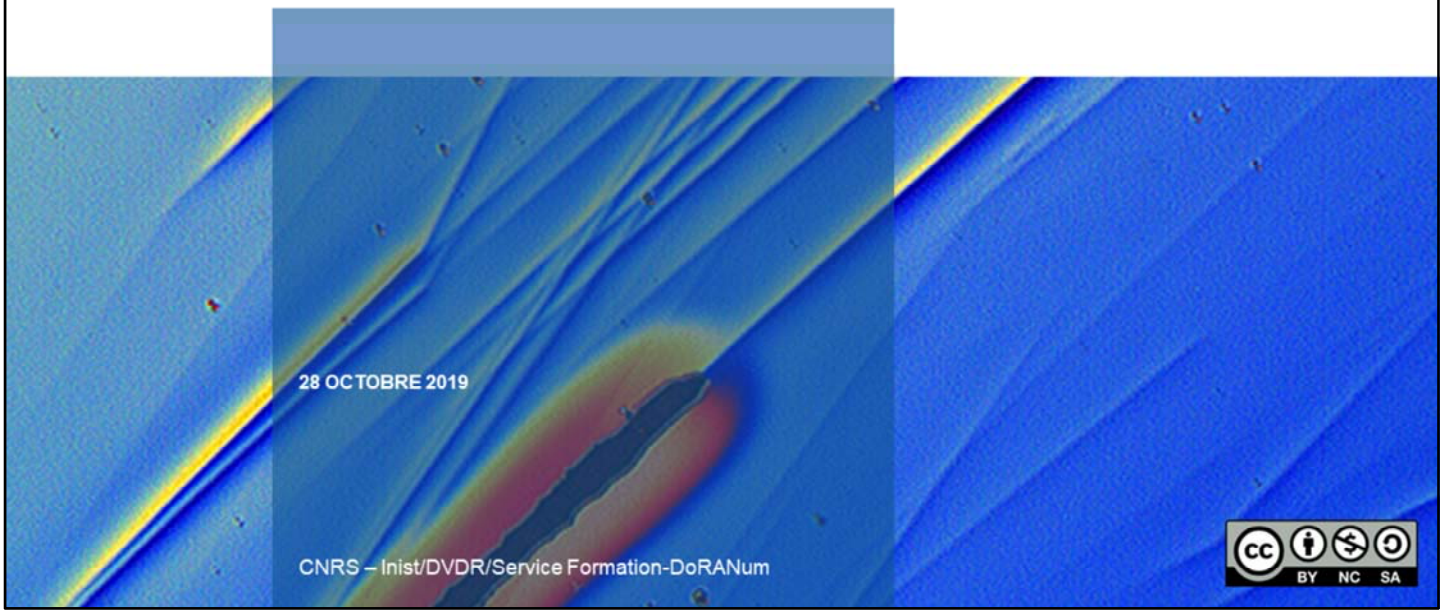
Introduction au cycle de vie des données, aux principes FAIR et au Plan de Gestion des Données (PGD)/ DMP

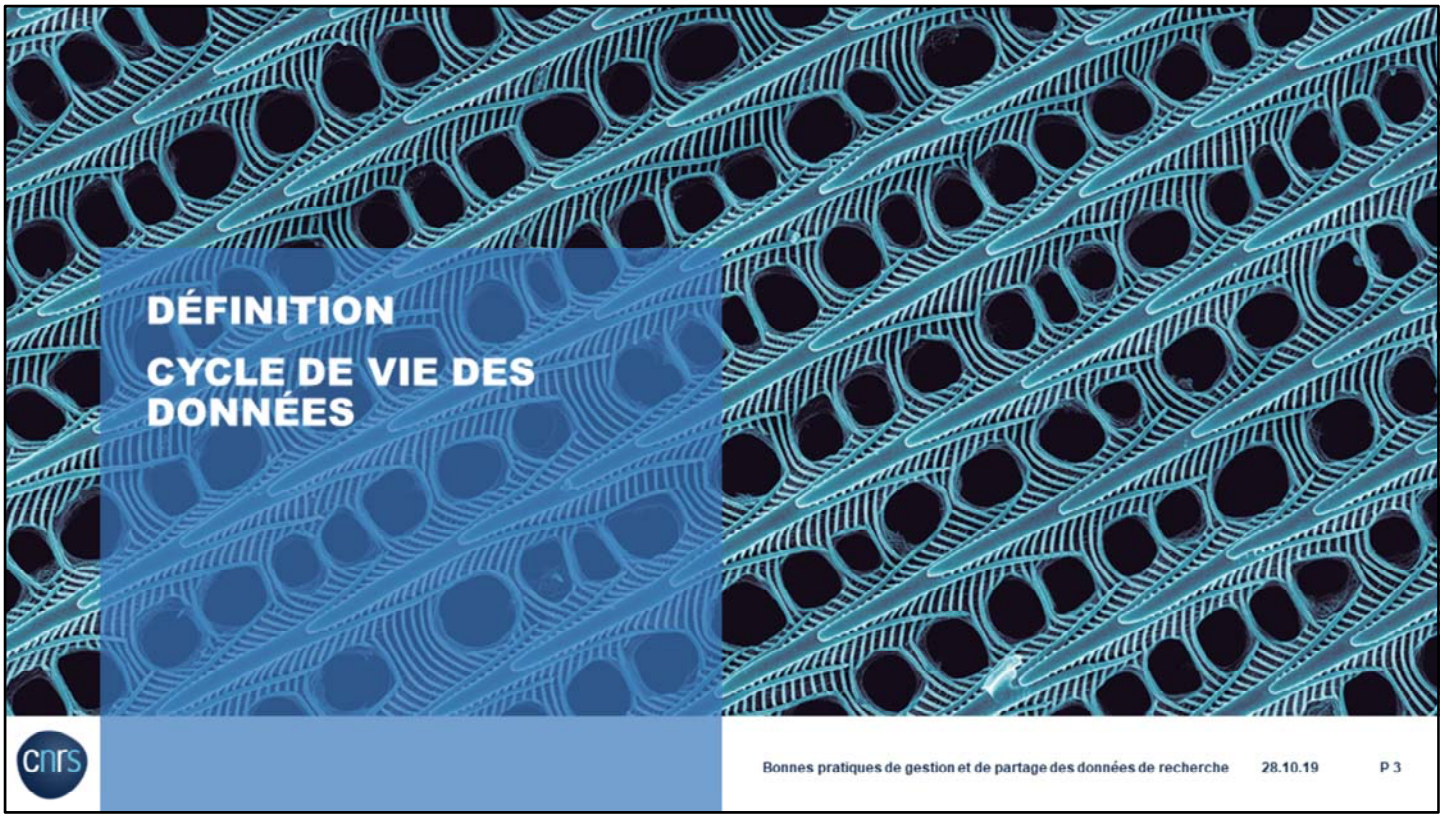


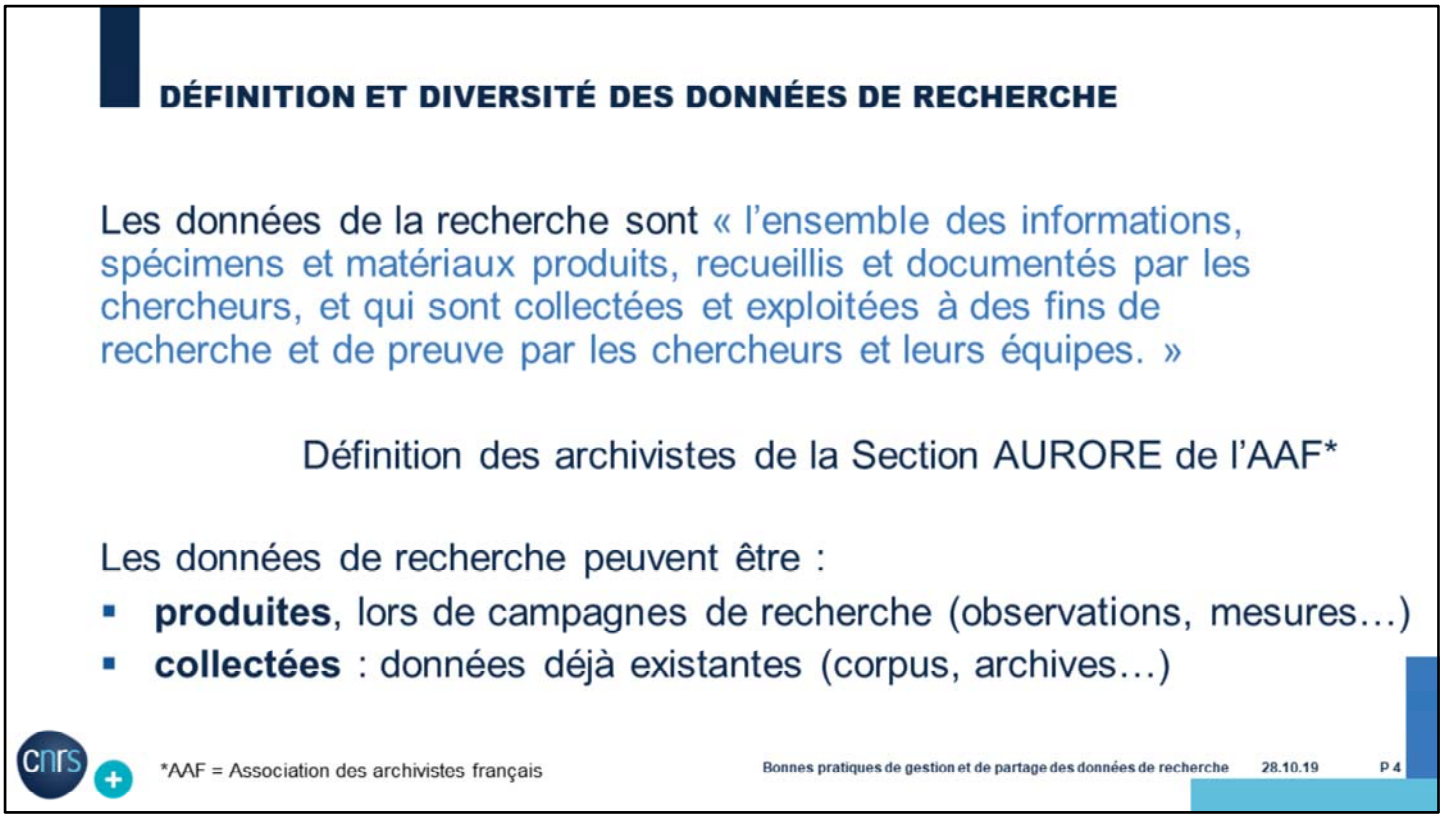
28 OCTOBRE 2019

CNRS – Inist/DVDR/Service Formation-DoRANum

SOMMAIRE

- Définition et cycle de vie de données
- Principes FAIR
- Plan de Gestion des Données de recherche (PGD / DMP)
- Aspects juridiques et éthiques, modalités de protection et d'accès
- Création, collecte et description des données de recherche
- Métadonnées
- Identifiants pérennes
- 3 étapes de sauvegarde tout au long du projet
- Stockage durant le projet
- Partage, diffusion des données, dépôt dans un entrepôt
- Archivage pérenne
- Réutilisation et valorisation des données
- A retenir

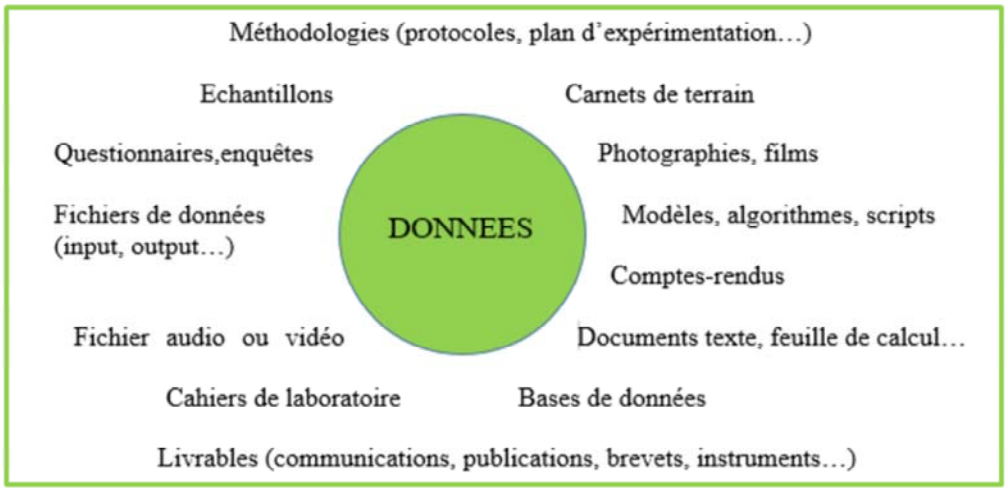

DÉFINITION
CYCLE DE VIE DES DONNÉES
cnrs
Bonnes pratiques de gestion et de partage des données de recherche 28.10.19 P 3



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. »

Ressources :
- 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/
- 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, données géospatiales, images, photographies, langages de programmation, matérielles et physiques, modèles, visualisations, 3D, numériques, textuelles, numérisations, scans, qualitatives, quantitatives, statistiques…

Ressources :

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/
Wikipedia – Données de la recherche - https://fr.wikipedia.org/wiki/Donn%C3%A9es_de_la_recherche

CYCLE DE VIE DES DONNÉES DE RECHERCHE

C'est l'ensemble des étapes
- de gestion,
- de conservation
- et de diffusion des données de recherche,

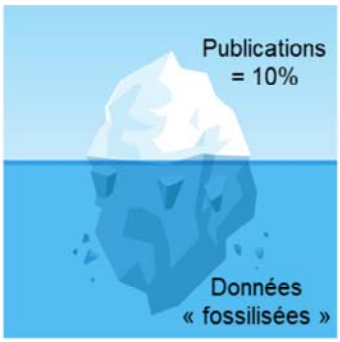
associées aux activités de recherche

D'après Research data lifecycle – UK Data Service
https://www.ukdataservice.ac.uk/manage-data/lifecycle

Source :
D'après Research data lifecycle – UK Data Service
https://www.ukdataservice.ac.uk/manage-data/lifecycle

POURQUOI GÉRER ET PARTAGER SES DONNÉES

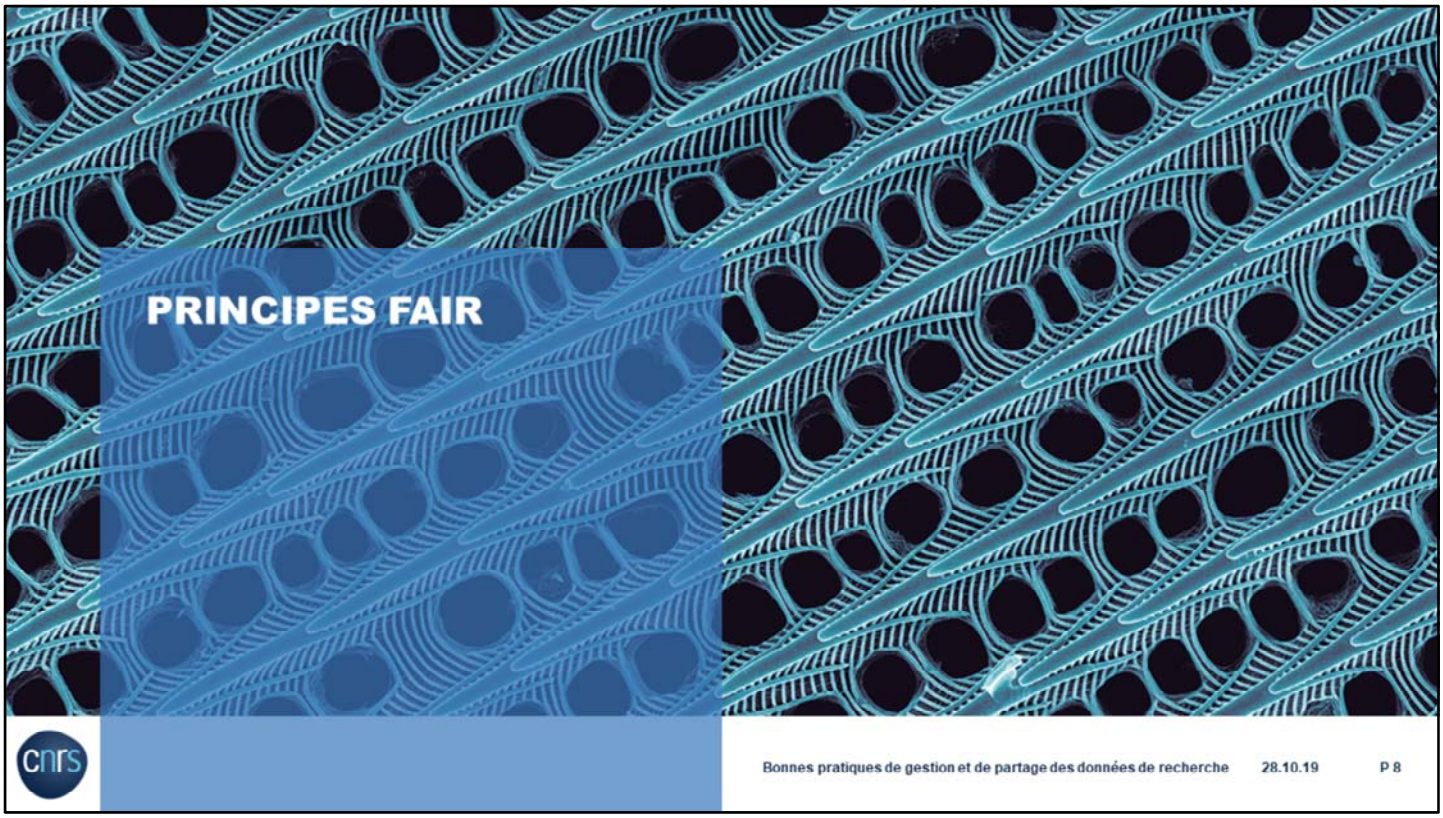
- Gestion nécessaire face à l'accroissement de la quantité de données
- Exhumation de données « fossilisées »
- Evite la perte de données uniques
- Gain de temps et d'argent
- Facilite la reproductibilité, la réutilisation et le croisement de données provenant de différentes disciplines



"Designed by wstudio / Freepik"

- 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…

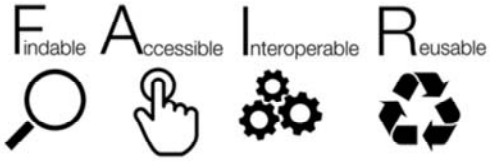
Source :
Durand-Barthez Manuel. Les données de la Recherche. 17 avril 2018.


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

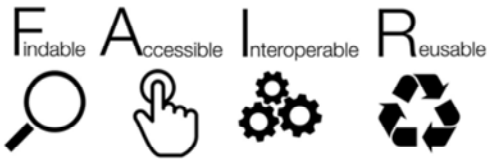


Principes admis par les différentes communautés scientifiques au niveau international, ainsi que par les financeurs (ex : Commission européenne, ANR, etc.)

