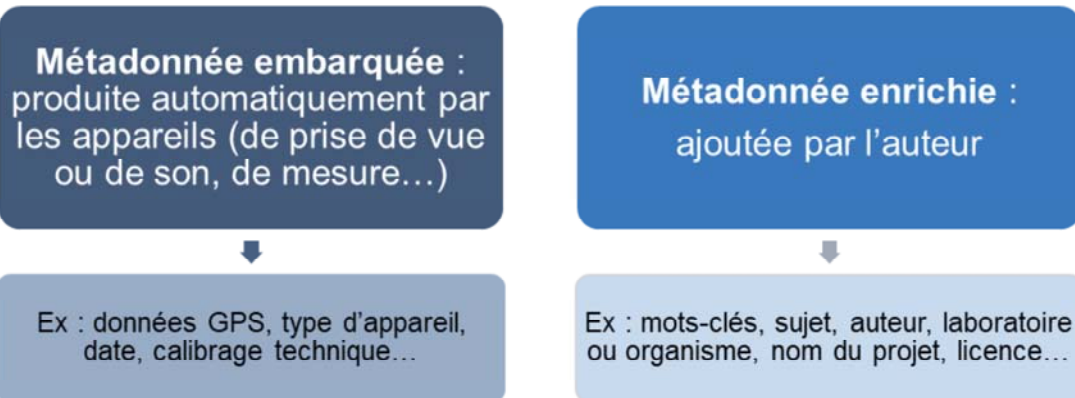
Sans métadonnées



Avec métadonnées

MÉTADONNÉES DÉFINITION - UTILISATION

- Différence entre métadonnées embarquées et métadonnées enrichies :



Tous ces éléments très importants devront être précisés lors de la rédaction du DMP ou lors du dépôt dans un entrepôt

MÉTADONNÉES ENRICHISSEMENT DES MÉTADONNÉES

- Utiliser des vocabulaires contrôlés disciplinaires utilisés couramment : (ontologies, lexiques, thesaurus...)
- Ex :
 - Codex de médicaments
 - Classifications taxonomiques
 - Nomenclature internationale des formules chimiques
 - ...
- Cela augmentera la capacité des données à être combinées avec d'autres données



UTILITÉ DES MÉTADONNÉES

- Comprendre l'origine des données et leur contexte de création ou de collecte
- Améliorer le moissonnage par les machines (moteur de recherche)
- Garantir l'interopérabilité
- Connaître les conditions de réutilisation et de partage des données
- Accéder à des informations très utiles lorsqu'on ne peut pas partager ses données ou lors du retrait des données

SCHÉMA ET STANDARD DE MÉTADONNÉES

Schéma :

- Organisation des métadonnées déclinée pour l'ensemble des jeux de données : hiérarchie, plan...
- Peut être créé par les partenaires du projet en fonction des besoins

Ressource :

- [DoRANum - Les schémas de métadonnées](#)

Standard :

- Schéma de métadonnées normalisé, reconnu et utilisé à grande échelle
- Il peut être imposé par
 - un organisme de tutelle
 - un entrepôt

Ressources :

- [DoRANum - Les standards de métadonnées : pourquoi, lequel ?](#)
- Répertoire de standards de métadonnées en Sciences de la Vie [FAIRsharing.org](#)
- [RDA Metadata Standards Directory](#) ou [Disciplinary Metadata](#)
- Informations fournies par les entrepôts de données sur les standards de métadonnées



Exemples de standards de Métadonnées :

- **Dublin Core (interdisciplinaire)** : description des ressources numériques. <http://dublincore.org/>
- **DataCite Metadata Schema** : métadonnées enregistrées dans le DataCite Metadata Store lors de la création d'un DOI pour un jeu de données. Permet l'identification précise et cohérente des données à des fins de citation et de réutilisation. <https://schema.datacite.org/>
- **CSMD-CCLRC Core Scientific Metadata Model** : domaines des sciences structurales (chimie, science des matériaux, sciences de la terre, biochimie) <http://icatproject-contrib.github.io/CSMD/>
- **DDI (Data Documentation Initiative)** : domaine des sciences sociales, comportementales et économiques. <http://www.ddialliance.org/>
- **DwC (Darwin Core)** : domaine de la biodiversité. <http://rs.tdwg.org/dwc/>
- **EML (Ecological Metadata Language)** : très développé en écologie. En grande partie conçu pour décrire des ressources numériques. Il peut également être utilisé pour décrire des ressources non numériques telles que des cartes papier ou d'autres médias. <https://knb.ecoinformatics.org/external/emlparser/docs/index.html>
- **MIDAS-Heritage** : domaine de l'architecture. <https://historicengland.org.uk/images-books/publications/midas-heritage/>



IDENTIFIANTS PÉRENNES



IDENTIFIANTS PÉRENNES

Attribuer un **identifiant auteur** (ORCID)

- Fait le lien avec ses productions scientifiques
- Permet d'être bien identifié et cité

Ressources :

- [DoRANum - Zoom sur ORCID et DOI](#)



Bonnes pratiques de gestion et de partage des données de recherche 16.09.19 P 40

Une équipe dédiée de l'Inist-CNRS se tient à disposition pour conseiller dans l'attribution de DOI aux données de recherche (abonnement annuel de 180 euros / an pour l'attribution de 500 DOI) et fournit :

- un accès à un espace test pour enregistrer temporairement des DOI et vérifier la compatibilité de ce service avec ses propres workflows
- un préfixe unique de DOI
- un accès à la plateforme Metadata Store de DataCite (MDS) pour commencer à créer les DOI
- une assistance à la création et à la conversion de métadonnées...
- un accompagnement dans l'utilisation des différents services proposés par DataCite.