Applicables tout au long du cycle de vie des données

PRINCIPES FAIR

« Aussi ouvert que possible,
aussi fermé que nécessaire »



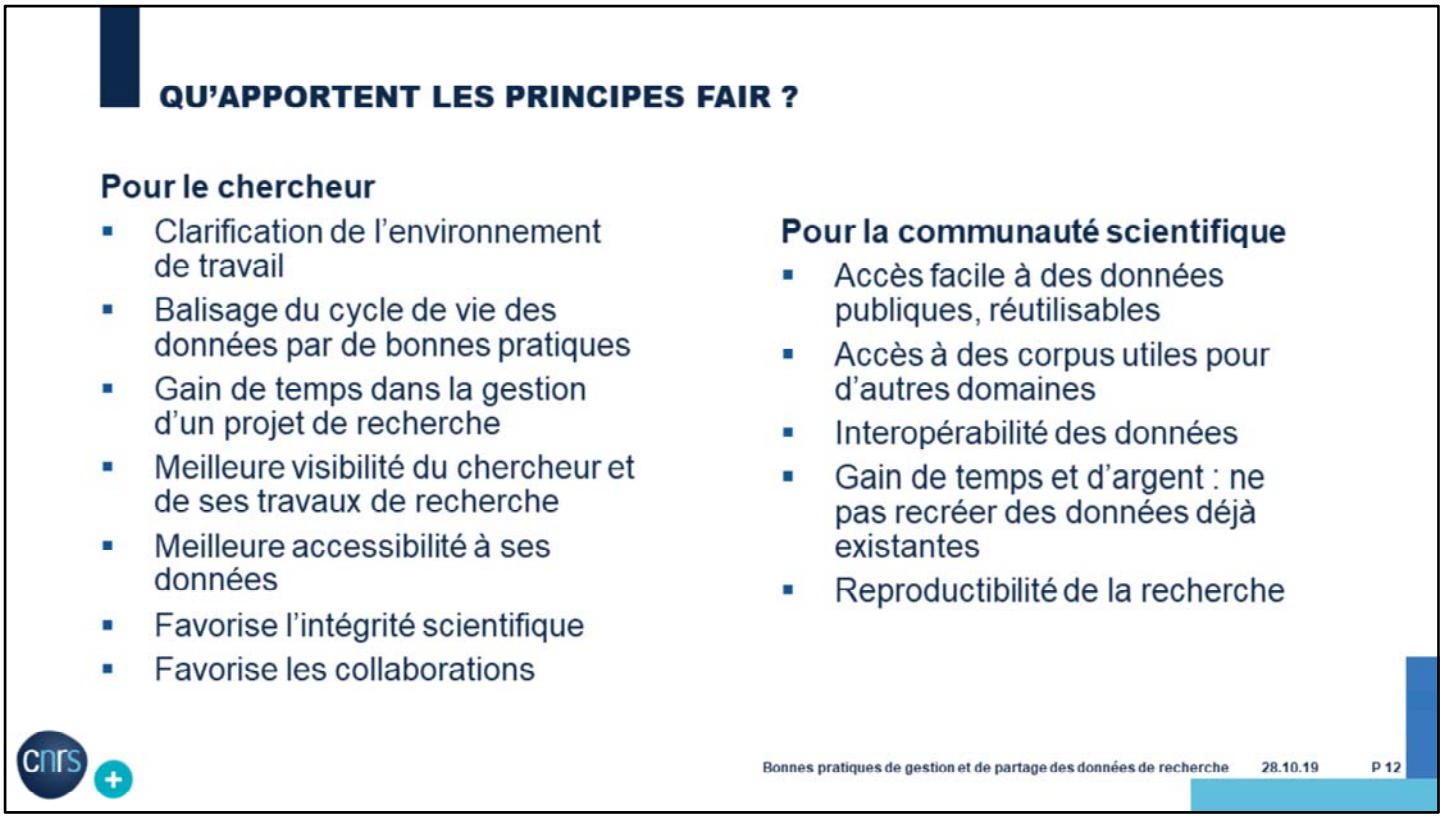
Principes FAIR :
Facile à trouver : facilite la découverte des données par les humains et les systèmes informatiques
- Données et métadonnées identifiées par un identifiant global unique et pérenne (ex. DOI)
- Métadonnées riches pour décrire les données (standards de métadonnées disciplinaires)
- Données 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



Ressources :

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

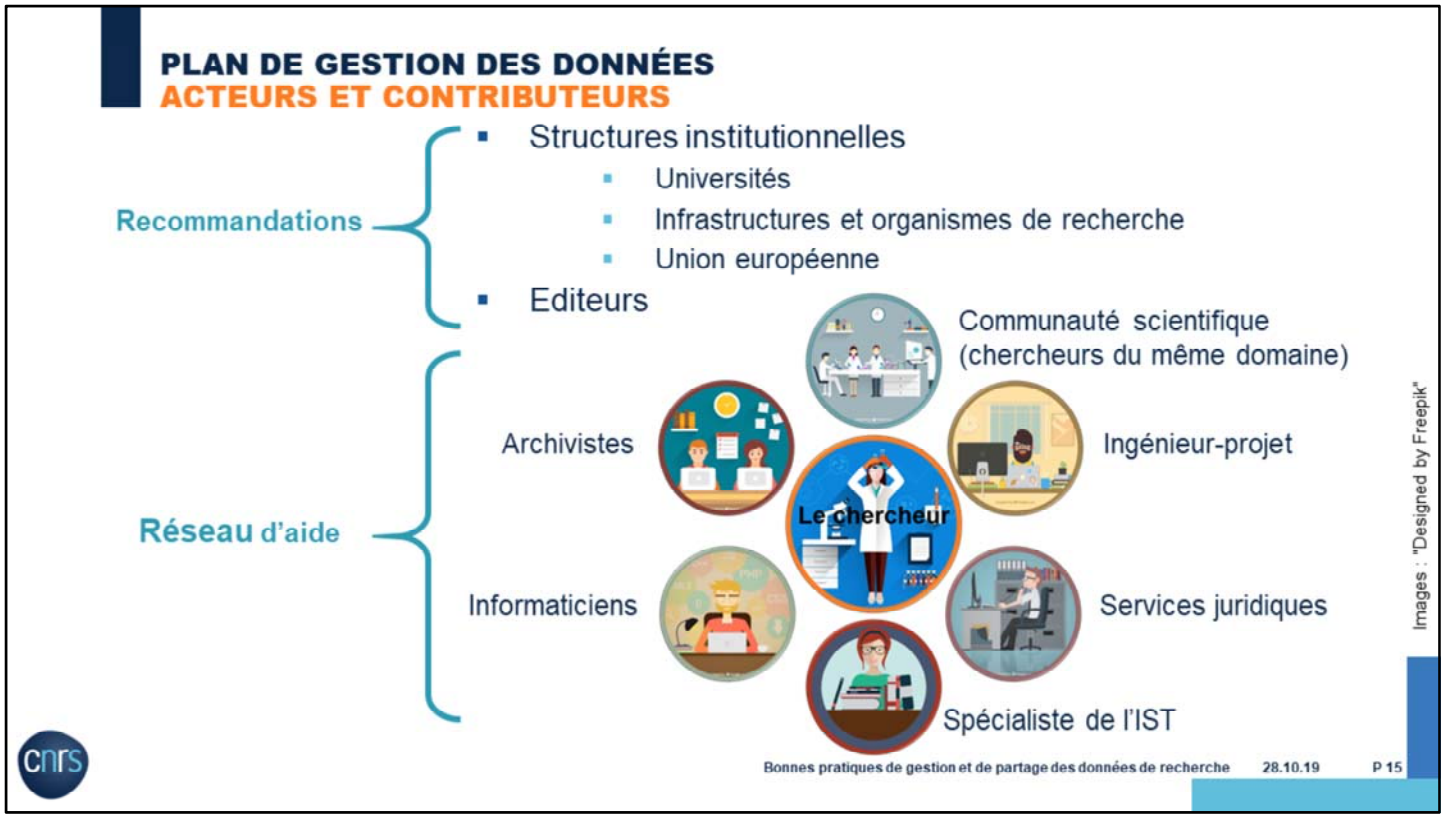

PLAN DE GESTION DES
DONNÉES DE RECHERCHE
(PGD OU DMP)
cnrs
Bonnes pratiques de gestion et de partage des données de recherche 28.10.19 P 13

PLAN DE GESTION DES DONNÉES
IMPORTANCE

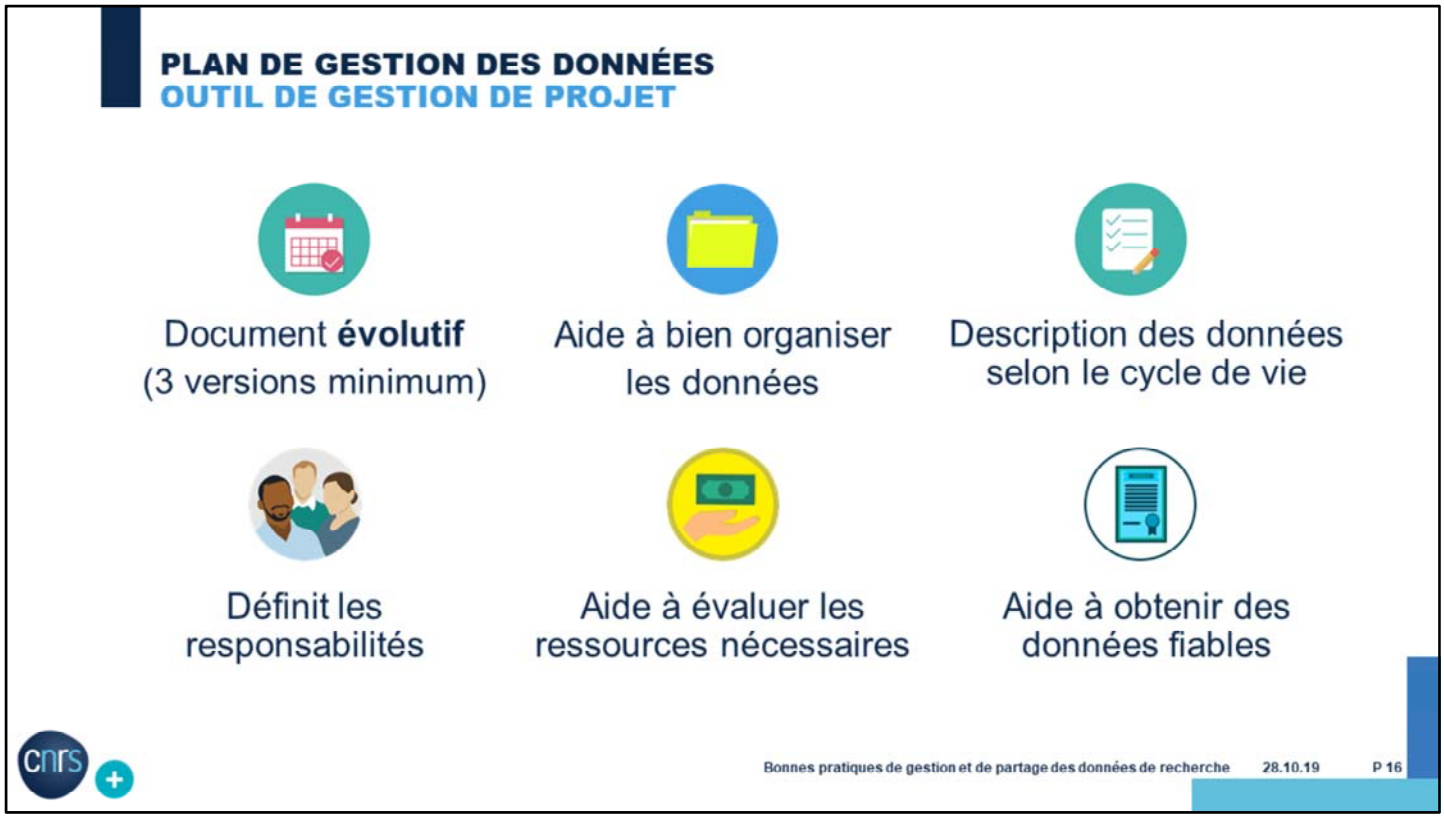
- **Echelle mondiale :** DMP de plus en plus recommandé ou exigé
- **Echelle européenne :** obligation pour Horizon 2020
- **En France :** obligation pour l'ANR
- **Echelle des organismes :** recommandations, politique d'établissements…
- **Echelle disciplinaire :** modèles de DMP par domaine

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 :
  - Financeur 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…)



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



- Le DMP est un document évolutif. Il faut commencer à le rédiger dès le début du projet, avec les éléments déjà connus ou prévus. D'ailleurs, il peut être demandé dès la soumission du projet. Ensuite compléter le DMP au fur et à mesure du projet.
Prévoir 3 versions au minimum :
    - Au début du projet
    - Au milieu du projet
    - A la fin du projet.

- Dans le DMP, 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 :
    - saisie des données,
    - production des métadonnées,
    - contrôle de la qualité des données,
    - stockage, partage et archivage des données,
    - mise à jour du DMP.

- Le DMP aide à bien organiser les données, tout au long du projet

- Il est demandé d'évaluer les ressources nécessaires (budget, temps alloué, personnels) permettant la mise en œuvre des actions décrites dans le DMP :
    - temps nécessaire à la préparation des données pour le stockage, le partage et l'archivage des données
    - coûts de matériel, rémunération des personnels
    - frais de stockage (serveurs dédiés, traitement, maintenance, sécurité, accès…), partage (site web, publication…) et d'archivage des données

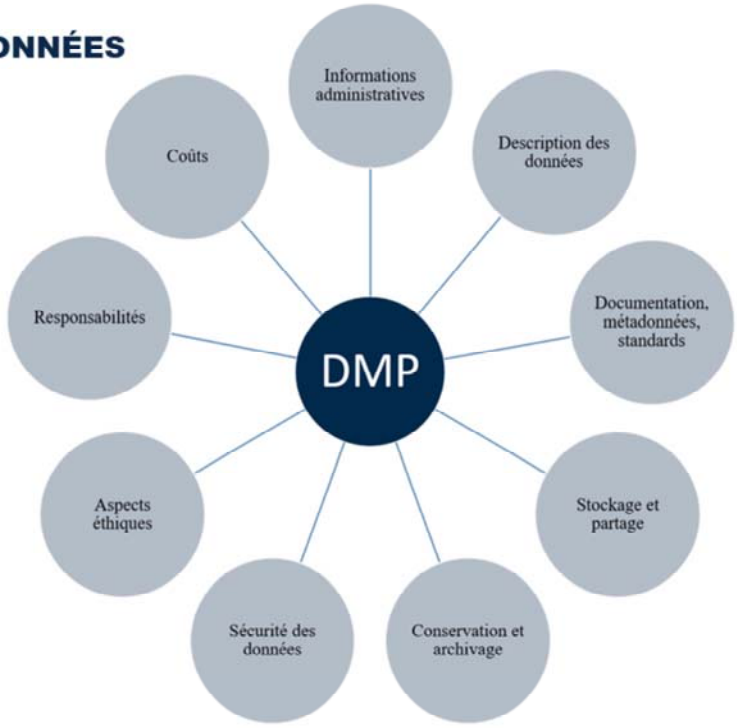
- Dans le DMP, il faut décrire la façon dont les données seront obtenues, traitées, organisées, stockées, sécurisées, préservées, partagées… (cycle de vie des données)

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

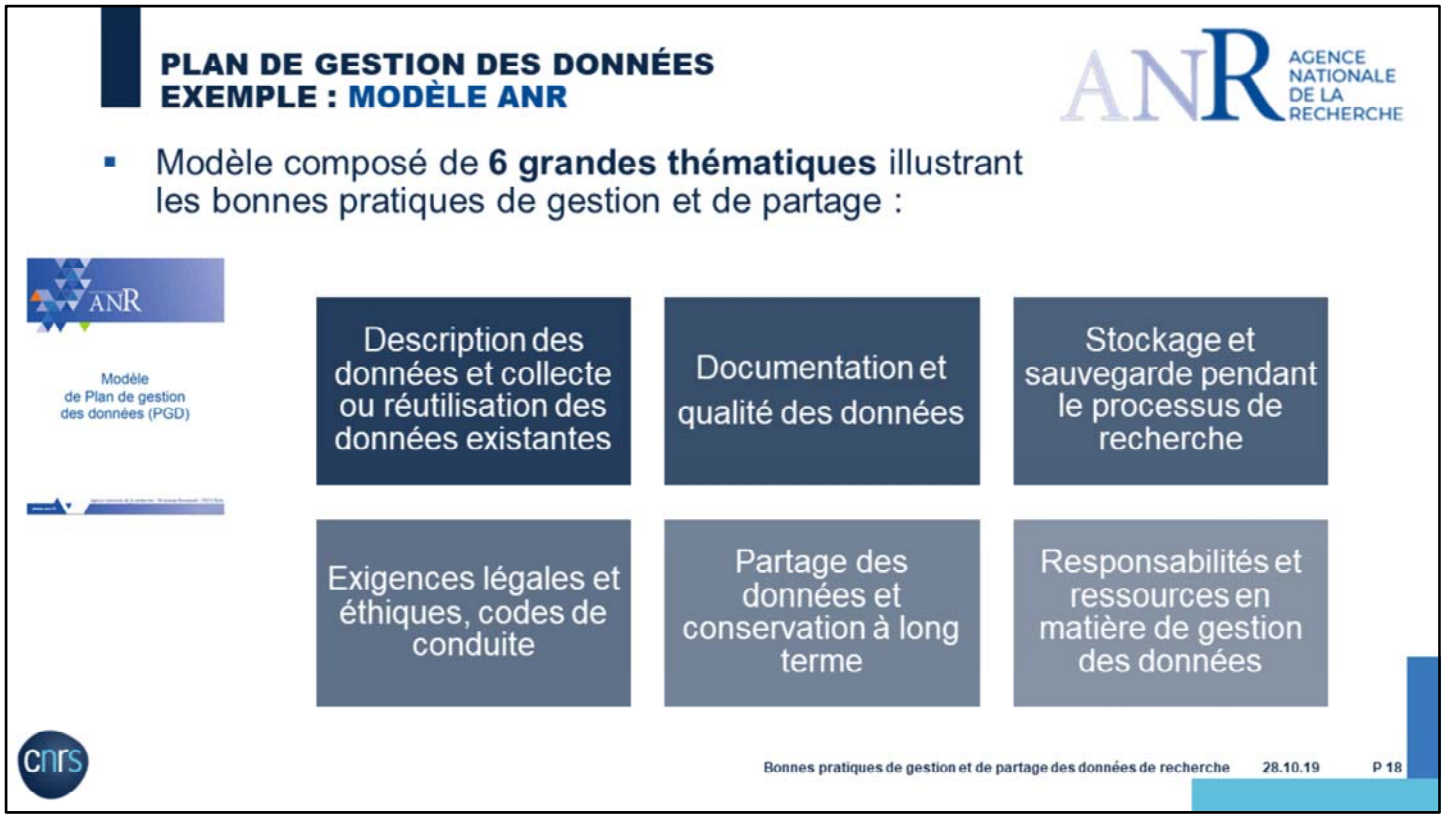
Ressources :
- UK Data Service – Data management costing tool and checklist : https://data-archive.ac.uk/media/247429/costingtool.pdf
- DoRANum - plan de gestion de données : https://doranum.fr/plan-gestion-donnees-dmp/
- Nathalie Reymonet, Magalie Moysan, Aurore Cartier, Renaud Délémontez – Réaliser un plan de gestion de données FAIR : https://archivesic.ccsd.cnrs.fr/sic_01690547/

PLAN DE GESTION DES DONNÉES
DIFFÉRENTS MODÈLES

- Il n'existe pas de trame unique, mais de nombreux modèles de DMP ont été établis par des organismes, instituts, financeurs à destination de leurs utilisateurs
- On y retrouve les mêmes éléments (schéma)



Ressource :
DoRANum – Plan de gestion de données : https://doranum.fr/plan-gestion-donnees-dmp/