IDENTIFIANTS PÉRENNES

Attribuer un **identifiant pérenne à chacun des jeux de données**

Les plus utilisés sont DOI et Handle)

Facilite le suivi, la localisation, l'accès et la citation des données lors de leur publication ou à des fins de réutilisation

Ressources :

- [Zoom sur ORCID et DOI](#)
- [PID OPIDoR](#) : service de l'Inist-CNRS, agence nationale [DataCite](#) pour l'attribution de DOI



Une équipe dédiée de l'Inist-CNRS se tient à disposition pour conseiller dans l'attribution de DOI aux données de recherche (abonnement annuel de 180 euros / an pour l'attribution de 500 DOI)

et fournit :

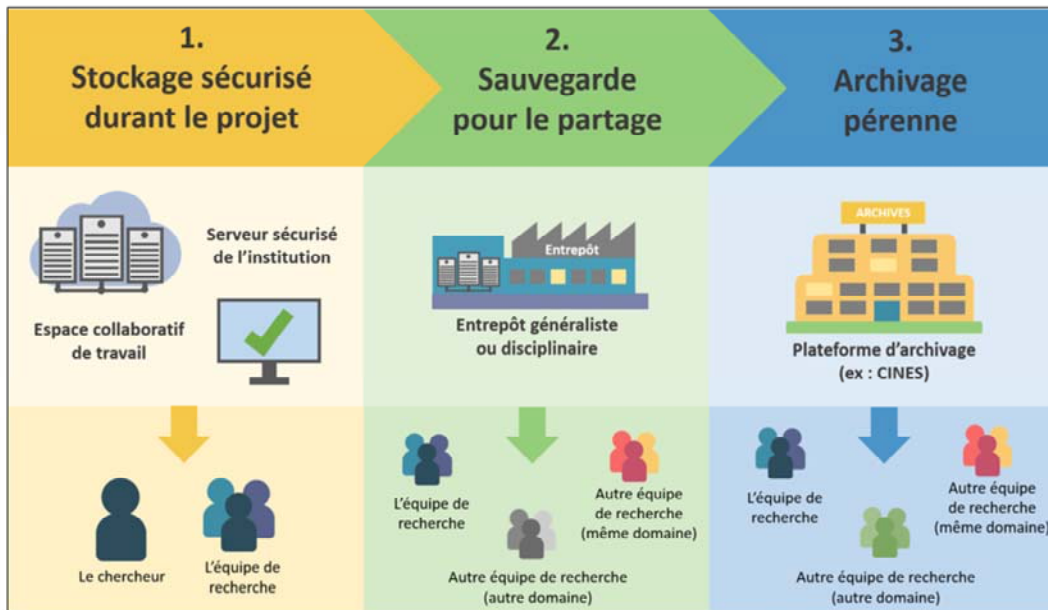
- un accès à un espace test pour enregistrer temporairement des DOI et vérifier la compatibilité de ce service avec ses propres workflows
- un préfixe unique de DOI
- un accès à la plateforme Metadata Store de DataCite (MDS) pour commencer à créer les DOI
- une assistance à la création et à la conversion de métadonnées...
- un accompagnement dans l'utilisation des différents services proposés par DataCite.



SAUVEGARDE DES DONNÉES TOUT AU LONG DU PROJET



3 ÉTAPES DE SAUVEGARDE DES DONNÉES





ETAPE 1
STOCKAGE SÉCURISÉ
DURANT LE PROJET



MESURES DE SAUVEGARDE TOUT AU LONG DU PROJET (STOCKAGE)

1.

Stockage sécurisé durant le projet



- Dupliquer et stocker les données à différents endroits sur différents supports

Règle du 3-2-1 :

- garder 3 exemplaires des données,
- sur 2 supports ou technologies différents,
- dont 1 se trouve hors site

- Définir l'hébergement, le stockage des données et la politique de sauvegarde associée
- Gérer les versions



Organiser et planifier les sauvegardes :

- Sélectionner les données à sauvegarder tout au long du projet
- Définir leur durée de conservation, en se posant par exemple ces questions :
 - Le contenu est-il valorisé ou toujours valorisable ?
 - Le fichier et son contenu seront-ils toujours compréhensibles ?
 - Le support physique utilisé va-t-il résister dans le temps ?
 - Le format de fichier utilisé sera-t-il toujours lisible par un logiciel ?
- Indiquer celles qui nécessitent d'être détruites pour des raisons contractuelles (exemple : données issues d'un projet mené avec une entreprise privée), légales (exemple : données personnelles) ou réglementaires
- Estimer la volumétrie des données
- Définir l'hébergement, le stockage des données et la politique de sauvegarde associée : serveurs locaux (machines virtuelles), cloud institutionnel avec accès sécurisé
- Gérer les versions :
 - Les différents états des données sont conservés en corrélation avec les différentes étapes de traitement
 - Permet de revenir à une version antérieure si besoin.

Source :

Université de Lille Sciences et Technologies -

https://indico.math.cnrs.fr/event/2317/contributions/1314/attachments/692/773/problematique_des_sauvegardes-journees_mathrice-20170928.pdf

NOMMAGE DES FICHIERS

La fiabilité d'accès passe par un nommage unique et précis des fichiers de données :

- 30 caractères maximum
- Noms de partenaires insérables si leur graphie est harmonisée entre les fichiers
- Numéros de versions le cas échéant
- Dates au format ISO : AAAA-MM-JJ

- Pas de caractères spéciaux ou accentués du type ùéàç+'@°[] :</>*/ »& !\$...
- Séparateurs : pas d'espace, pas de mots vides, éventuellement Majuscules ou underscore _
- Pas de dénomination vague : divers, autres, à classer...



Source :

Arnould Pierre-Yves, Jacquemot-Perbal Marie-Christine. Guide de bonnes pratiques : Gestion et valorisation des données de recherche. 1er février 2016.

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

FORMATS DE FICHIERS

Formats ouverts et non propriétaires

Opter pour des formats de fichiers les plus ouverts possible (non propriétaires), standardisés et pérennes

Exemples :

- Privilégier .csv à .xlsx
- Privilégier .odt à .docx
- Privilégier .jpg à .tif

Choix du format

Le choix d'un format peut être guidé par :

- les recommandations de son institution
- les usages de la communauté scientifique de la discipline
- les logiciels ou équipements utilisés

Il faudra le justifier dans le DMP

Ressource :

[DoRANum - Tableau format ouvert ou fermé](#)





ETAPE 2

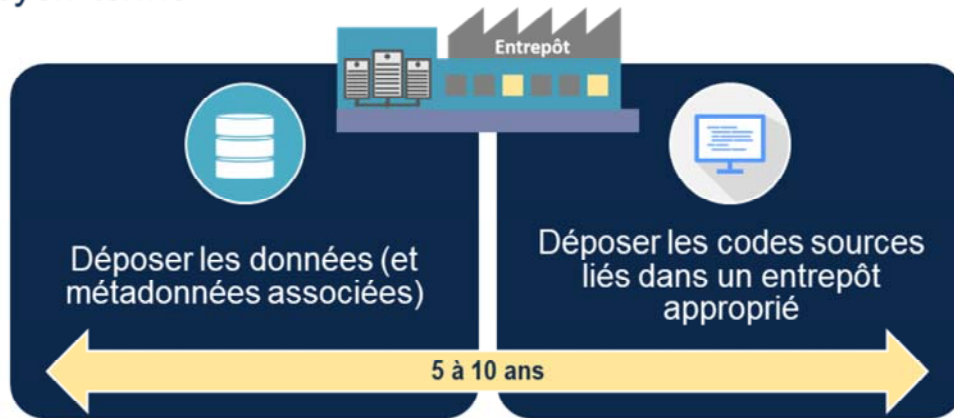
PARTAGE, DIFFUSION DES DONNÉES

DÉPÔT DANS UN ENTREPÔT



DÉPÔT DANS UN ENTREPÔT

- Permet le partage et la réutilisation optimale des données sur le court et le moyen terme



Ressources :

- Pour les logiciels, déposer les codes sources dans [HAL](#) (lien avec [Software Heritage](#), archive de logiciels)

PRÉPARATION DES DONNÉES POUR LE PARTAGE DANS UN ENTREPÔT

- Sélectionner les données à partager
- Vérifier la compatibilité et l'interopérabilité des formats de données
- Migrer si nécessaire vers un format adapté, le plus ouvert possible
- Préparer si nécessaire les codes sources (ex : scripts) qui permettront de lire et traiter les données
- Compléter et enrichir les métadonnées (en fonction de l'entrepôt choisi)
 - Si ce n'est pas déjà fait, choisir un standard de métadonnées
 - S'il n'en existe pas d'adapté, créer un schéma de métadonnées
 - Compléter les champs pour chaque jeu de données, suivant le standard adopté

CHOIX DE L'ENTREPÔT

- Il est souvent recommandé par son institution, son financeur ou sa communauté scientifique
Ex : Data Inra
- Il est parfois imposé par un éditeur
Ex : Gene Expression Omnibus
- S'il n'y a pas de recommandation, le choisir dans un annuaire, en fonction de ses besoins

Ressources :

- Annuaire d'entrepôts : re3data, OAD, OpenDOAR



Exemples d'entrepôts imposés par l'éditeur :

Séquençage haut débit (ADN / ARN) :

- Gene Expression Omnibus (GEO) - <https://www.ncbi.nlm.nih.gov/geo/>
- Sequence Read Archive (SRA) - <https://www.ncbi.nlm.nih.gov/sra>

Source :

Pierre Poulain - Retour d'expériences sur la

publication de données en biologie – 27 novembre 2018 - https://qt-donnees2018.sciencesconf.org/data/08_donnees_biologie_Poulain_2018.pdf

PRINCIPAUX CRITÈRES DE CHOIX D'UN ENTREPÔT

- Discipline
- Type de données acceptées
- Qualité des métadonnées
- Certification
- Entrepôt de confiance
- Pérennité des données
- Génération d'un identifiant pérenne
- Gestion des versions



Un entrepôt digne de confiance devrait permettre et assurer :

- le repérage et l'identification des données
- la recherche, la citation et le téléchargement des données
- la gestion des versions des jeux de données
- le référencement d'informations pertinentes complémentaires, telles que d'autres jeux de données et publications
- l'accès ouvert à des informations mises à jour, y compris sur des données non publiées, protégées, rétractées ou supprimées : métadonnées archivées sur le long terme, même si les données correspondantes ne sont plus disponibles
- la récupération des métadonnées par les machines
- l'accès aux données dans des conditions bien définies (licences)
- l'authenticité et l'intégrité des données
- la confidentialité et le respect des droits des personnes et créateurs de données
- la pérennité des données et métadonnées

Source :

Science Europe - Practical guide to the international alignment of research data management - novembre 2018 - https://www.scienceeurope.org/wp-content/uploads/2018/12/SE_RDM_Practical_Guide_Final.pdf

ENTREPÔT GÉNÉRALISTE : ZENODO

- **Zenodo** : permet de déposer des ensembles de données et de les lier aux publications
- Recherche pouvant se faire sur différents types de documents : jeux de données, images, posters, présentations, publications, vidéos/audios
- Affecte des DOIs aux objets et expose les métadonnées de description pour le moissonnage
- Permet de définir 3 niveaux d'accès :
 - « open access »
 - « restricted access » avec le courriel à contacter pour obtenir le document
 - « closed access »

Ressource :

- [Déposer ses données de recherche dans Zenodo](#)



Source :

Aventurier P. M. Zenodo, un entrepôt de données. 16 septembre 2013.