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
OUTIL DMP OPIDOR

DMP OPIDoR

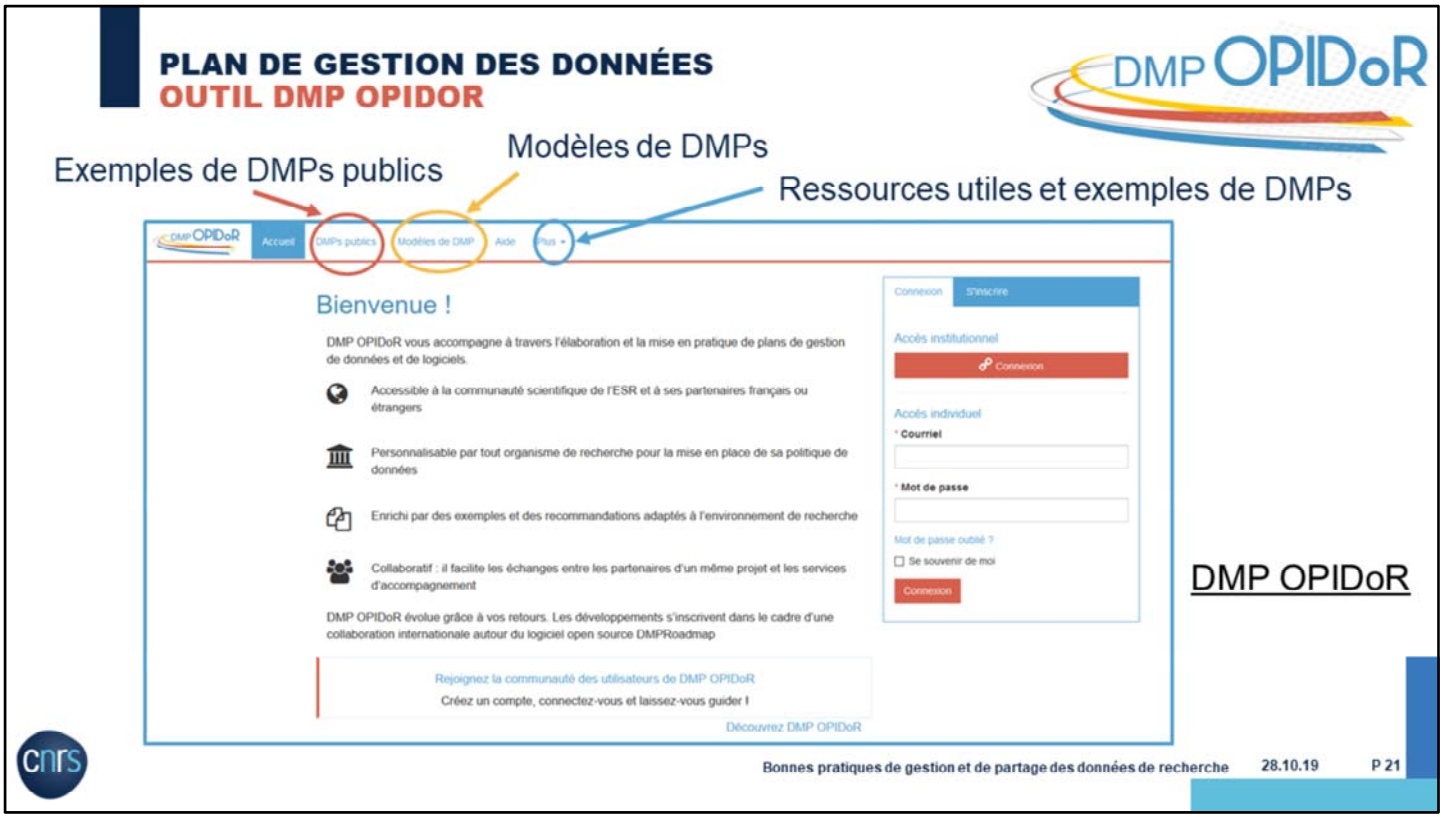
Data Management Plan pour une Optimisation du Partage et de l'Interopérabilité des Données de la Recherche

https://dmp.opidor.fr/

- Outil collaboratif en ligne d'aide à la rédaction de DMP
- Accessible à l'ensemble de la communauté scientifique de l'ESR et à ses partenaires français ou étrangers
- Outil basé sur le code commun DMPRoadmap (Digital Curation Center/UK et l'UC3/USA)

PLAN DE GESTION DES DONNÉES
OUTIL DMP OPIDOR

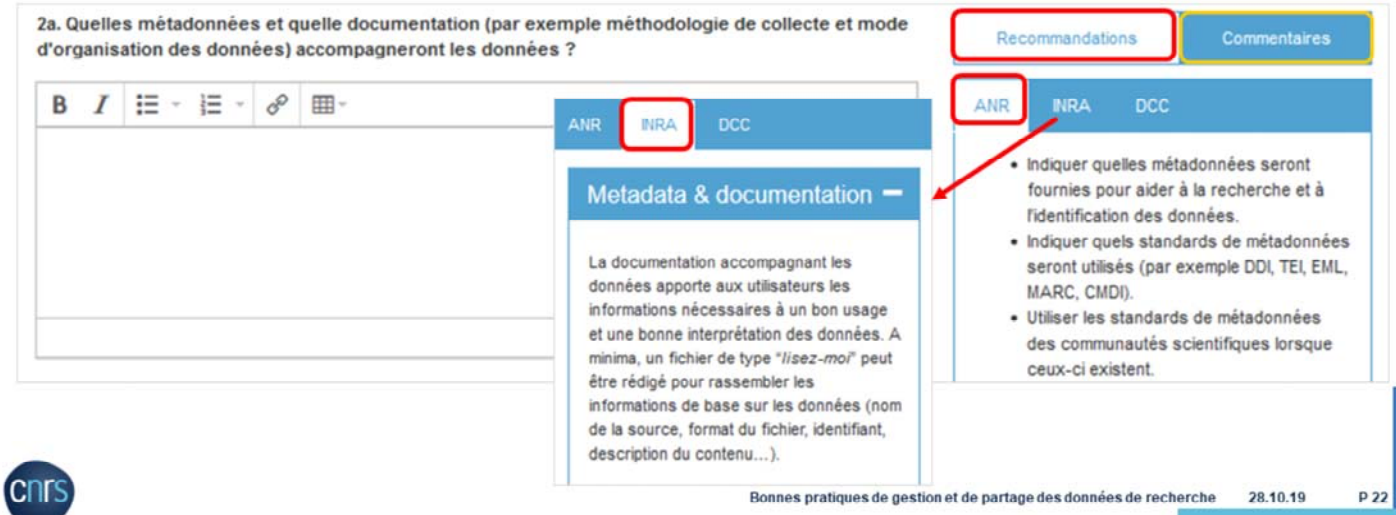
- Création /rédaction d'un DMP à partir d'un modèle (plusieurs modèles existants : ANR, H2020, INRA, CIRAD…)
- Exemples de DMPs
- Partage du DMP avec un / des collaborateur(s)
- Possibilité d'ajout de commentaires par les collaborateurs
- Définition du niveau de visibilité du DMP
- Possibilité de personnalisation pour les organismes de recherche pour la mise en place de leur politique de données
- Possibilité de demander une assistance conseil auprès des services d'appui de son organisme de recherche (s'il existe)
- Téléchargement des DMPs sous différents formats (docx, pdf,…)

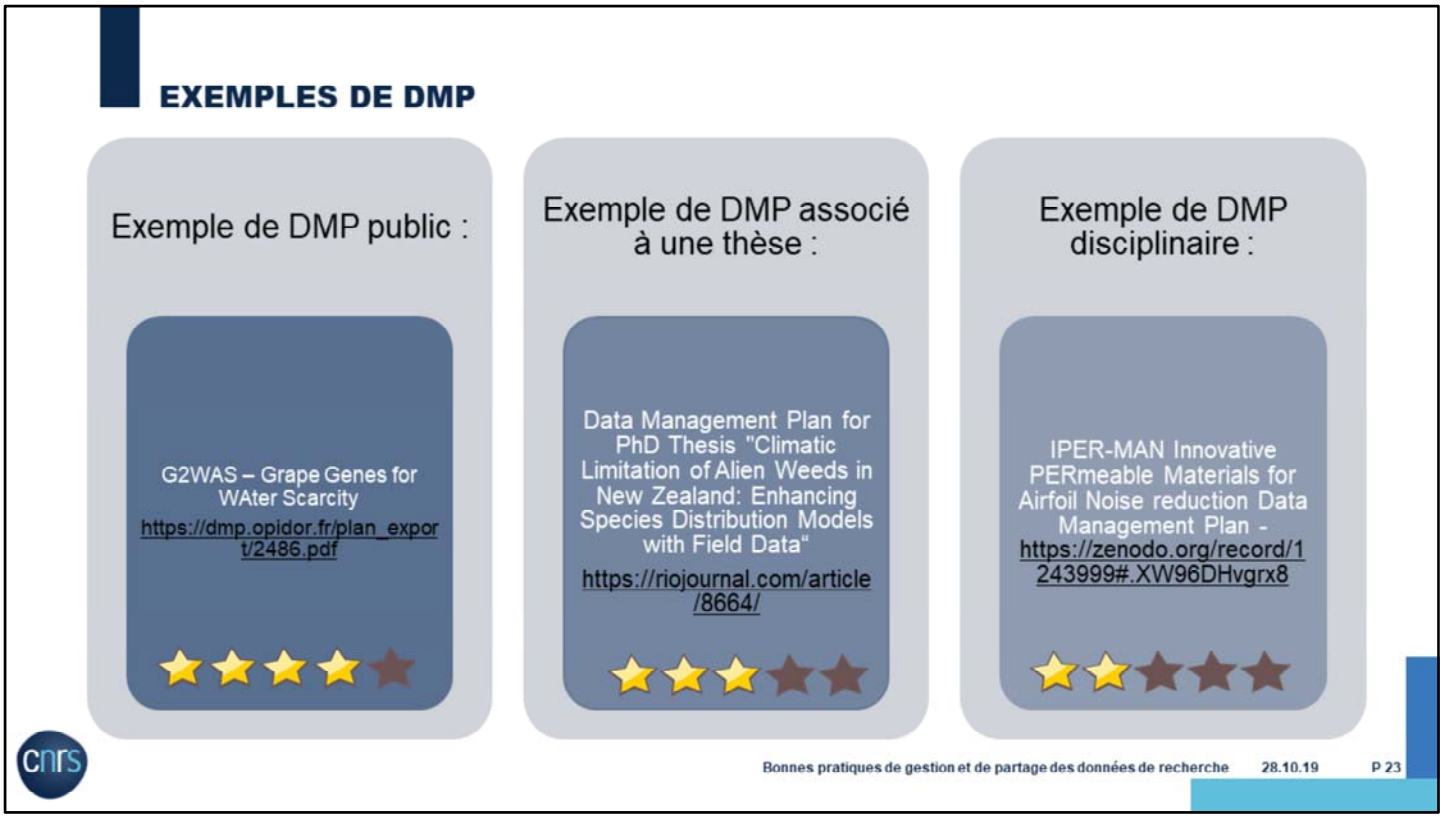


Dans l'onglet « DMPs publics », vous pouvez consulter les DMPS rédigés dans DMP OPIDoR qui ont été rendus publics par leur(s) auteur(s).
Dans l'onglet « Plus », vous pouvez consulter d'autres exemples de DMPs, disponibles par ailleurs, mais aussi accéder à des ressources utiles, guides et autres sites sur la gestion des données de la recherche.

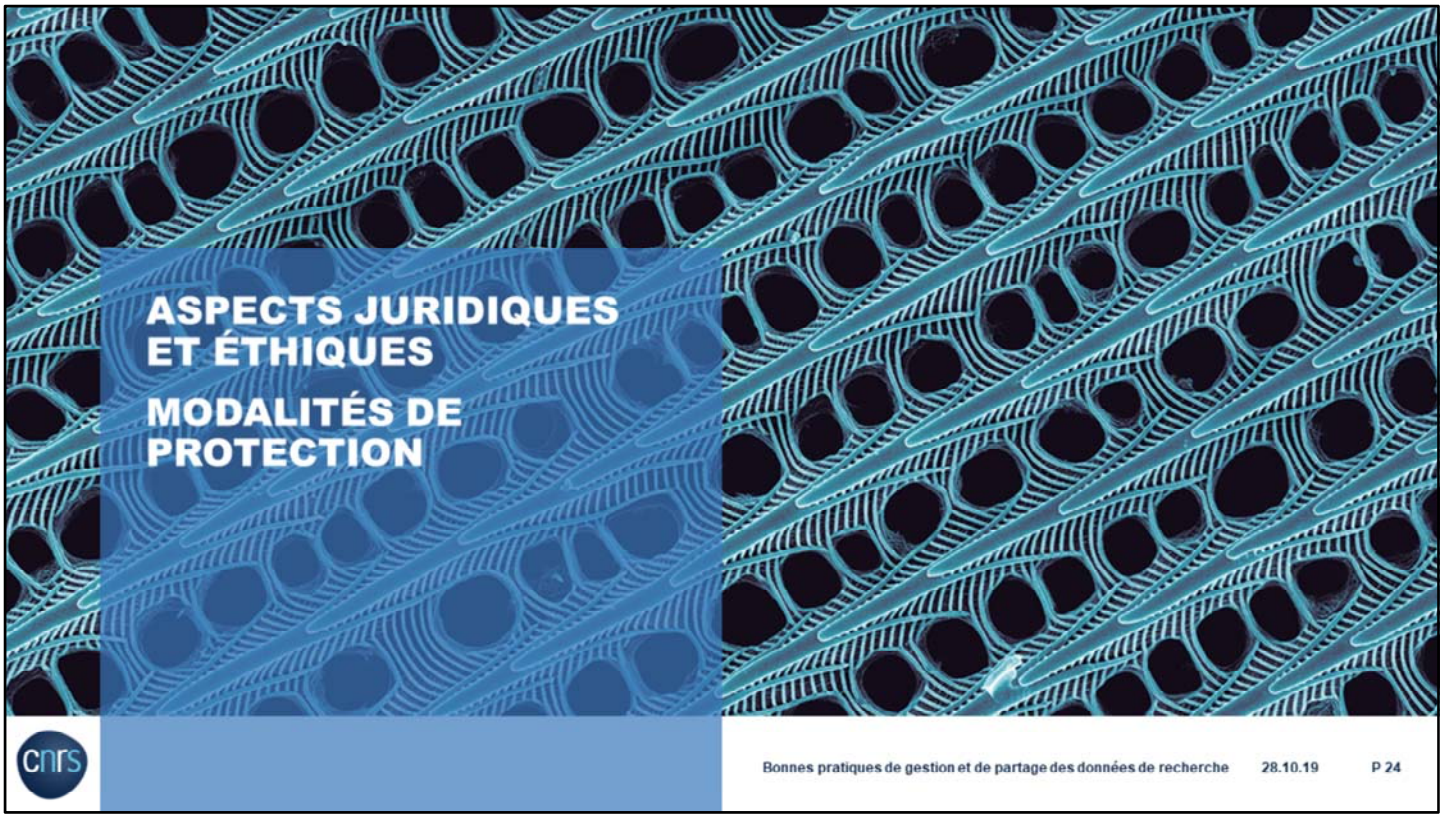
PLAN DE GESTION DES DONNÉES
OUTIL DMP OPIDOR

- Possibilité de consulter les recommandations adaptées à l'environnement de recherche



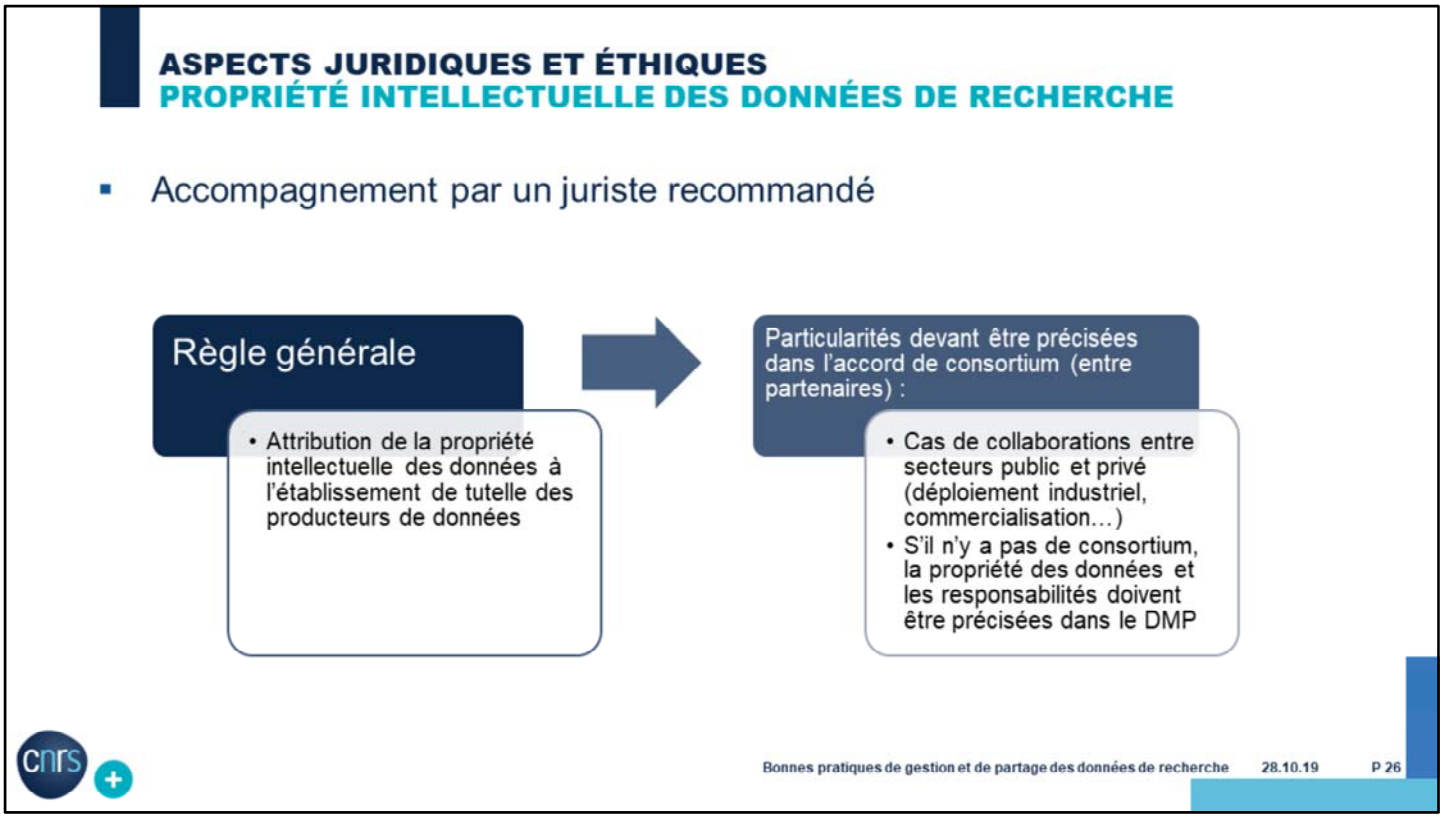


A noter que les DMPs publics rédigés avec DMP OPIDoR ne sont pas évalués, ce sont les auteurs qui les mettent à disposition. Exemple : le DMP G2WAS


ASPECTS JURIDIQUES
ET ÉTHIQUES
MODALITÉS DE
PROTECTION
cnrs
Bonnes pratiques de gestion et de partage des données de recherche 28.10.19 P 24

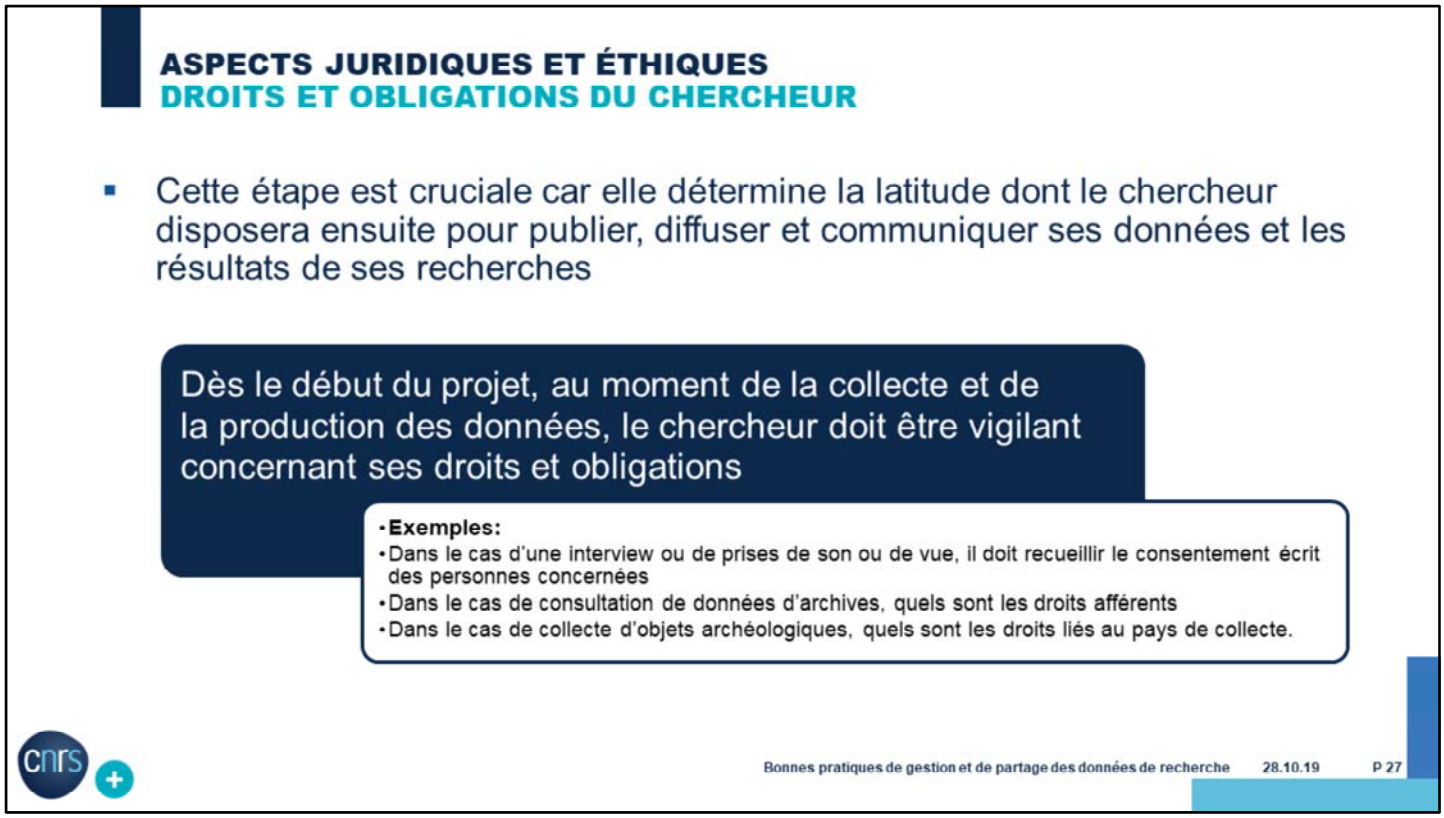
CYCLE DE VIE DES DONNÉES DE RECHERCHE
CRÉATION, COLLECTE ET DESCRIPTION

D'après Research data lifecycle – UK Data Service
https://www.ukdataservice.ac.uk/manage-data/lifecycle