<https://ist.blogs.inra.fr/technologies/2013/09/16/zenodo-un-entrepot-de-donnees/>

ENTREPÔT MULTIDISCIPLINAIRE : NAKALA

- **Nakala** : entrepôt en Sciences Humaines et Sociales
- Permet de déposer, documenter et diffuser les données de recherche
- Services d'accès aux données elles-mêmes
- Services de présentation des métadonnées
- Moissonnable et optimisé pour Isidore
- Les producteurs de données soulagés de la gestion purement technique, peuvent se consacrer à la valorisation scientifique de leurs données









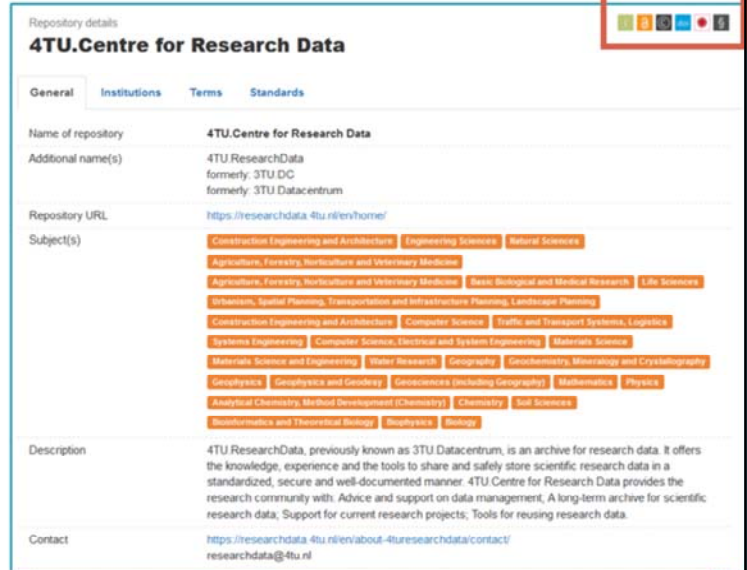
Isidore : plateforme et moteur de recherche permettant d'effectuer des requêtes et d'accéder aux données numériques de la recherche en sciences humaines et sociales développée par l'infrastructure TGIR Huma-Num.

RECHERCHE D'ENTREPÔTS DISCIPLINAIRES

Dans **re3data**,
exemples d'entrepôt répondant à de nombreux critères :

- [4TU.Centre for Research Data](#)
- [CLARIN repository at the University of Tübingen](#)
- [NASA Socioeconomic Data and Applications Center](#)
- [PANGAEA](#)

-  Informations complémentaires fournies
-  Libre accès aux données
-  Conditions d'utilisation et licence fournis
-  Génération d'un DOI
-  Certification
-  Politique de l'entrepôt fournie



Repository details

4TU.Centre for Research Data

General | Institutions | Terms | Standards

Name of repository: 4TU.Centre for Research Data

Additional name(s): 4TU ResearchData
formerly: 3TU.DC
formerly: 3TU Datacentrum

Repository URL: <https://researchdata.4tu.nl/en/home/>

Subject(s):

- Construction Engineering and Architecture
- Engineering Sciences
- Natural Sciences
- Agriculture, Forestry, Horticulture and Veterinary Medicine
- Basic Biological and Medical Research
- Life Sciences
- Urbanism, Spatial Planning, Transportation and Infrastructure Planning, Landscape Planning
- Construction Engineering and Architecture
- Computer Science
- Traffic and Transport Systems, Logistics
- Systems Engineering
- Computer Science, Electrical and System Engineering
- Materials Science
- Metals Science and Engineering
- Water Research
- Geography
- Geochemistry, Mineralogy and Crystallography
- Geophysics
- Geophysics and Geodesy
- Geosciences (Including Geography)
- Mathematics
- Physics
- Analytical Chemistry, Method Development (Chemistry)
- Chemistry
- Soil Sciences
- Bioinformatics and Theoretical Biology
- Biophysics
- Biology

Description: 4TU ResearchData, previously known as 3TU Datacentrum, is an archive for research data. It offers the knowledge, experience and the tools to share and safely store scientific research data in a standardized, secure and well-documented manner. 4TU Centre for Research Data provides the research community with: Advice and support on data management, A long-term archive for scientific research data; Support for current research projects; Tools for reusing research data.

Contact: <https://researchdata.4tu.nl/en/about-4turesearchdata/contact/>
researchdata@4tu.nl

Possibilité de consulter l'annuaire par sujet, type de contenu, pays. La recherche graphique par sujet est matérialisée par une succession de cercles concentriques, du plus général vers le plus spécifique.