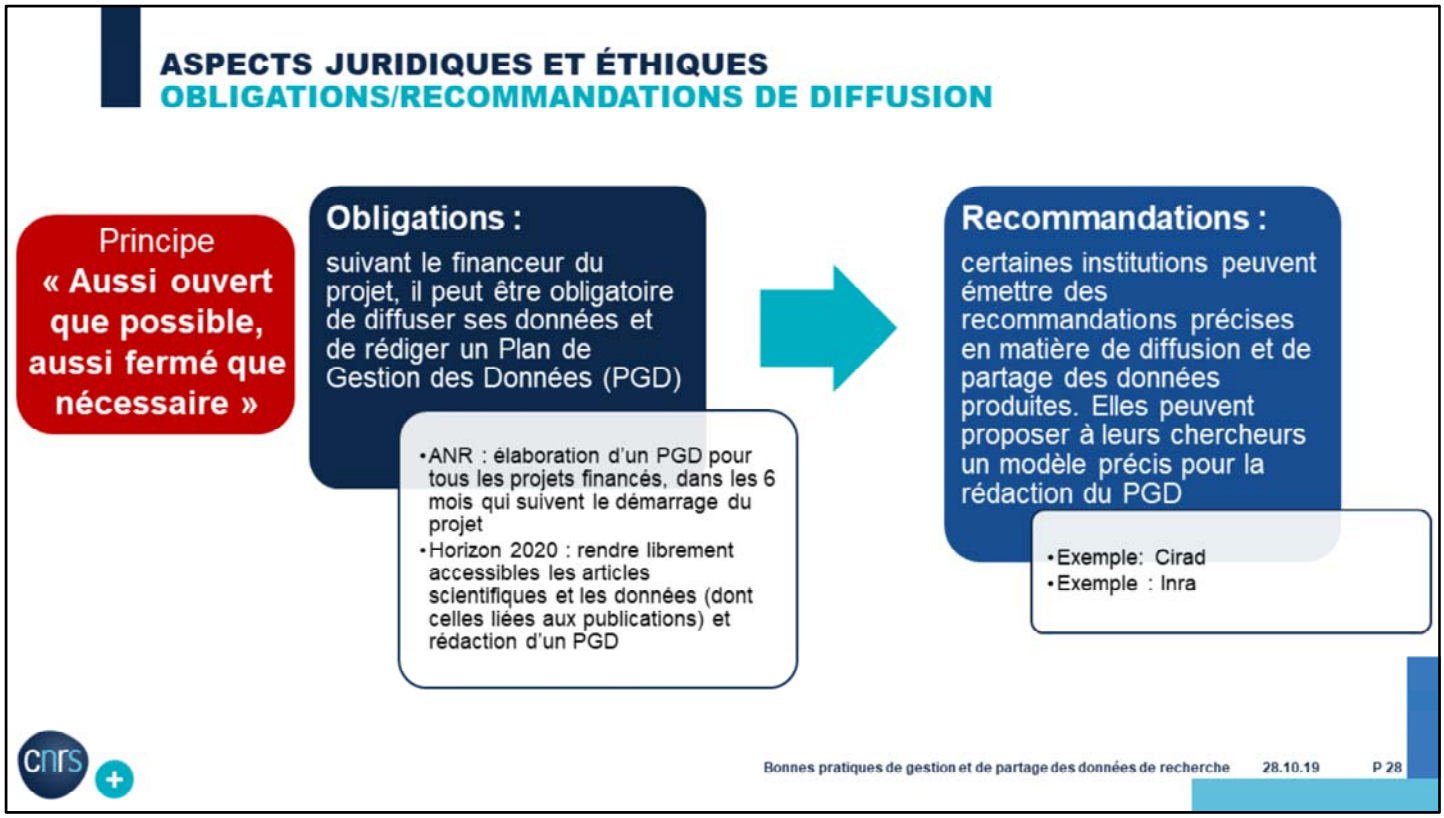
Il ne s'agit pas du même droit que les publications (droit d'auteur principalement). Les données relèvent d'un régime lié au droit des bases de données. Dans ce cas, le droit de propriété appartient légalement au « producteur » de la base de données, compris au sens de la personne qui réalise l'investissement 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.

Ressources:
- La diffusion numérique des données en SHS – Guide des bonnes pratiques éthiques et juridiques : https://hal-amu.archives-ouvertes.fr/page/guide-de-bonnes-pratiques
- Qui a les droits, quelles obligations ? https://espacechercheurs.enpc.fr/sites/default/files/logigramme_a_plat_2.pdf



Ressource :

- Fiches pratiques 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



Ressources :

- Ouverture des données de recherche. Guide d'analyse juridique en France : http://dx.doi.org/10.15454/1.481273124091092E12
- La diffusion numérique des données en SHS – Guide des bonnes pratiques éthiques et juridiques : https://hal-amu.archives-ouvertes.fr/page/guide-de-bonnes-pratiques
- Qui a les droits, quelles obligations ? : https://espacechercheurs.enpc.fr/sites/default/files/logigramme_a_plat_2.pdf

ASPECTS JURIDIQUES ET ÉTHIQUES
COMMUNICABILITÉ DES DONNÉES

La communicabilité des données peut être conditionnée par

- la nature ou le type des données
- l'origine des données
- leur(s) utilisation(s)

Elle peut être **empêchée temporairement ou définitivement.**

Toute restriction doit être mentionnée et expliquée dans le DMP

Communication obligatoire pour certaines disciplines	Communication sous conditions	Communication interdite par principe
• Données géographiques • Données environnementales…	• Données protégées par le droit d'auteur ou par contrat • Données personnelles • Statistiques…	• Secrets professionnels • Secrets défense • Sécurité de l'établissement…

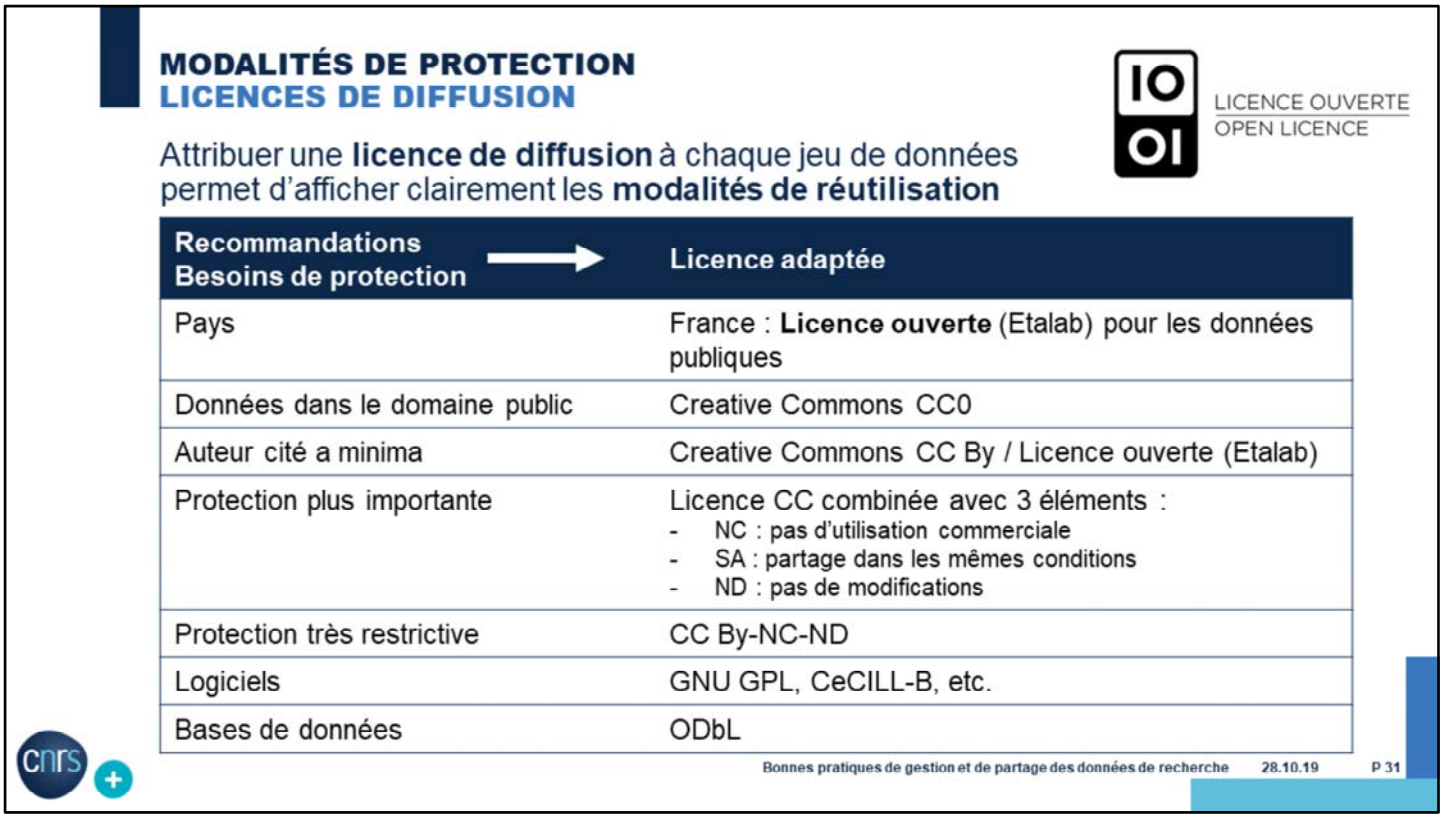
MODALITÉS DE PROTECTION
ACCÈS ET LICENCES

- 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



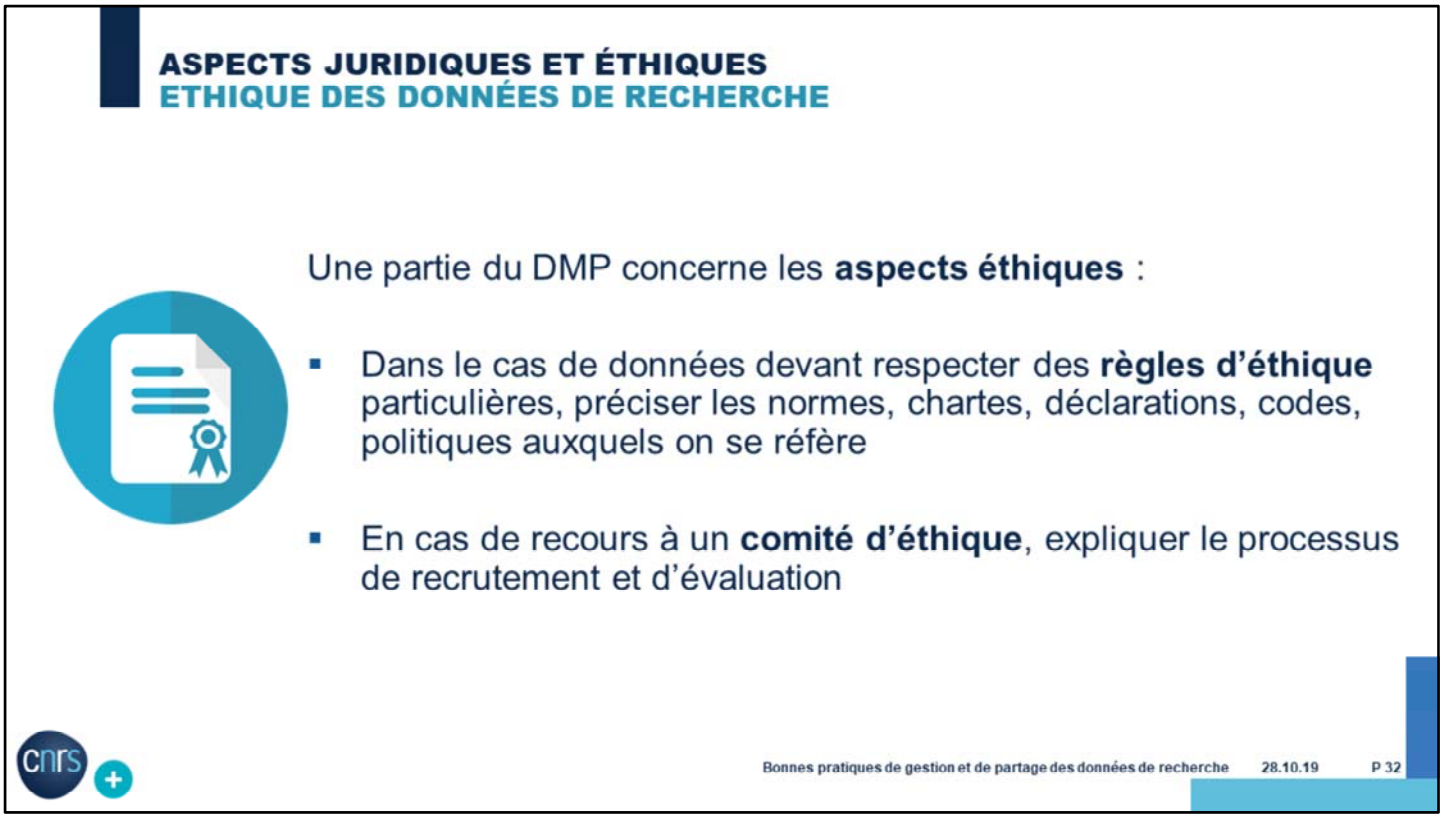
Ressources :
Licences de réutilisation : https://www.data.gouv.fr/fr/licences

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/



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-scientifque-aout-2014.pdf

Ressources :

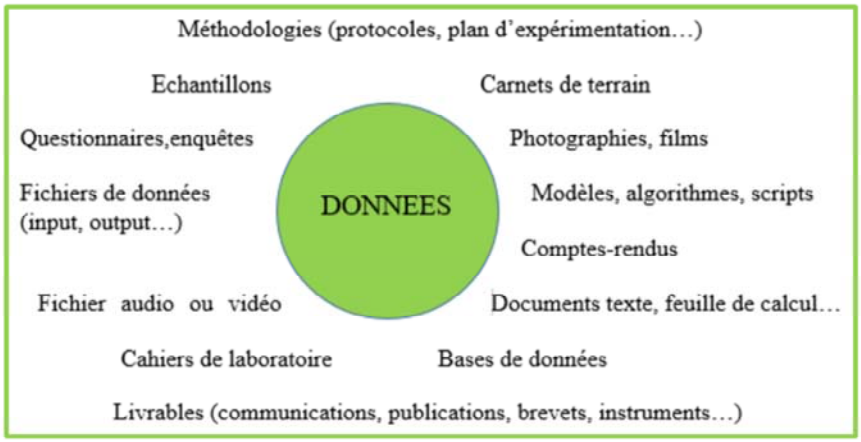
- Politique en matière d'éthique et d'intégrité scientifique (ANR) : http://www.agence-nationale-recherche.fr/fileadmin/documents/2014/Politique-ethique-integrite-scientifque-aout-2014.pdf
- La diffusion numérique des données en SHS – Guide des bonnes pratiques éthiques et juridiques : https://hal-amu.archives-ouvertes.fr/page/guide-de-bonnes-pratiques
- Question éthique et droit en SHS : https://ethiquedroit.hypotheses.org/


CRÉATION, COLLECTE ET DESCRIPTION DES DONNÉES DE RECHERCHE
cnrs
Bonnes pratiques de gestion et de partage des données de recherche 28.10.19 P 33

CYCLE DE VIE DES DONNÉES DE RECHERCHE
CRÉATION, COLLECTE ET DESCRIPTION

D'après Research data lifecycle – UK Data Service
https://www.ukdataservice.ac.uk/manage-data/lifecycle

CRÉATION, COLLECTE ET DESCRIPTION
RAPPEL : DIVERSITÉ DES DONNÉES DE RECHERCHE



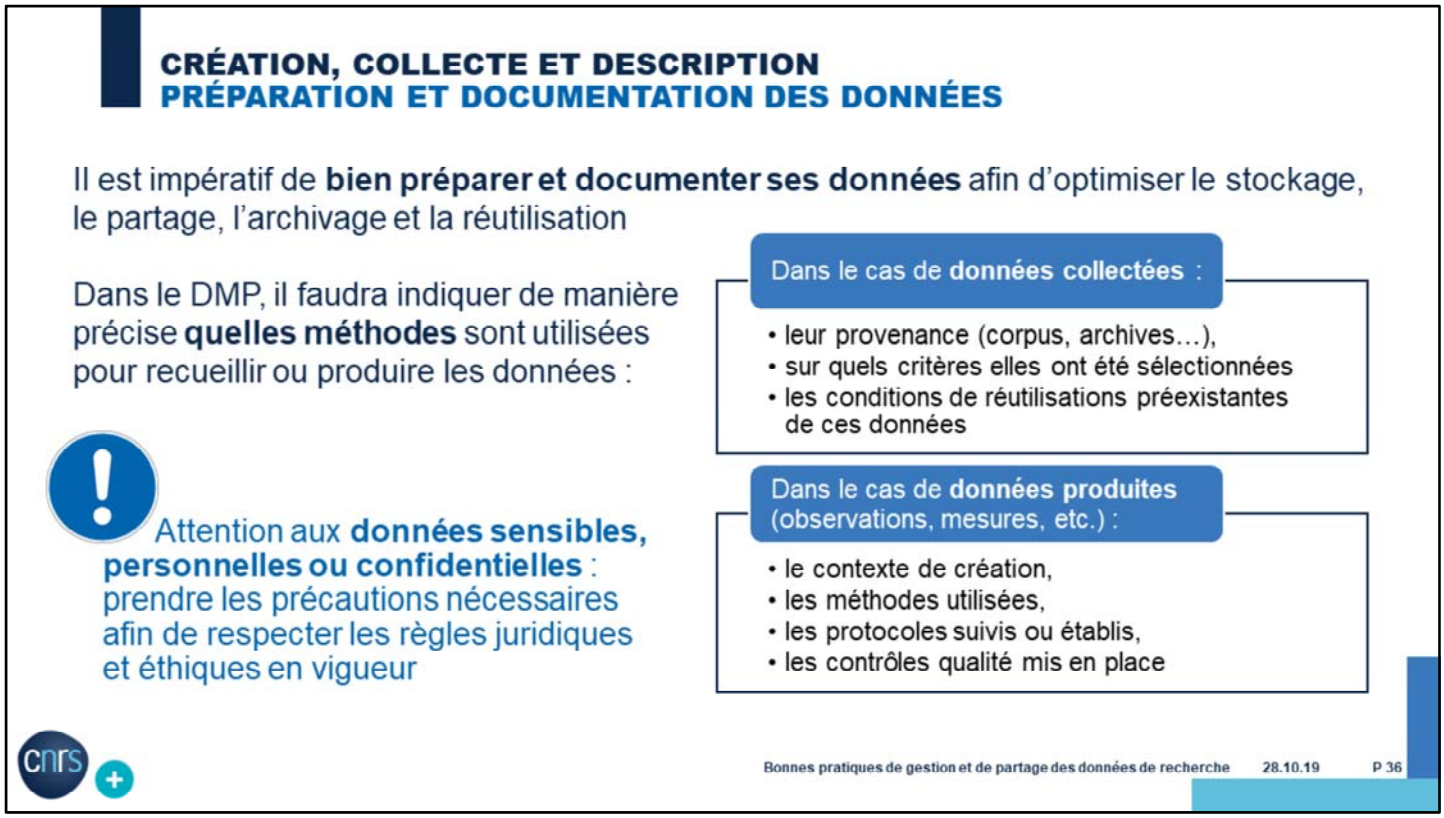
Les données de recherche peuvent être :

- **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/



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.

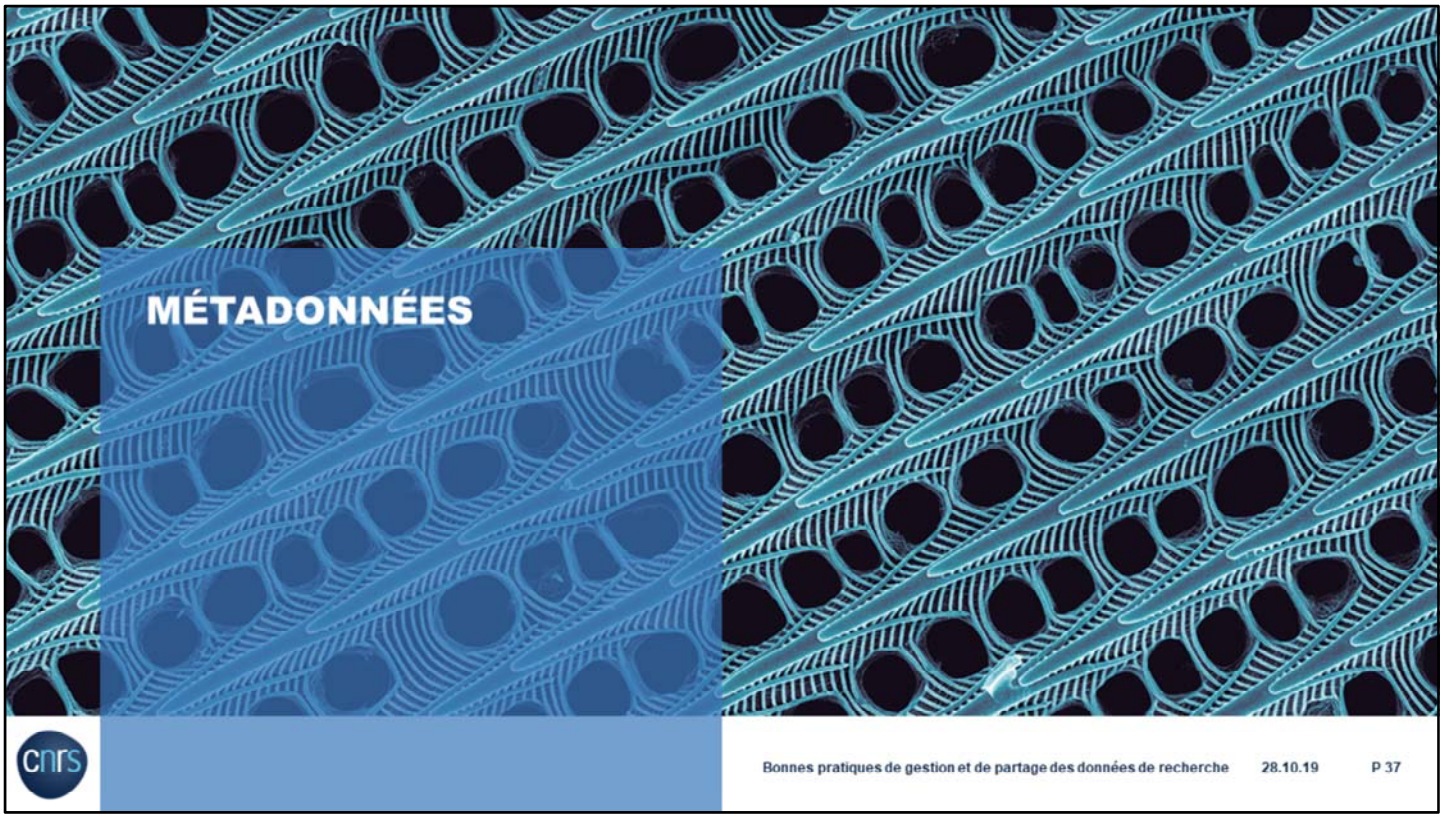
La charte nationale de déontologie des métiers de la recherche souligne ces points essentiels :

- 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 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


MÉTADONNÉES
cnrs
Bonnes pratiques de gestion et de partage des données de recherche 28.10.19 P 37

CYCLE DE VIE DES DONNÉES DE RECHERCHE
LES MÉTADONNÉES

Au moment du partage puis de l'archivage pérenne, des métadonnées spécifiques seront à renseigner

D'après Research data lifecycle – UK Data Service
https://www.ukdataservice.ac.uk/manage-data/lifecycle

Il est recommandé de renseigner les métadonnées au fur et à mesure de l'avancée du projet

MÉTADONNÉES
DÉFINITION - UTILISATION

- Les métadonnées (MTD) permettent de décrire plus précisément les données
- La boite de conserve = jeu de données / l'étiquette = métadonnées

Sans métadonnées

Avec métadonnées

MÉTADONNÉES
MÉTADONNÉES EMBARQUÉES ET ENRICHIES

- Il existe deux types de métadonnées : les MTD embarquées et enrichies

Métadonnée embarquée : produite automatiquement par les appareils (de prise de vue ou de son, de mesure…)

Ex : données GPS, type d'appareil, date, calibrage technique, etc.

Métadonnée enrichie : ajoutée par l'auteur

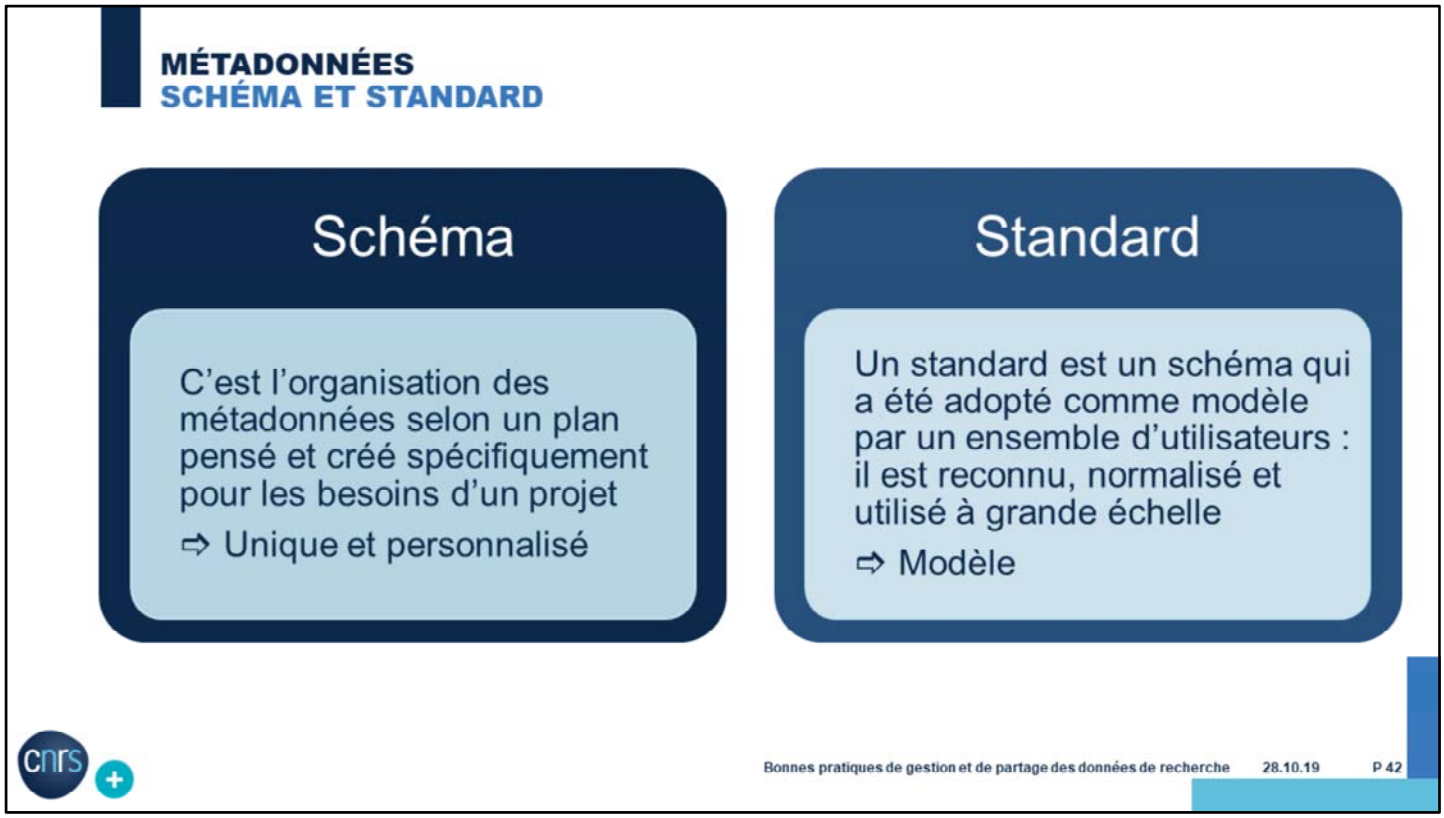
Ex : mots-clés, sujet, auteur, laboratoire ou organisme, nom du projet, licence, etc.

MÉTADONNÉES
RENSEIGNEMENT DES MÉTADONNÉES

- Les métadonnées sont à renseigner dans l'idéal au fur et à mesure

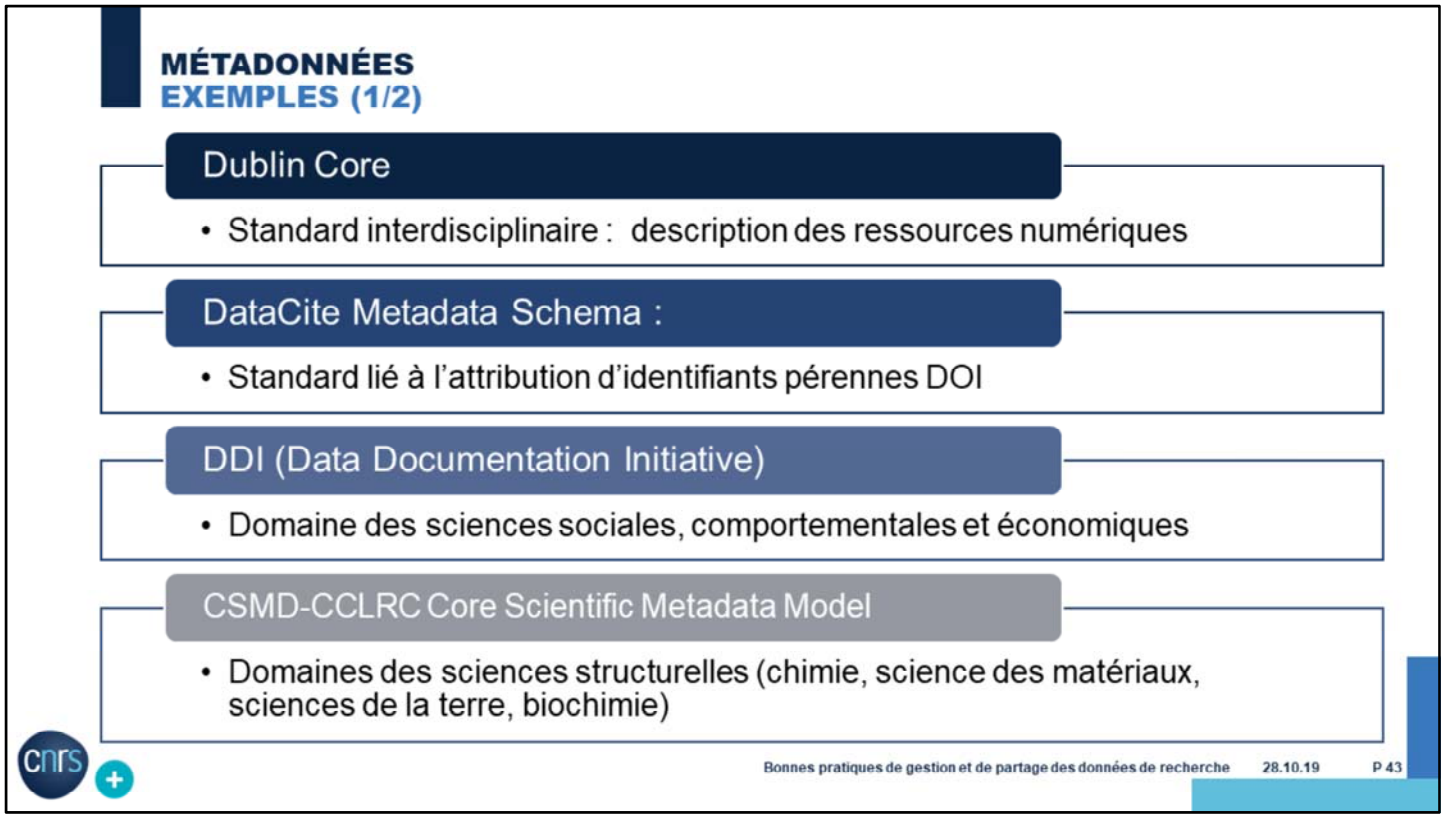
- Compléter les MTD embarquées par des MTD enrichies

- Utiliser un **standard** de MTD dans votre discipline ou adapté à vos besoins, et s'il n'en existe pas, créer un **schéma** de métadonnées



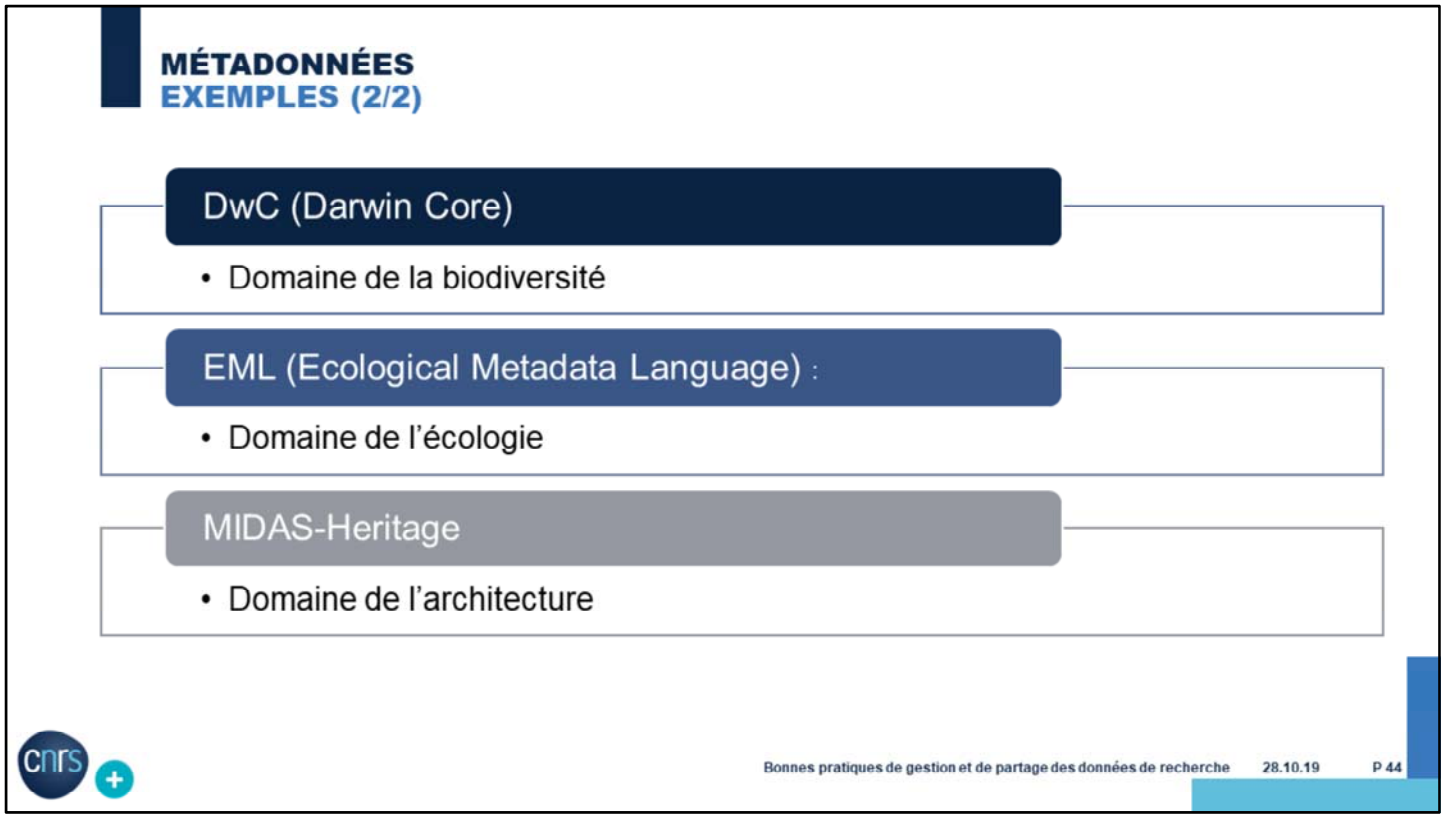
Ressources :

- DoRANum – Les schémas de métadonnées : https://doranum.fr/metadonnees-standards-formats/schemas-metadonnees/
- DoRANum – Les standards de métadonnées : pourquoi, lequel ? : https://doranum.fr/metadonnees-standards-formats/standard-metadonnees/
- Répertoire de standards de métadonnées en Sciences de la Vie FAIRsharing.org : https://fairsharing.org/
- RDA Metadata Standards Directory : http://rd-alliance.github.io/metadata-directory/standards/
- Disciplinary Metadata : http://www.dcc.ac.uk/drupal/resources/metadata-standards
- Consulter également les 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 structurelles (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/



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 structurelles (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/

MÉTADONNÉES
EXEMPLE : LE DUBLIN CORE

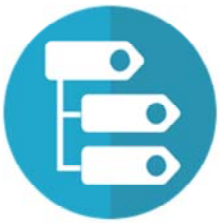
- Le Dublin Core est un standard international et multidisciplinaire
- Il comporte 15 éléments (= minimum exigé)

Élément	Élément (anglais)	Commentaire
1. Titre (métadonnée)	Title	Nom donné à la ressource
2. Créateur (métadonnée)	Creator	Nom de la personne, de l'organisation ou du service responsable de la création du contenu de la ressource
3. Sujet (métadonnée) ou mots clés	Subject	Thème du contenu de la ressource (mots clés, expressions, codes de classification)
4. Description (métadonnée)	Description	Présentation du contenu de la ressource (résumé, table des matières, représentation graphique du contenu, texte libre)
5. Éditeur	Publisher	Nom de la personne, de l'organisation ou du service responsable de la mise à disposition ou de la diffusion de la ressource

Source : wikipédia https://fr.wikipedia.org/wiki/Dublin_Core

MÉTADONNÉES
ENRICHISSEMENT DES MÉTADONNÉES

- Utiliser des vocabulaires contrôlés disciplinaires utilisés couramment : (ontologies, lexiques, thesaurus…)
- Exemples :
  - Codex de médicaments
  - Classifications taxonomiques
  - Nomenclature internationale des formules chimiques