ENTREPÔT DISCIPLINAIRE : DRYAD

Dryad : entrepôt en Sciences de la Vie, Agronomie, Géosciences, Anthropologie et Sciences comportementales

The screenshot shows the Dryad website interface. At the top, there is a navigation bar with links for 'About', 'For researchers', 'For organizations', 'Contact us', 'Log in', and 'Sign up'. Below this, a featured article is displayed with the title 'Data from: Latitudinal patterns of herbivore pressure in a temperate herb support the biotic interactions hypothesis' by Baskett CA and Schemske DW, published on February 28, 2018. A 'Submit data now' button and a 'How and why?' link are visible. A search bar is present with a 'Go' button. A section titled 'Files in this package' lists the title 'Herbivory on Phytolacca americana from Michigan to Florida in field', which has been downloaded 28 times. Two files are listed for download: 'README.pdf (64.42 Kb)' and 'baskett_herbiv_lat_PHAM.csv (15.14 Kb)'. A red arrow points from the 'baskett_herbiv_lat_PHAM.csv' link to the corresponding row in the adjacent table.

| | A | B | C | D | E | F |
|----|------------|------------|----------|----------|--------------|---------------------------|
| 1 | study_year | population | latitude | plant_ID | leaf_age | herbivory,marked_interval |
| 2 | 2014 | HV,37.1818 | HV1 | mature | 1.613089141 | 34 |
| 3 | 2014 | HV,37.1818 | HV1 | young | 25 | 34 |
| 4 | 2014 | HV,37.1818 | HV10 | mature | -0.588768116 | 34 |
| 5 | 2014 | HV,37.1818 | HV10 | young | 40 | 34 |
| 6 | 2014 | HV,37.1818 | HV11 | mature | -0.640116398 | 34 |
| 7 | 2014 | HV,37.1818 | HV11 | young | 0 | 34 |
| 8 | 2014 | HV,37.1818 | HV12 | mature | 18.55026049 | 34 |
| 9 | 2014 | HV,37.1818 | HV12 | young | 1.515151515 | 34 |
| 10 | 2014 | HV,37.1818 | HV13 | mature | 25.47731804 | 34 |
| 11 | 2014 | HV,37.1818 | HV13 | young | 10.93290043 | 34 |
| 12 | 2014 | HV,37.1818 | HV2 | young | 2.5 | 34 |
| 13 | 2014 | HV,37.1818 | HV3 | mature | 0.657981803 | 34 |
| 14 | 2014 | HV,37.1818 | HV3 | young | 0 | 34 |
| 15 | 2014 | HV,37.1818 | HV5 | mature | 0.073229772 | 34 |
| 16 | 2014 | HV,37.1818 | HV5 | young | 33.33333333 | 34 |
| 17 | 2014 | HV,37.1818 | HV6 | mature | 0.99845286 | 34 |
| 18 | 2014 | HV,37.1818 | HV6 | young | 3.75 | 34 |
| 19 | 2014 | HV,37.1818 | HV7 | mature | 0.524662742 | 34 |
| 20 | 2014 | HV,37.1818 | HV7 | young | 0.3125 | 34 |



ENTREPÔT DISCIPLINAIRE : DRYAD

Dryad : métadonnées

| | |
|------------------|--|
| Funding | National Science Foundation (United States), grant number DGE-0802267 and DGE-1424871 |
| Pageviews | 43 |
| Keywords | herbivory , plant defense , plant-insect interactions , latitudinal gradient , biotic interactions hypothesis , geographic variation , evolutionary ecology , Lepidoptera , threshold , climatic gradient , palatability , common garden |
| Scientific Names | <i>Phytolacca americana</i> , <i>Spodoptera exigua</i> , <i>Argyrotaenia velutinana</i> , <i>Asciodes gordialis</i> , <i>Psara obscuralis</i> , <i>Disclisioprocta stellata</i> , <i>Spodoptera eridania</i> |
| Spatial Coverage | Eastern USA |

Abstract

The longstanding biotic interactions hypothesis predicts that herbivore pressure declines with latitude, but the evidence is mixed. To address gaps in previous studies, we measured herbivory and defense in the same system, quantified defense with bioassays, and considered effects of leaf age. We quantified herbivory and defense of young and mature leaves along a continental gradient in eastern North America in the native herb *Phytolacca americana*. Herbivory in the field declined with latitude and was strongly correlated with Lepidopteran abundance. Laboratory bioassays revealed that leaf palatability was positively correlated with latitude of origin. Young leaves were more damaged than mature leaves at lower latitudes in the field, but less palatable in bioassays. Both defense and palatability displayed non-linear latitudinal patterns, suggesting potential mechanisms based on biological or climatic thresholds. In sum, observational and experimental studies find patterns consistent with high herbivore pressure and stronger plant defenses at lower latitudes.

[Show Full Metadata](#)

| Show simple item record | |
|-------------------------|--|
| dc:contributor.author | Baskett, Carina A.  |
| dc:contributor.author | Schemske, Douglas W. |
| dc:coverage.spatial | Eastern USA |
| dc:date.accessioned | 2018-02-28T17:28:45Z |
| dc:date.available | 2018-02-28T17:28:45Z |
| dc:date.issued | 2018-02-23 |
| dc:identifier | doi:10.5061/dryad.11715 |
| dc:identifier.citation | Baskett CA, Schemske DW (2018) Latitudinal patterns of herbivore pressure in a temperate herb support the biotic interactions hypothesis. <i>Ecology Letters</i> 21(4): 578-587. |
| dc:identifier.uri | http://hdl.handle.net/10255/dryad.165862 |
| dc:description | The longstanding biotic interactions hypothesis predicts that herbivore pressure declines with latitude, but the evidence is mixed. To address gaps in previous studies, we measured herbivory and defense in the same system, quantified defense with bioassays, and considered effects of leaf age. We quantified herbivory and defense of young and mature leaves along a continental gradient in eastern North America in the native herb <i>Phytolacca americana</i> . Herbivory in the field declined with latitude and was strongly correlated with Lepidopteran abundance. Laboratory bioassays revealed that leaf palatability was positively correlated with latitude of origin. Young leaves were more damaged than mature leaves at lower latitudes in the field, but less palatable in bioassays. Both defense and palatability displayed non-linear latitudinal patterns, suggesting potential mechanisms based on biological or climatic thresholds. In sum, observational and experimental studies find patterns consistent with high herbivore pressure and stronger plant defenses at lower latitudes. |
| dc:relation.haspart | doi:10.5061/dryad.11715/1 |
| dc:relation.haspart | doi:10.5061/dryad.11715/2 |



Dans cet exemple, les métadonnées suivant les standards utilisés sont Dublin Core et Darwin Core.