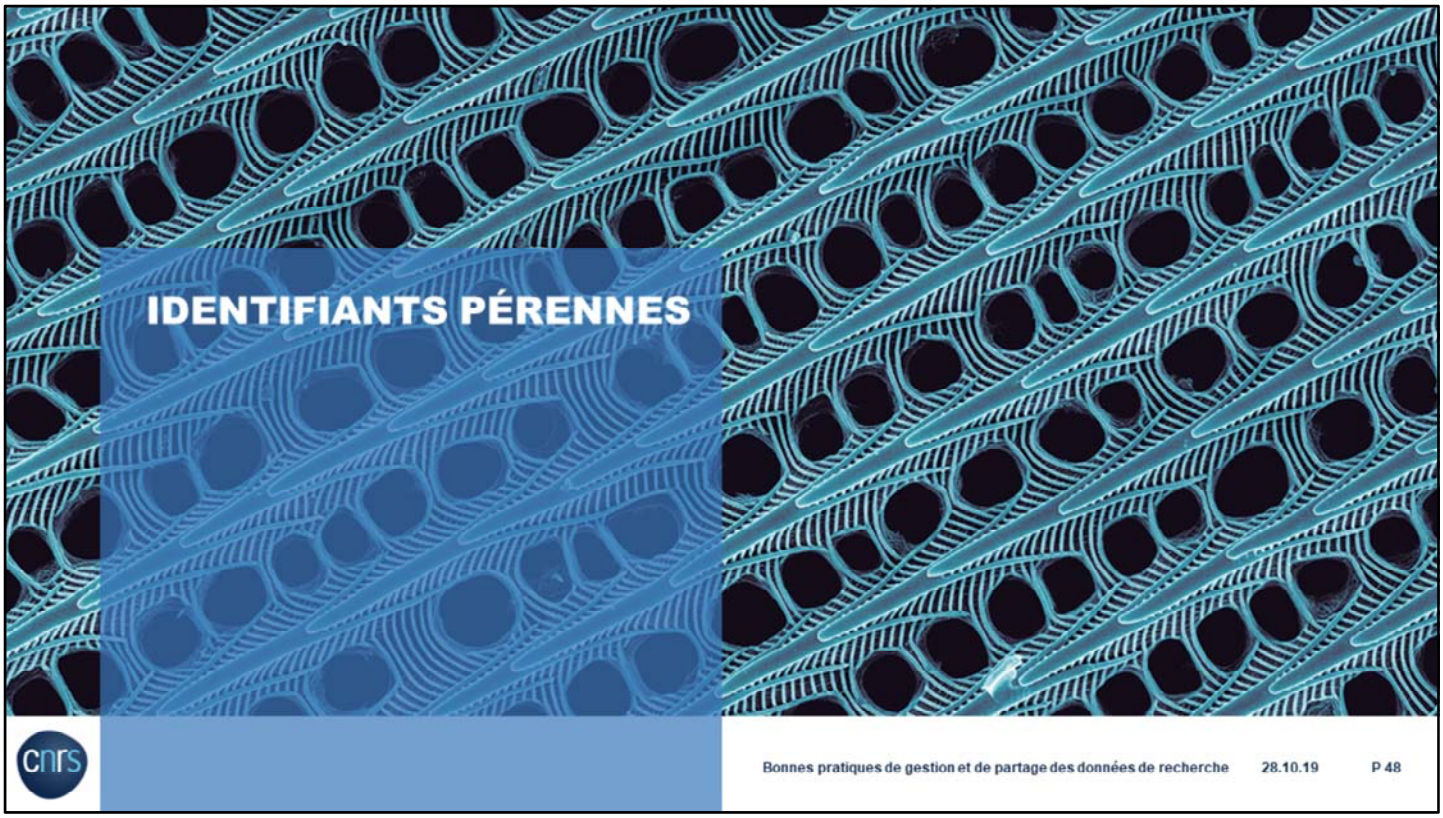


- Cela augmentera la capacité des données à être combinées avec d'autres données

MÉTADONNÉES
UTILITÉ

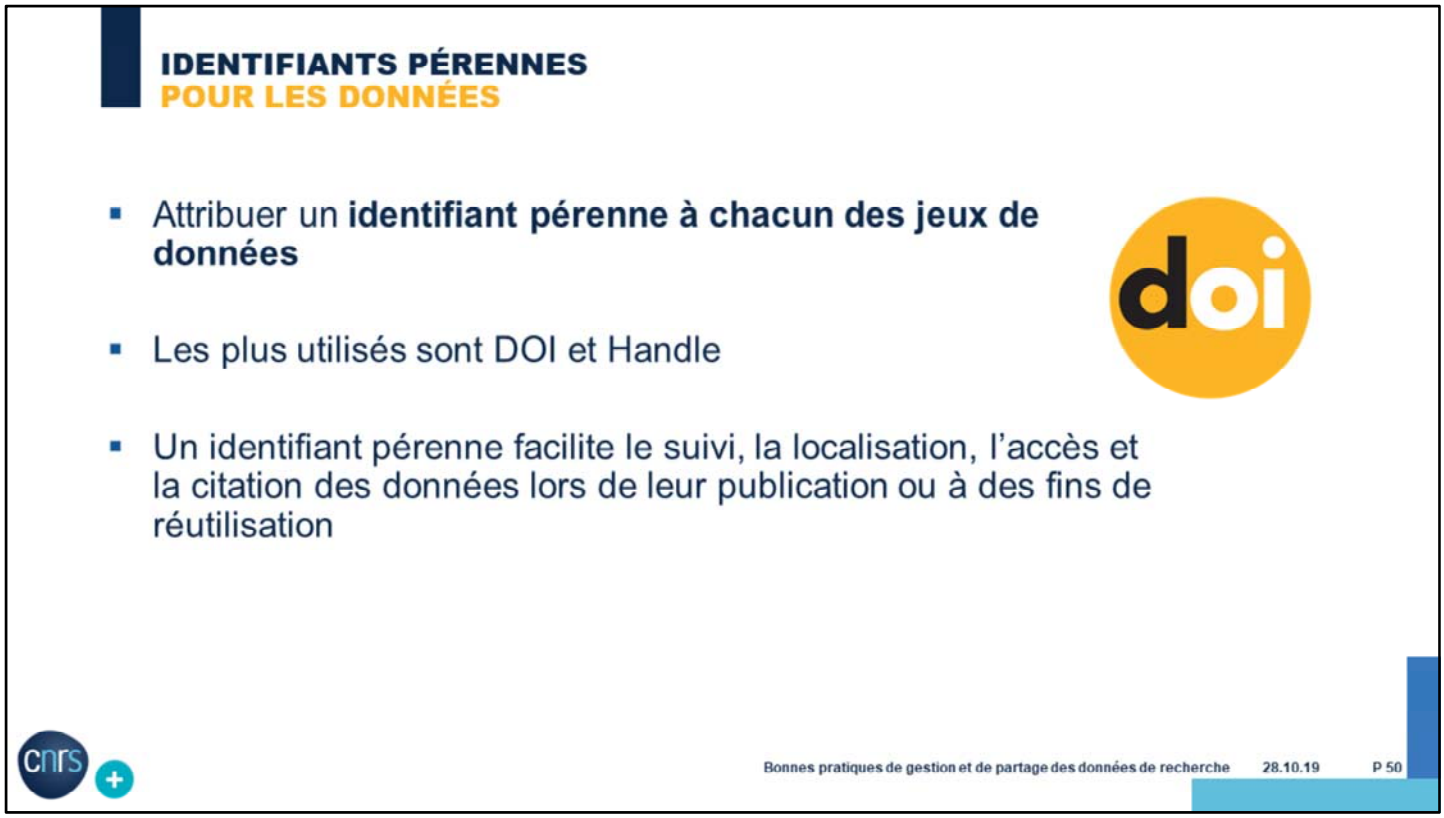
Les métadonnées sont utiles pour :

- 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
- Connaitre 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


IDENTIFIANTS PÉRENNES
cnrs
Bonnes pratiques de gestion et de partage des données de recherche 28.10.19 P 48

CYCLE DE VIE DES DONNÉES DE RECHERCHE
IDENTIFIANTS PÉRENNES

D'après Research data lifecycle – UK Data Service
https://www.ukdataservice.ac.uk/manage-data/lifecycle



Une équipe dédiée de l'Inist-CNRS se tient à disposition pour conseiller dans l'attribution de DOI aux données de recherche 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.

Ressources :

- DoRANum : Zoom sur ORCID et DOI (https://doranum.fr/identifiants-perennes-pid/zoom-orcid-doi/)
- PID OPIDoR (https://opidor.fr/identifier/) : service de l 'Inist-CNRS, agence nationale DataCite pour l'attribution de DOI (https://www.datacite.org/)

IDENTIFIANTS PÉRENNES
IDENTIFIANT AUTEUR

Attribuer un **identifiant auteur** (ORCID)

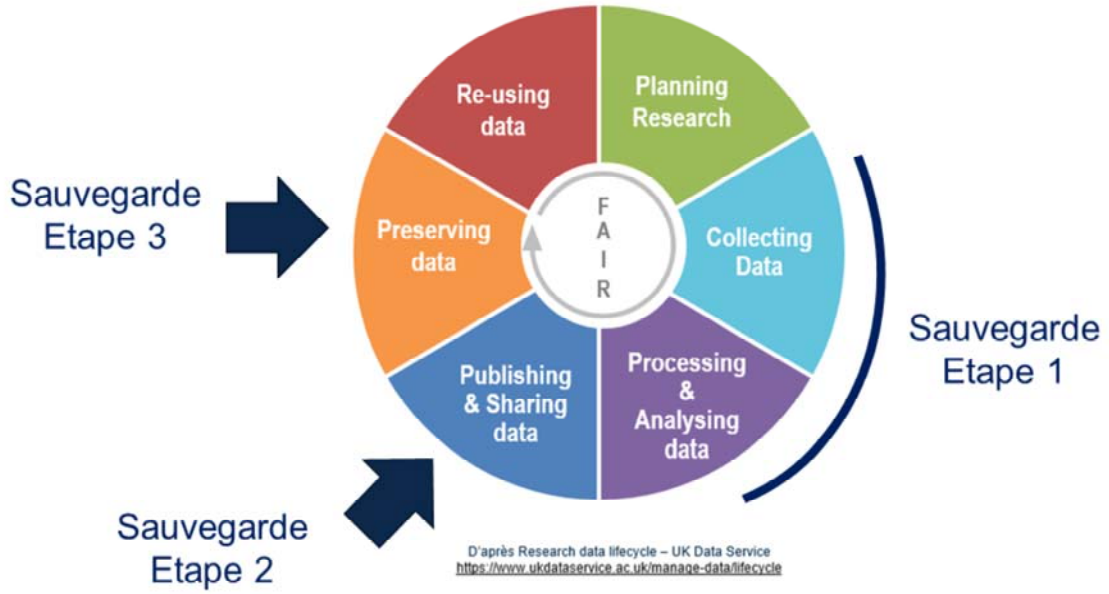
- Fait le lien avec ses productions scientifiques
- Permet d'être bien identifié et cité

Ressources :
DoRANum : Zoom sur ORCID et DOI (https://doranum.fr/identifiants-perennes-pid/zoom-orcid-doi/)


SAUVEGARDE DES DONNÉES TOUT AU LONG DU PROJET
cnrs
Bonnes pratiques de gestion et de partage des données de recherche 28.10.19 P 52

3 ÉTAPES DE SAUVEGARDE DES DONNÉES

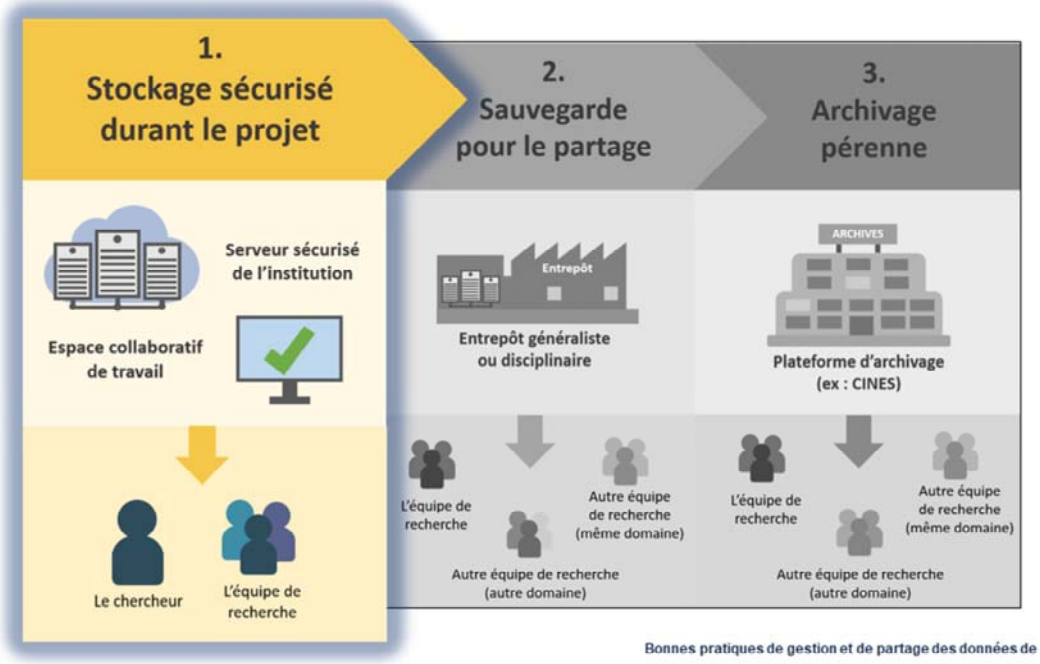
CYCLE DE VIE DES DONNÉES DE RECHERCHE
SAUVEGARDE DES DONNÉES

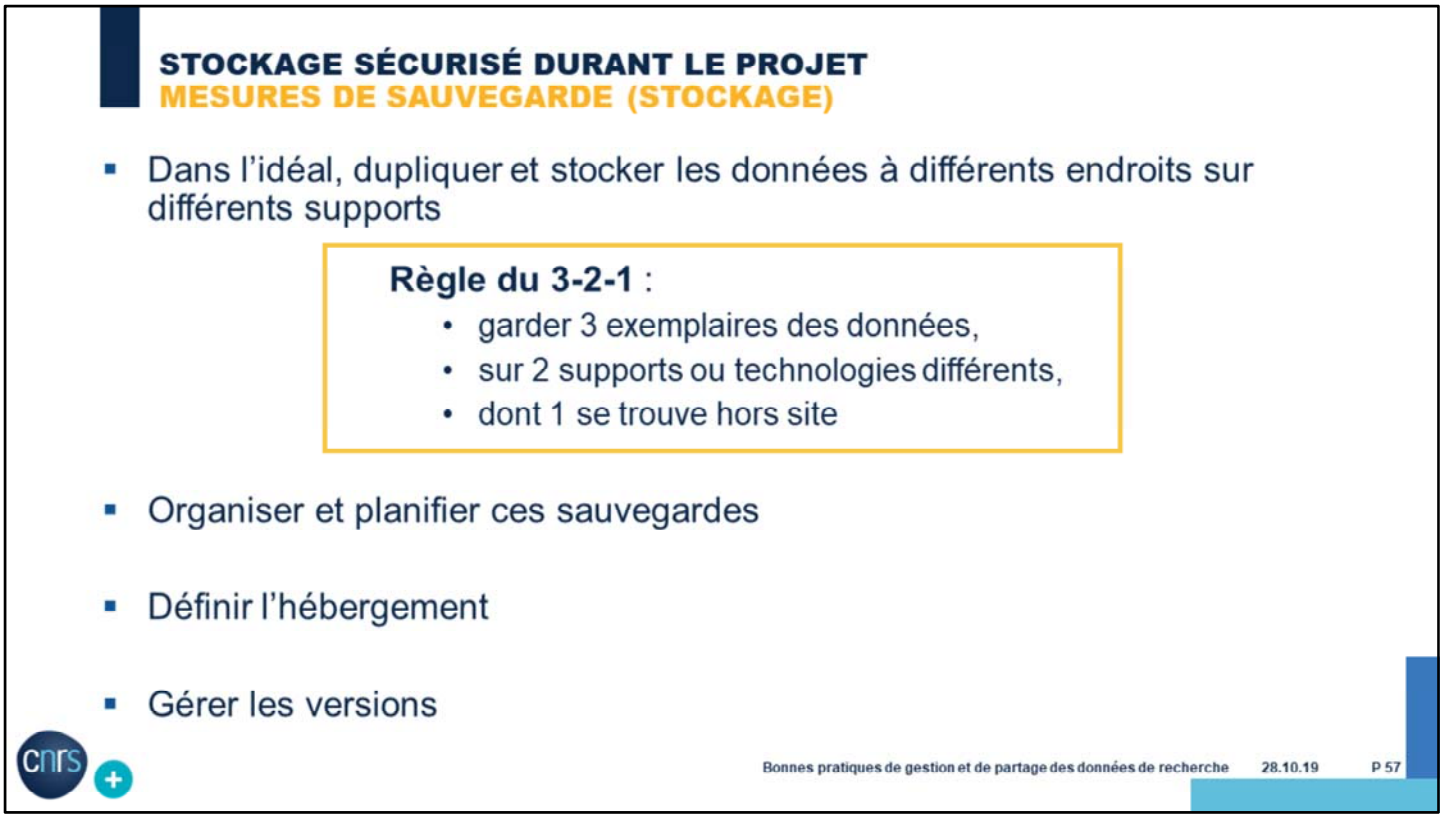


D'après Research data lifecycle – UK Data Service
https://www.ukdataservice.ac.uk/manage-data/lifecycle