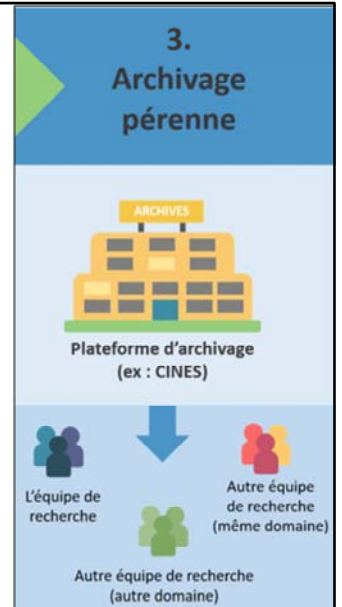


ETAPE 3 ARCHIVAGE PÉRENNE



ARCHIVAGE PÉRENNE

- Les données à archiver doivent présenter une valeur scientifique reconnue par la communauté dont elles proviennent
- L'archivage garantit une conservation des données pour au moins 30 ans
- L'archivage numérique pérenne des documents électroniques consiste à conserver le document et l'information qu'il contient :
 - Dans son aspect physique comme dans son aspect intellectuel
 - Sur le très long terme
 - De manière à ce qu'il soit en permanence accessible et compréhensible



Source :

CINES - L'archivage numérique, qu'est-ce que c'est ? -

<https://www.cines.fr/archivage/un-concept-des-problematiques/>

ARCHIVAGE PÉRENNE PLATEFORME D'ARCHIVAGE DU CINES

- Le CINES est l'opérateur mandaté par le Ministère pour opérer la mission d'archivage pérenne pour l'Enseignement Supérieur et la Recherche. Il développe différentes solutions, en particulier la plateforme d'archivage pérenne PAC
- Selon son institution, sa discipline ou l'entrepôt choisi, il existe déjà des partenariats avec le CINES, proposant un accompagnement pour l'archivage.

Ex : Huma-Num en SHS

ARCHIVAGE PÉRENNE PRÉPARATION DES DONNÉES À ARCHIVER

- Sélectionner les jeux de données (et métadonnées associées) à conserver à long terme (peuvent être différents des jeux de données partagés)
- Certaines données ne peuvent être archivées qu'après un traitement adapté
Exemples :
 - Données personnelles nécessitant une anonymisation
 - Données sensibles : des conditions d'accès doivent être ajoutées
- Vérifier la validité des formats de fichiers de données avec l'outil FACILE mis en place par le CINES
- Documenter également les logiciels permettant l'accès aux données
- Compléter et enrichir si besoin les métadonnées
 - Les données doivent posséder une description minimale imposée par le CINES

Ressources :

- Référentiel de gestion des archives de la recherche
- Le concept d'archivage numérique pérenne
- Les formats de fichiers
- Software Heritage : archive de logiciels



Tester ses fichiers avec l'Outil FACILE du CINES : opération de contrôle de validité et/ou d'éligibilité générales.

Exemple : 2 étapes pour les fichiers PDF :

- Correction avec PDFtk : il peut subsister une croix rouge dans la case « Valide ». Il faut passer à l'étape suivante avec l'outil Ghostscript pour avoir une interopérabilité optimale.
- Conversion en PDF/A à l'aide de Ghostscript (logiciel libre permettant le traitement des formats de fichiers PostScript et PDF) : Le CINES recommande de visualiser cette nouvelle version pour vérifier qu'elle est conforme à l'originale.

Sources :

Huma-Num - Archivage à long terme - <https://www.huma-num.fr/services-et-outils/archiver>

CINES – FACILE - <https://facile.cines.fr/>



RÉUTILISATION ET VALORISATION DES DONNÉES



RÉUTILISATION ET CITATION DES DONNÉES

- Choisir une licence de diffusion adaptée à ses besoins
- Respecter la propriété intellectuelle des auteurs telle que mentionnée dans la licence
- Citer les données si la licence l'exige (il est recommandé de toujours citer ses sources)
- Lier les données aux publications

Ressources pour rechercher des jeux de données :

- [Metadata Search](#), [re3data](#), [OAD](#), [OpenDOAR](#)



Dès lors que les données issues d'une activité de recherche financée au moins pour moitié par des dotations de l'État, 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. Sauf accord de l'administration, la réutilisation des informations publiques est soumise à la condition que ces dernières ne soient pas altérées, que leur sens ne soit pas dénaturé et que leurs sources et la date de leur dernière mise à jour soient mentionnées.

Source :

DIST-CNRS. Le travail de la science et le numérique : données, plateformes, Publications. Une analyse systémique de la Loi numérique. 24 janvier 2017.

http://www.cnrs.fr/dist/z-outils/documents/20170203_analyse%20syst%C3%A9mique_vf.pdf

DATA JOURNAL – DATA PAPER

- Un **data paper** est une publication qui décrit des jeux de données de recherche
C'est un article à part entière, suivant le même processus éditorial que les articles scientifiques classiques
Il permet d'accéder à tous les éléments d'un projet
Ex. de data paper : [Tracking vegetation phenology across diverse North American biomes using PhenoCam imagery](#)
- Deux possibilité de publication d'un data paper :
 - dans un **data journal** (revue dédiée à ce type de publication)
 - dans une revue classique
- Ex. de data journal : [Journal of Physical and Chemical Reference Data \(AIP Publishing\)](#)

Ressources :

- [Data papers et data journals](#)
- [Exemples de revues publiant des data papers](#)



Structure d'un data paper :

Partie descriptive :

- Éléments communs aux articles classiques : titre, résumé, mots-clés...
- Éléments spécifiques aux données : types de données, formats, processus et méthodes de production, métadonnées, réutilisation...

Accès aux données : intégrées dans l'article ou déposées dans un entrepôt

L'identifiant des données (exemple DOI) permet d'établir le lien du data paper vers les données.

EXPOSITION ET VISUALISATION DES DONNÉES

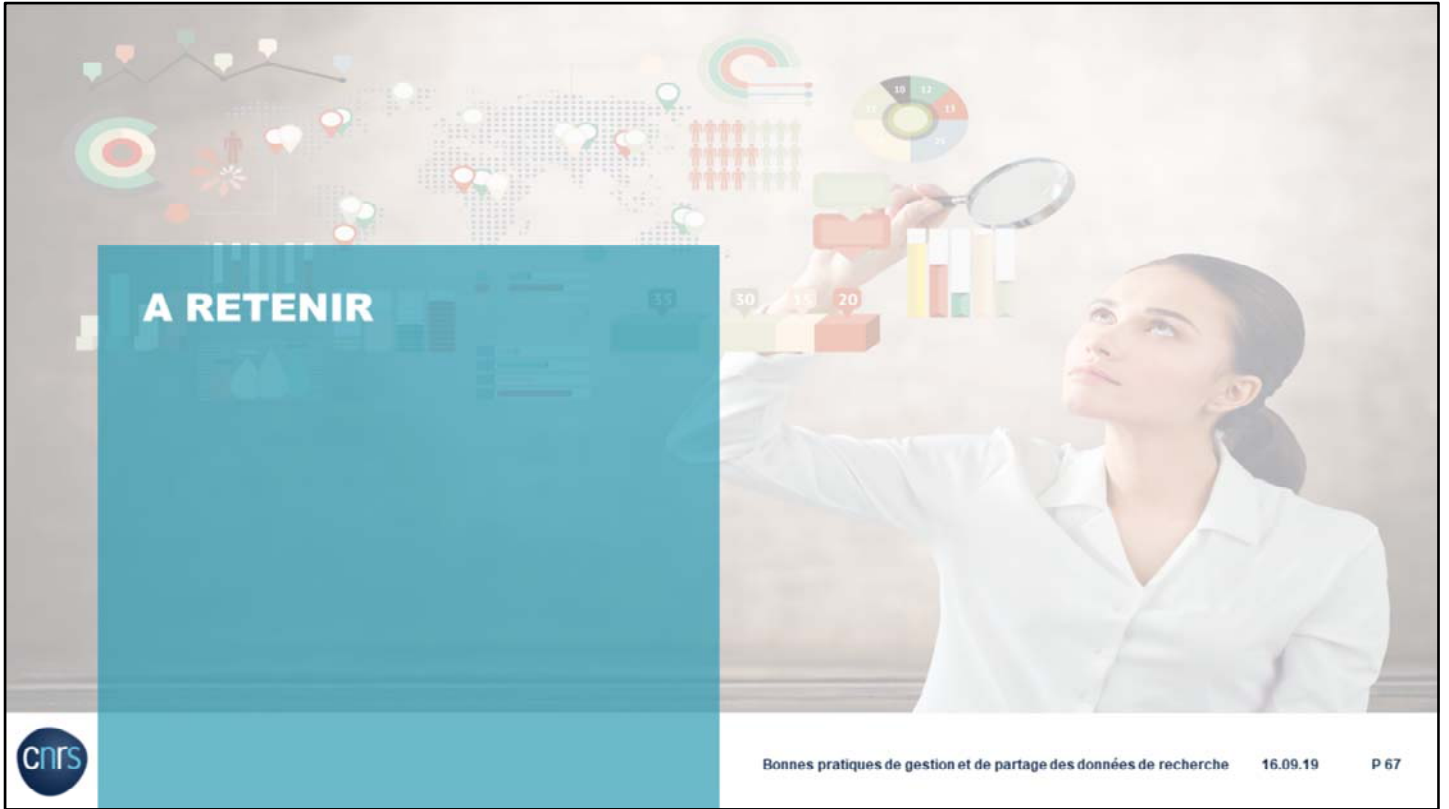
- En complément du dépôt dans un entrepôt, il peut s'avérer utile, surtout dans le cas de données nombreuses et complexes, de les exposer sous forme visuelle (cartographies, graphiques, etc.) via une plateforme
- L'outil Omeka est un logiciel libre couramment utilisé

The screenshot shows a webpage with a red header containing the title "LA MINUTE ACCÉDER AUX DONNÉES PAR LA VISUALISATION" and the date "PUBLIÉ LE 27/11/2017 (MIS À JOUR LE 28/05/2018)". Below the header, there is a navigation bar with "AUTEUR" (Océanica) and "CATÉGORIES + TAGS" (Accès et visualisation, Vidéo, Réseaux). The main content area is titled "OBJECTIFS" and lists two bullet points: "• Décrire l'impact de la visualisation dans les outils aux données de recherche" and "• Analyser le rôle de la visualisation dans les outils aux données de recherche". Below this, there is a question: "En quoi la visualisation facilite-t-elle l'accès aux données? Dans quels cas la visualisation est-elle nécessaire?". A video player is embedded, showing a network diagram with nodes and connections. To the right of the video player, there is a search bar and a "Rechercher" button. Below the video player, there are three sections: "QUELQUES EXEMPLES D'ENTREPÔTS AYANT RECOURS À LA VISUALISATION" (listing ODS, DataCamp, and DataCamp), "QUELQUES EXEMPLES DE PROJETS UTILISANT LA VISUALISATION" (listing DataCamp, DataCamp, and DataCamp), and "QUELQUES OUTILS DE VISUALISATION" (listing Gephi, Tableau, and Tableau).