ETAPE 1
STOCKAGE SÉCURISÉ
DURANT LE PROJET

3 ÉTAPES DE SAUVEGARDE DES DONNÉES

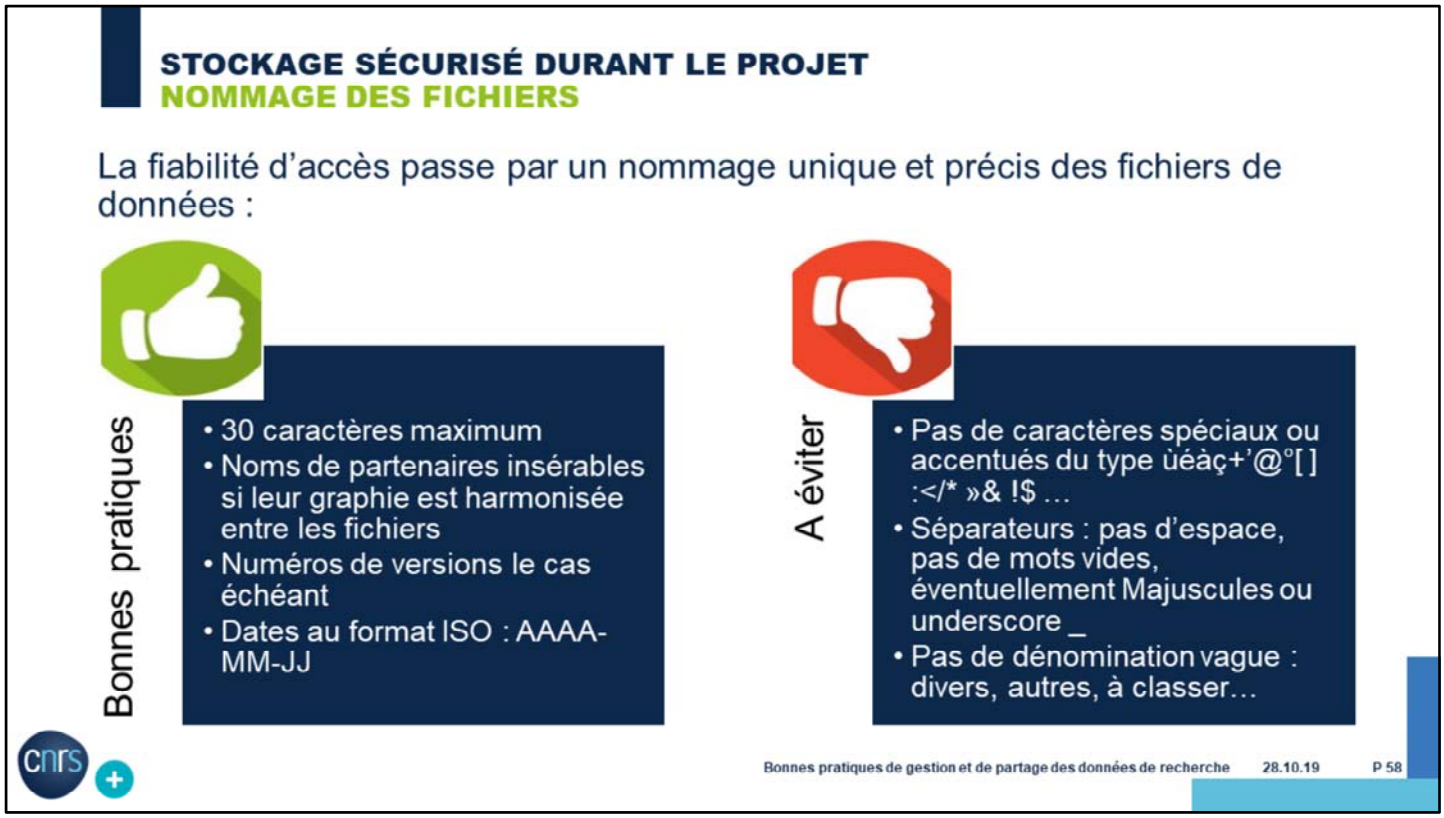




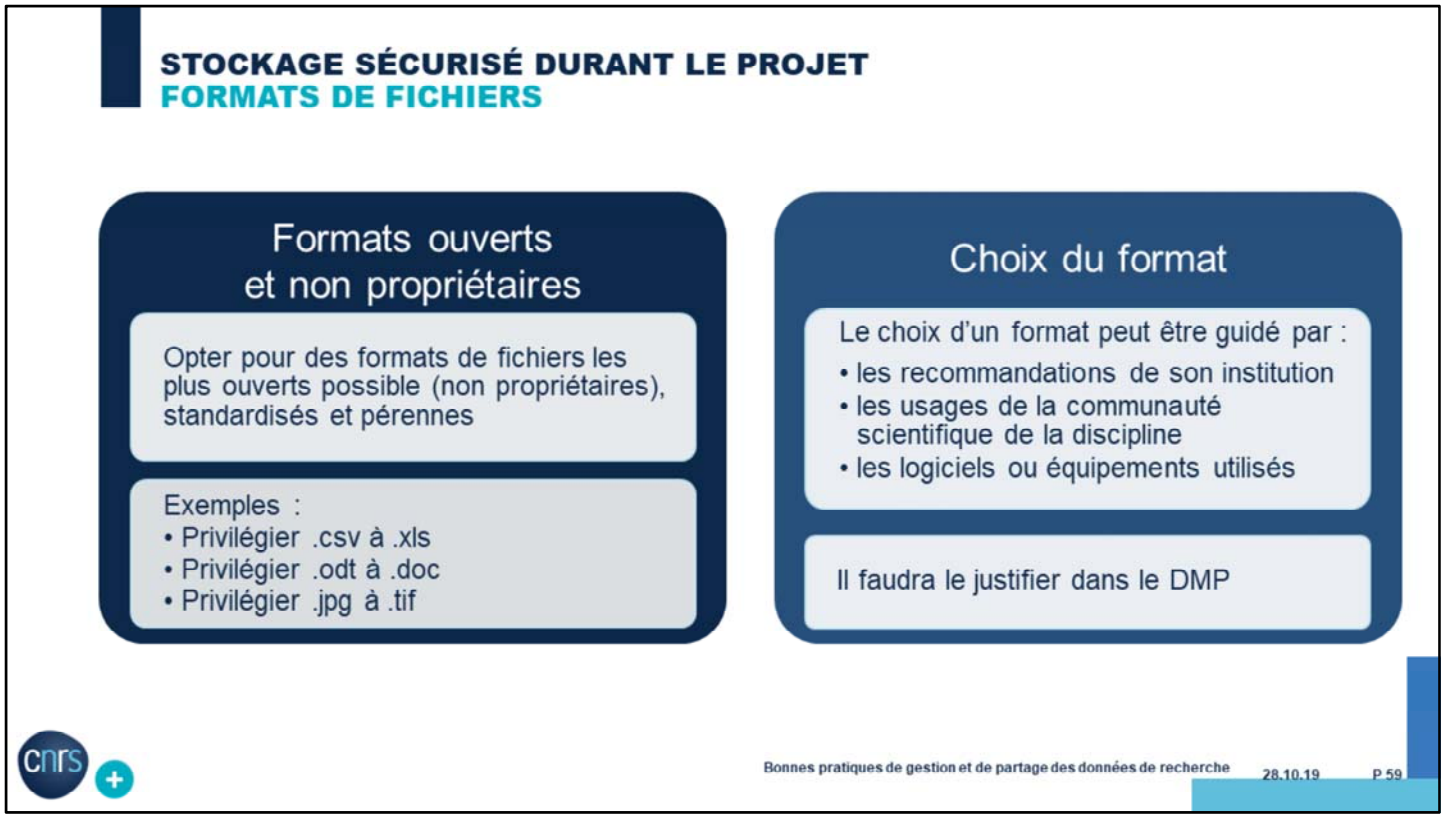
Organiser et planifier les sauvegardes :

- 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



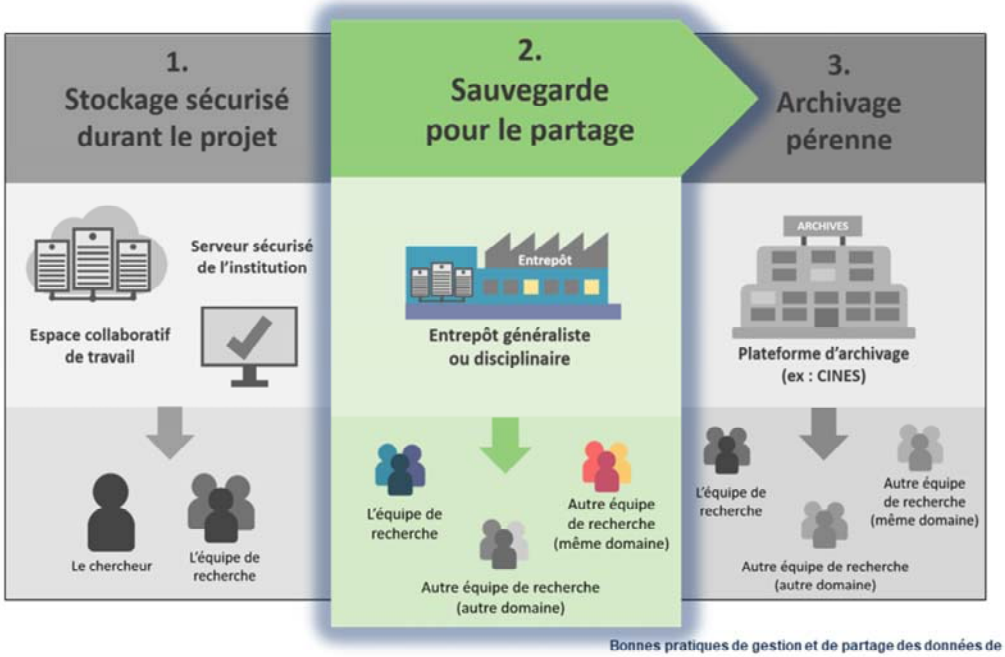
Source :
Arnould Pierre-Yves, Jacquemot-Perbal Marie-Christine. Guide de bonnes pratiques : Gestion et valorisation des données de recherche. 1er février 2016. https://ordar.otelo.univ-lorraine.fr/record?id=10.24396/ORDAR-1



Ressources :
DoRANum : Tableau format ouvert ou fermé (https://doranum.fr/stockage-archivage/quiz-format-ouvert-ou-ferme/)

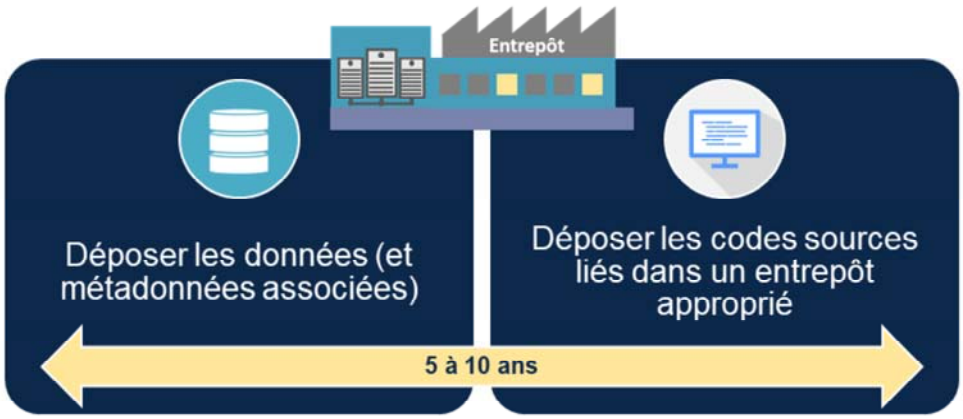

ETAPE 2
PARTAGE, DIFFUSION DES DONNÉES
DÉPÔT DANS UN ENTREPÔT
cnrs
Bonnes pratiques de gestion et de partage des données de recherche 28.10.19 P 60

3 ÉTAPES DE SAUVEGARDE DES DONNÉES



SAUVEGARDE POUR LE PARTAGE
DÉPÔT DANS UN ENTREPÔT

- Permet le partage et la réutilisation optimale des données sur le court et le moyen terme (5 à 10 ans)

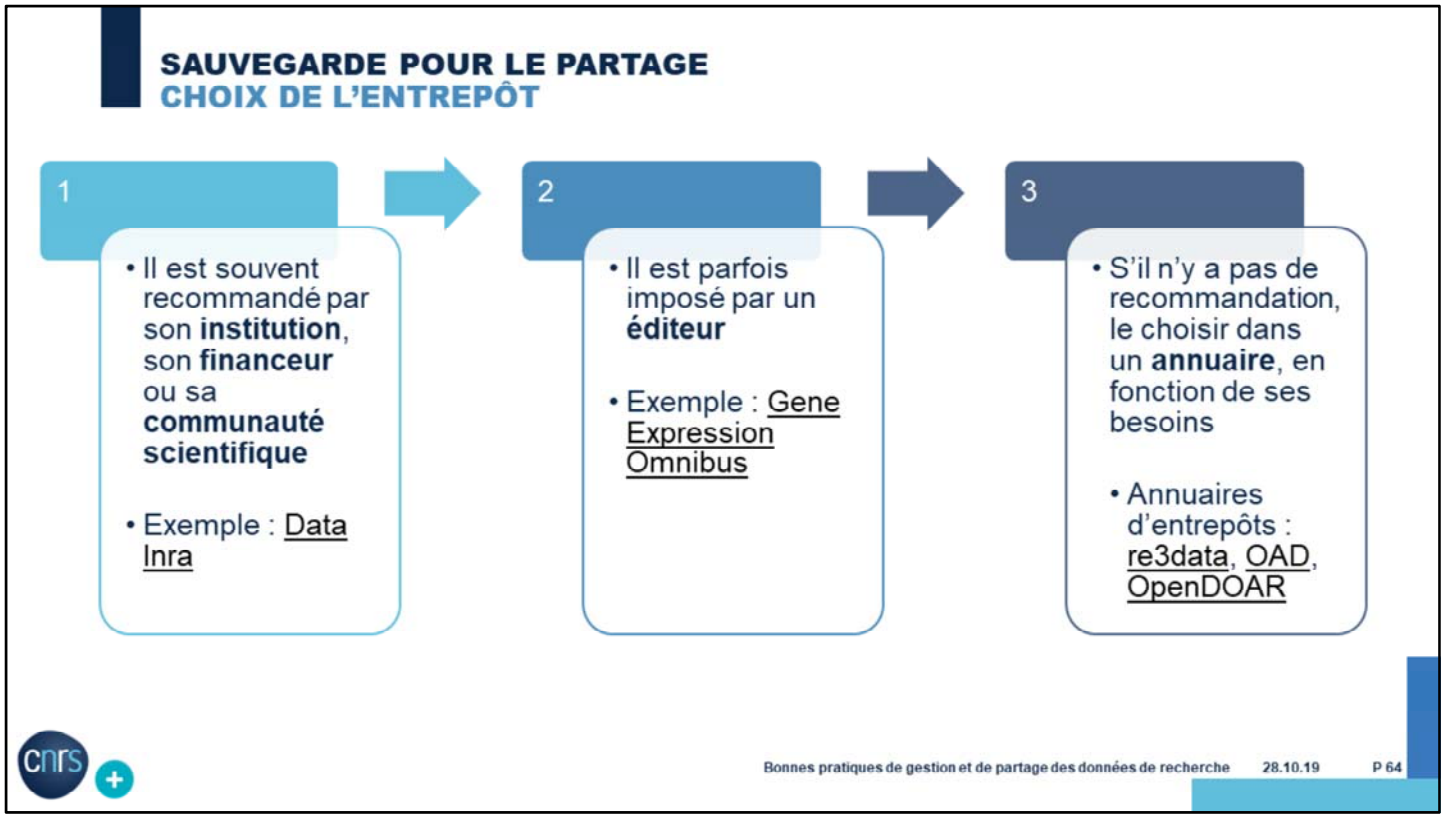


Ressources :
- Pour les logiciels, déposer les codes sources dans HAL (https://hal.archives-ouvertes.fr/), lien avec Software Heritage, archive de logiciels (https://www.softwareheritage.org/?lang=fr)

SAUVEGARDE POUR LE PARTAGE
PRÉPARATION DES DONNÉES

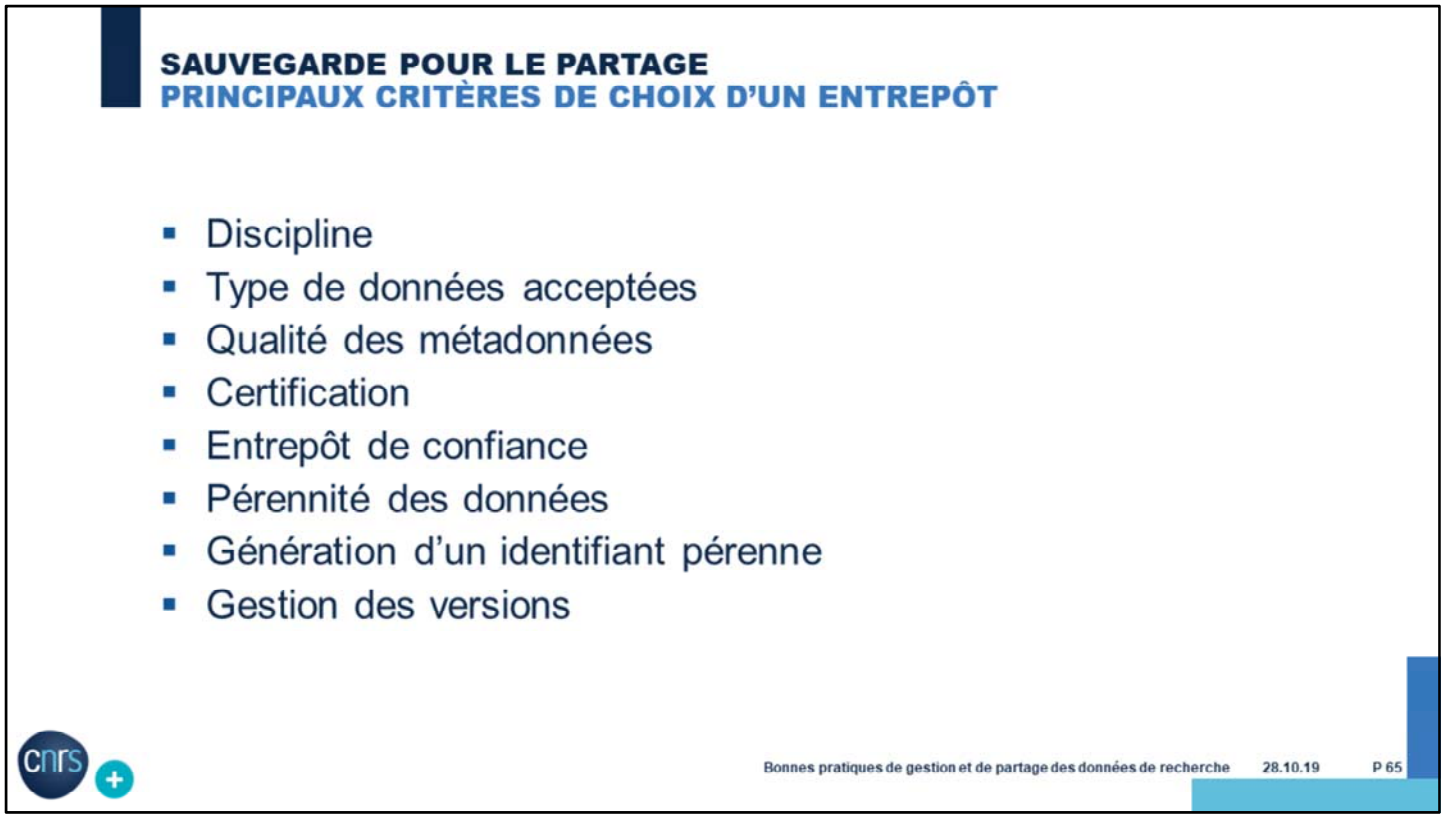
Check-list

- 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



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://gt-donnees2018.sciencesconf.org/data/08_donnees_biologie_Poulain_2018.pdf



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

https://datadryad.org/

https://www.nakala.fr/

https://zenodo.org/
https://doranum.fr/depot-entrepots/depot-donnees-recherche-zenodo/
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/

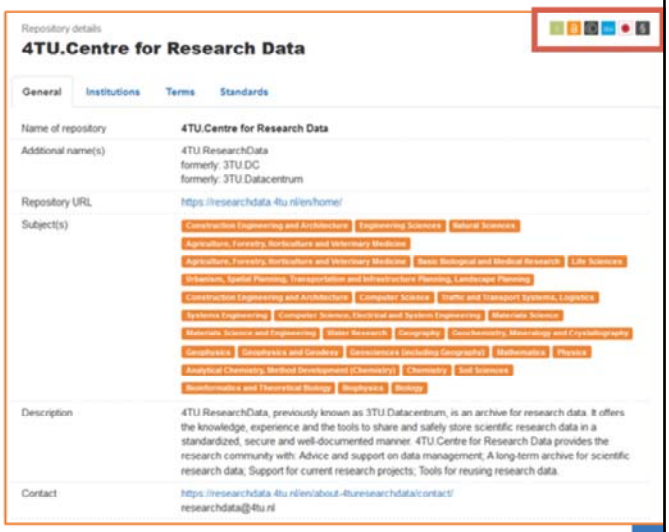
SAUVEGARDE POUR LE PARTAGE
RECHERCHE D'ENTREPÔTS

Dans **re3data**, exemple de recherche à partir des critères suivants :

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

Plusieurs entrepôts répondent à ces critères :

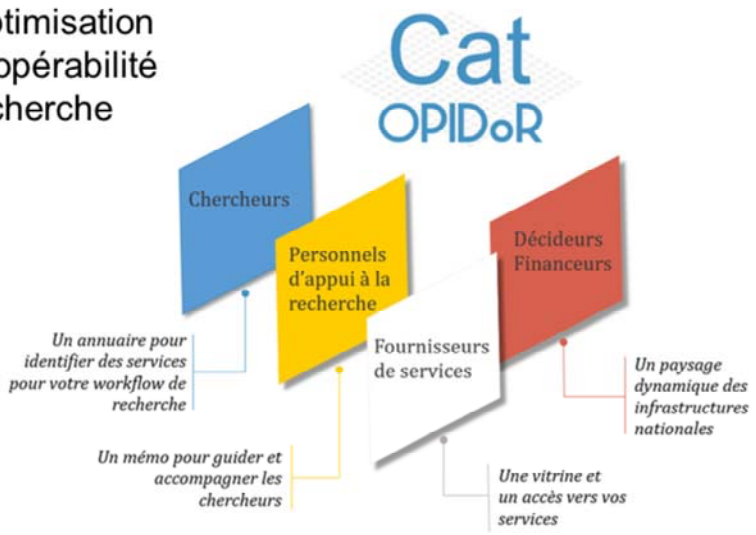
- **4TU.Centre for Research Data**
- CLARIN repository at the University of Tübingen
- NASA Socioeconomic Data and Applications Center
- PANGAEA



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.