PLATEFORME COREA : EXEMPLE D'UTILISATION D'OMEKA

- Bibliothèque numérique pour la documentation archéologique du Centre Camille Julian
- Permet de naviguer dans des corpus et ressources en archéologie

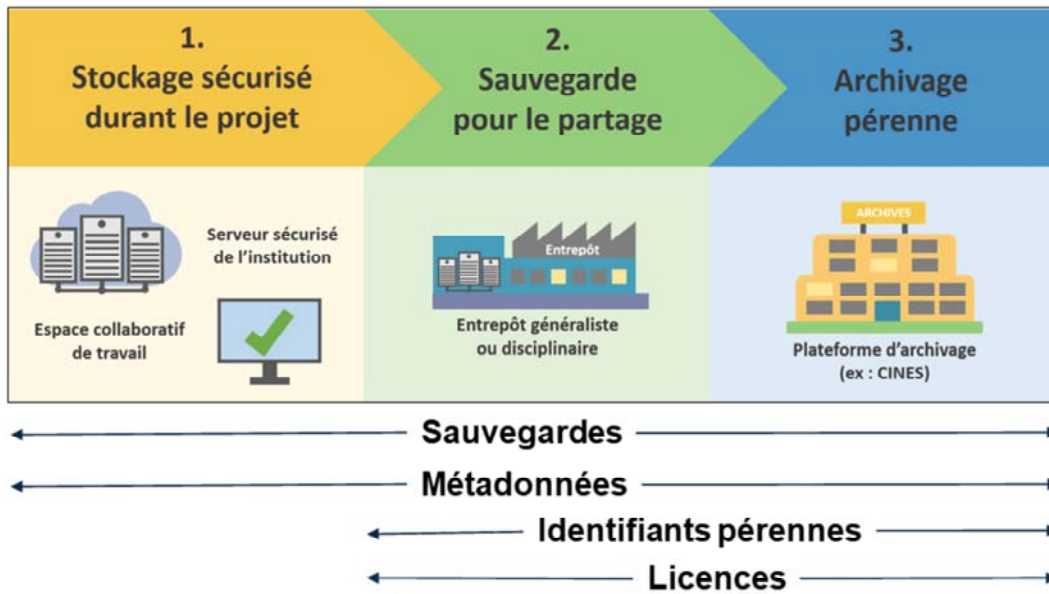




A RETENIR



3 ÉTAPES DE SAUVEGARDE DES DONNÉES





WEBOGRAPHIE

- Arnould Pierre-Yves, Jacquemot-Perbal Marie-Christine. Guide de bonnes pratiques : Gestion et valorisation des données de recherche. 1er février 2016. <https://ordar.otelo.univ-lorraine.fr/record?id=10.24396/ORDAR-1>
- Cocaud Sylvie , L'Hostis Dominique. Pourquoi et comment rédiger un plan de gestion de données ? 5 avril 2019. <https://prodinra.inra.fr/?locale=fr#!ConsultNotice:447192>
- Delplanque Catherine , Lamrini Nawale, Leclère Fabrice, Maurel Lionel, et al. Fiches pratiques à destination des chercheurs sur le Règlement Général pour la Protection des Données. <http://www.u-plum.fr/actualites/467-fiches-pratiques-sur-le-reglement-general-pour-la-protection-des-donnees>
- Durand-Barthez Manuel. Les données de la Recherche. 17 avril 2018. <http://urfist.chartes.psl.eu/ressources/les-donnees-de-la-recherche>

WEBOGRAPHIE

- Ginouvès Véronique, Gras Isabelle, et al. La diffusion numérique des données en SHS – Guide des bonnes pratiques éthiques et juridiques. Octobre 2018 <https://hal-amu.archives-ouvertes.fr/page/guide-de-bonnes-pratiques>
- Inist-CNRS, GIS Urfist – DoRANum - <https://doranum.fr/>
- Laï Paolo. Le cycle de vie des données de la recherche. Séminaire Intégrité et partage de la science Université Grenoble Alpes. 10-13 décembre 2018. https://datadoct2018.sciencesconf.org/data/program/UGA_SummerSchool_IPS_20181210_LAI.pdf
- Maurel Lionel. La réutilisation des données de la recherche après la loi pour une République numérique. Décembre 2017. <https://hal.archives-ouvertes.fr/hal-01908766>

WEBOGRAPHIE (SUITE)

- Rivet Alain, Bachèlerie Marie-Laure, Denis-Meyere Auriane, Tisserand Delphine - Traçabilité des activités de recherche et gestion des connaissances - Guide pratique de mise en place – 2018 - http://qualite-en-recherche.cnrs.fr/IMG/pdf/guide_tracabilite_activites_recherche_gestion_connaissances.pdf
- Science Europe. Guide pratique pour une harmonisation internationale de la gestion des données de recherche. juillet 2019. <https://www.ouvrirlascience.fr/science-europe-guide-pratique-pour-une-harmonisation-internationale-de-la-gestion-des-donnees-de-recherche/>
- Stérin Anne-Laure. Diffuser des données de la recherche dans le respect du droit et de l'éthique – Comment faire lorsqu'on n'est pas juriste ? octobre 2018. <https://hal-amu.archives-ouvertes.fr/hal-02050510>



Merci de votre attention

contact@doranum.fr

info-opidor@inist.fr

<https://doranum.fr>

<https://dmp.opidor.fr>

www.cnrs.fr