SAUVEGARDE POUR LE PARTAGE
OUTIL CAT OPIDOR

Catalogue pour une Optimisation du Partage et de l'Interopérabilité des Données de la Recherche



https://cat.opidor.fr/

https://cat.opidor.fr/

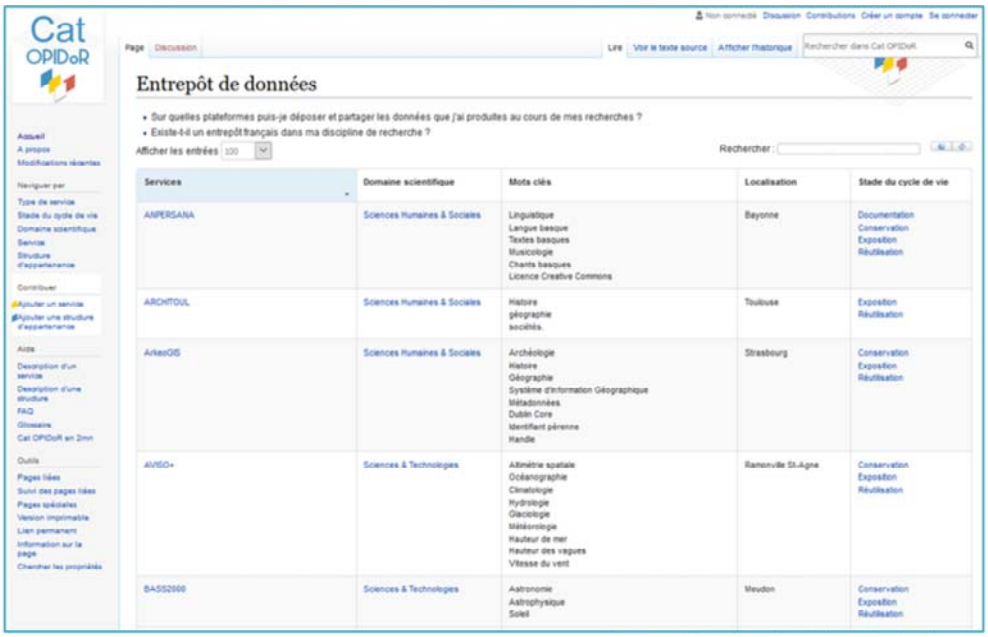
SAUVEGARDE POUR LE PARTAGE
OUTIL CAT OPIDOR

- Recense et décrit les **services français** dédiés aux données scientifiques
- Proposé sous forme d'un wiki, cet **outil collaboratif** ouvert à tous permet de repérer et ajouter des services utiles dans le cadre d'un projet de recherche
- Cat OPIDoR présente par **domaine scientifique** :
  - des sites d'information,
  - de formation,
  - des outils de gestion,
  - des plateformes,

  pour accompagner les chercheurs sur l'ensemble des étapes clés de la gestion, collecte, stockage, conservation et ouverture des données

SAUVEGARDE POUR LE PARTAGE
OUTIL CAT OPIDOR

SAUVEGARDE POUR LE PARTAGE
OUTIL CAT OPIDOR

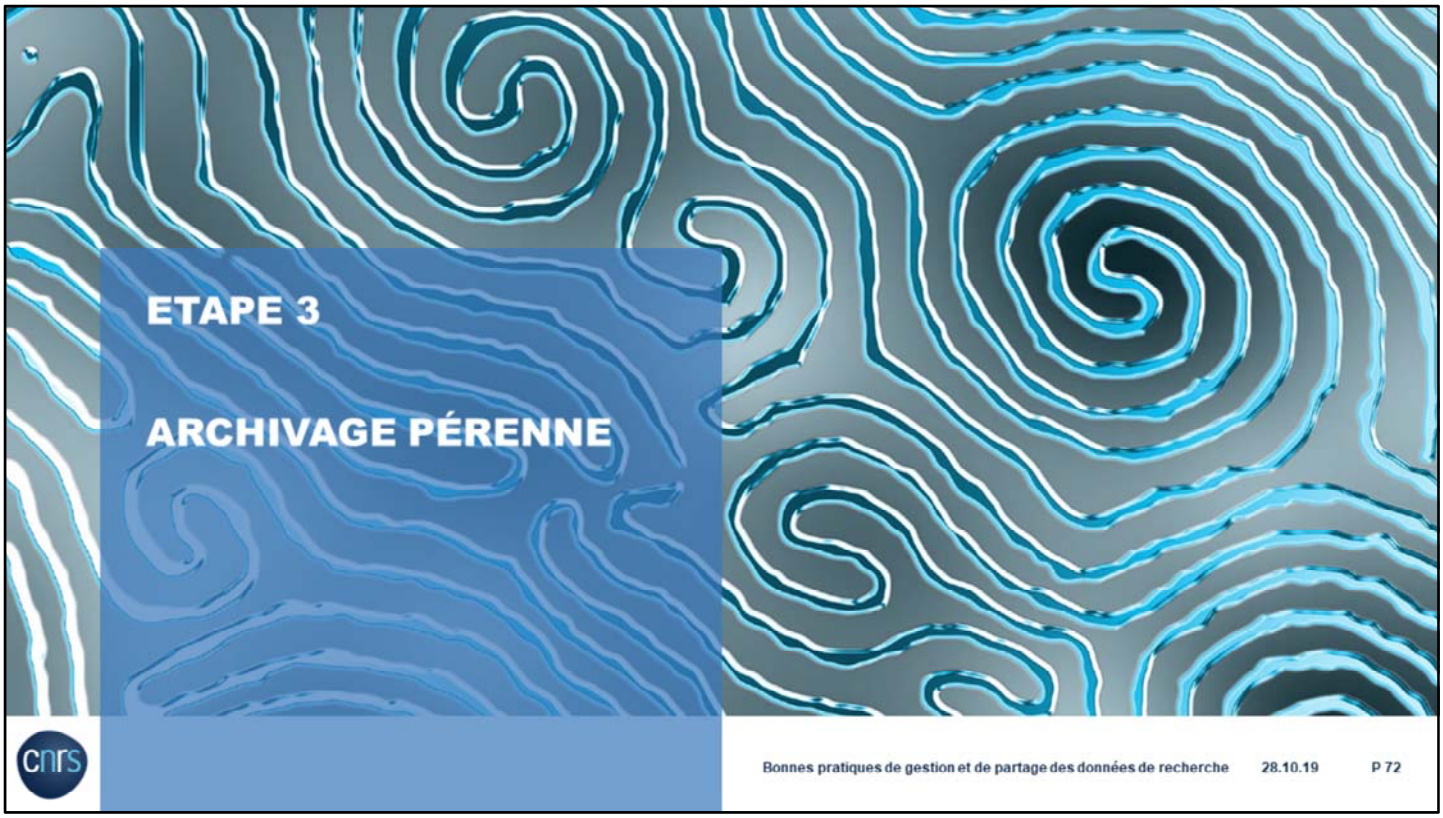


Cat OPIDoR

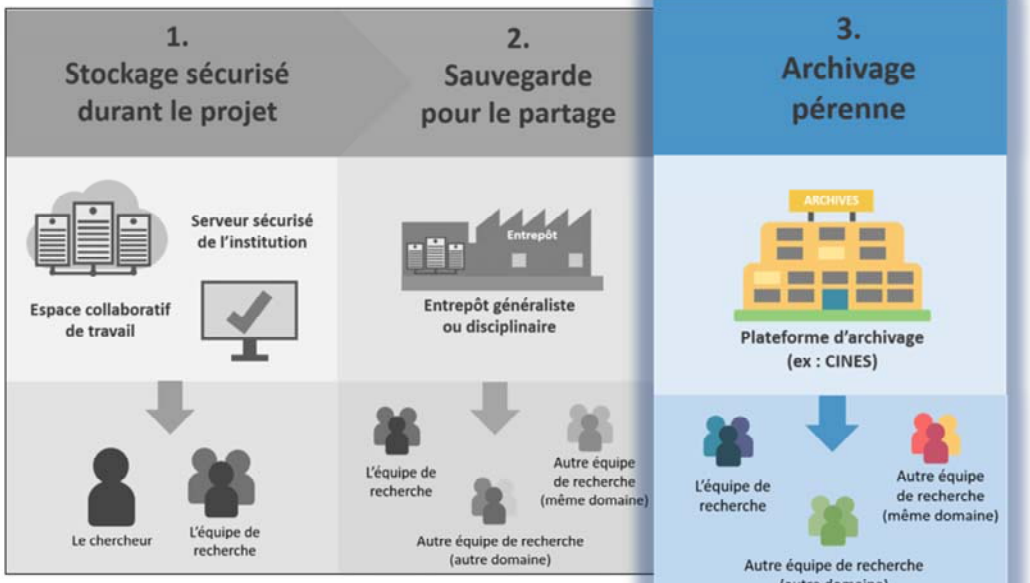
Entrepôt de données

- Sur quelles plateformes puis-je déposer et partager les données que j'ai produites au cours de mes recherches ?
- Existe-t-il un entrepôt français dans ma discipline de recherche ?

Services	Domaine scientifique	Mots clés	Localisation	Stade du cycle de vie
ANPERSANA	Sciences Humaines & Sociales	Linguistique Langue basque Textes basques Musicologie Chants basques Licence Creative Commons	Bayonne	Documentation Conservation Exposition Réutilisation
ARCHITOUL	Sciences Humaines & Sociales	Histoire géographie sociétés	Toulouse	Exposition Réutilisation
ArkeoGIS	Sciences Humaines & Sociales	Archéologie Histoire Géographie Système d'Information Géographique Métadonnées Dublin Core Identifiant pérenne Handle	Strasbourg	Conservation Exposition Réutilisation
AVISO+	Sciences & Technologies	Altimétrie spatiale Océanographie Climatologie Hydrologie Glaciologie Météorologie Hauteur de mer Hauteur des vagues Vitesse du vent	Ramonville St-Agne	Conservation Exposition Réutilisation
BASS2000	Sciences & Technologies	Astronomie Astrophysique Soleil	Meudon	Conservation Exposition Réutilisation


ETAPE 3
ARCHIVAGE PÉRENNE
cnrs
Bonnes pratiques de gestion et de partage des données de recherche 28.10.19 P 72

3 ÉTAPES DE SAUVEGARDE DES DONNÉES



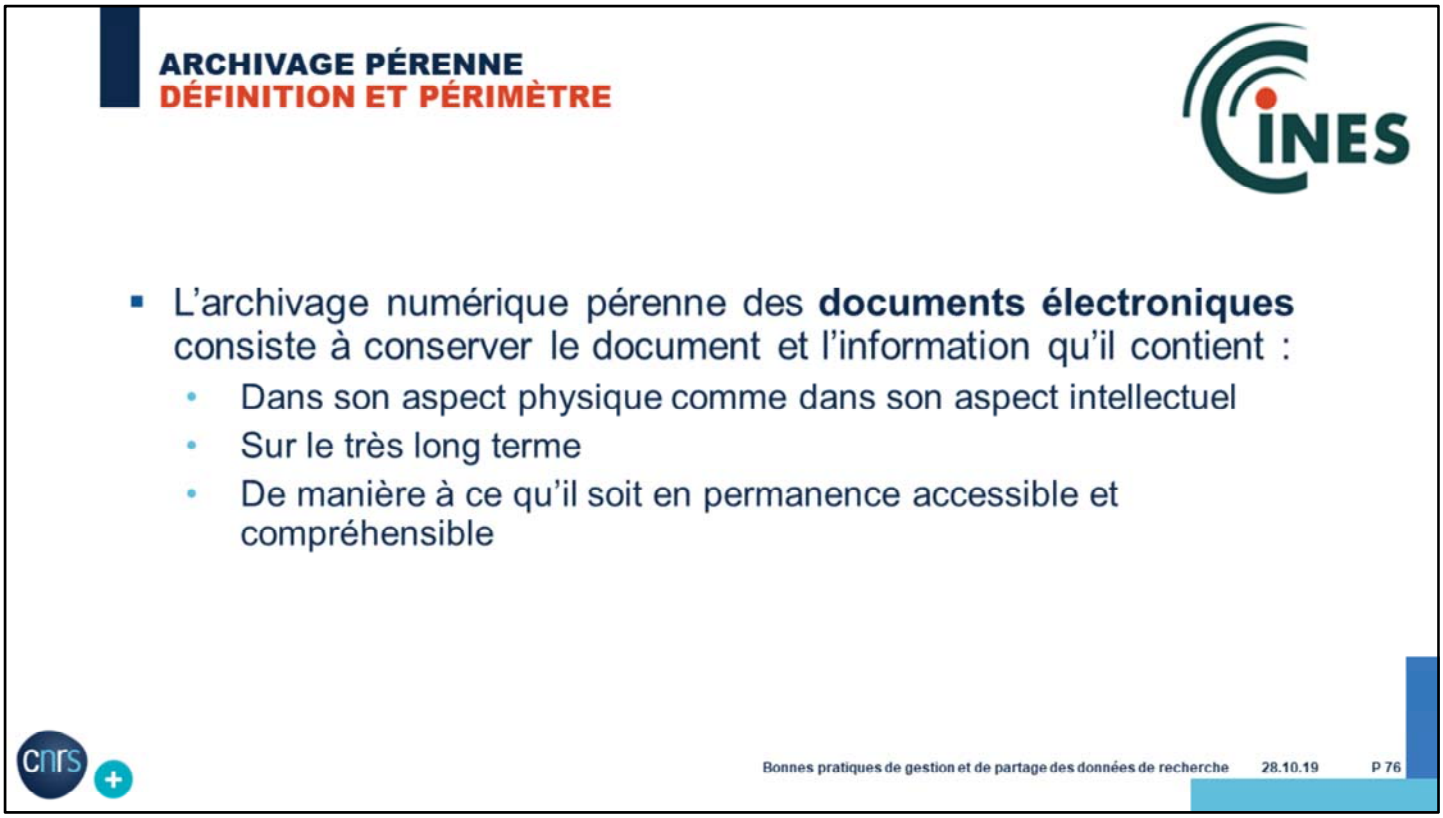
ARCHIVAGE PÉRENNE
PLATEFORME D'ARCHIVAGE DU CINES

- 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 **PAC**, la Plateforme d'Archivage au CINES

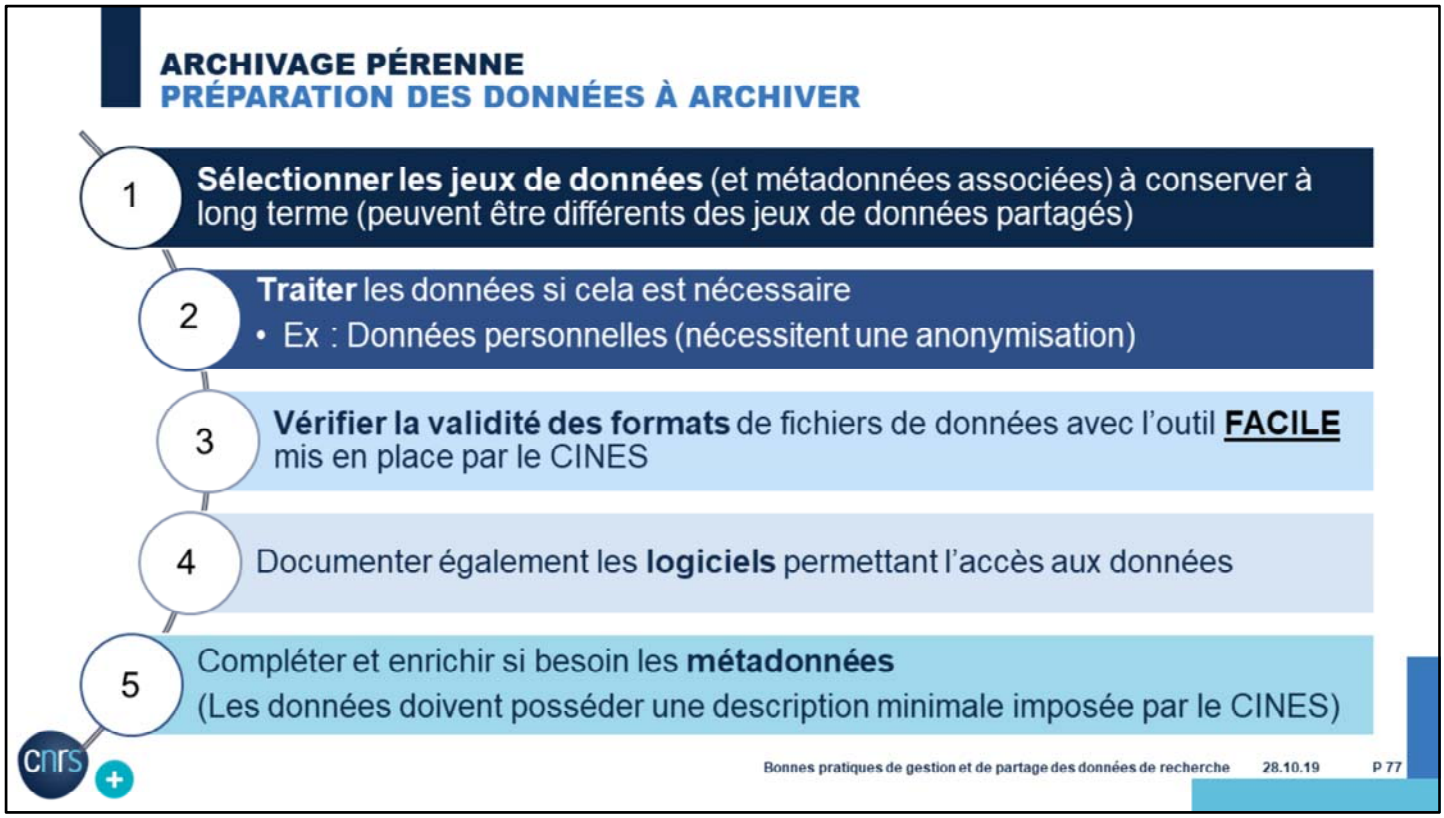
- Selon son institution, sa discipline ou l'entrepôt choisi, il existe déjà des partenariats avec le CINES, proposant un accompagnement pour l'archivage.

  Ex : Huma-Num en SHS

Source :
CINES - L'archivage numérique, qu'est-ce que c'est ? -
https://www.cines.fr/archivage/un-concept-des-problematiques/



Source :
CINES - L'archivage numérique, qu'est-ce que c'est ? -
https://www.cines.fr/archivage/un-concept-des-problematiques/



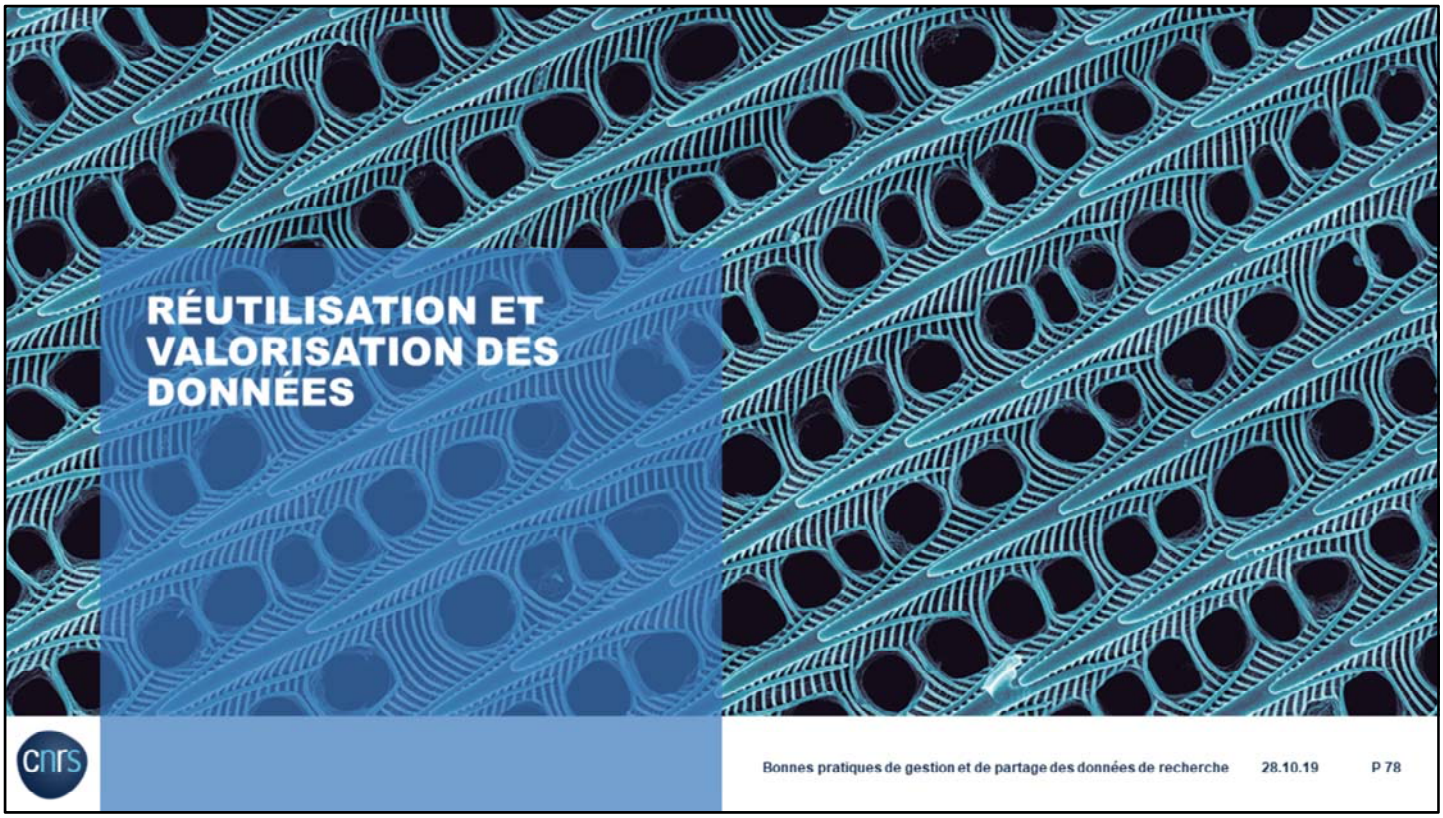
Tester ses fichiers avec l'Outil FACILE du CINES : opération de contrôle de validité et/ou d'éligibilité générales.
Exemple : 2 étapes pour les fichiers PDF :

- 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/

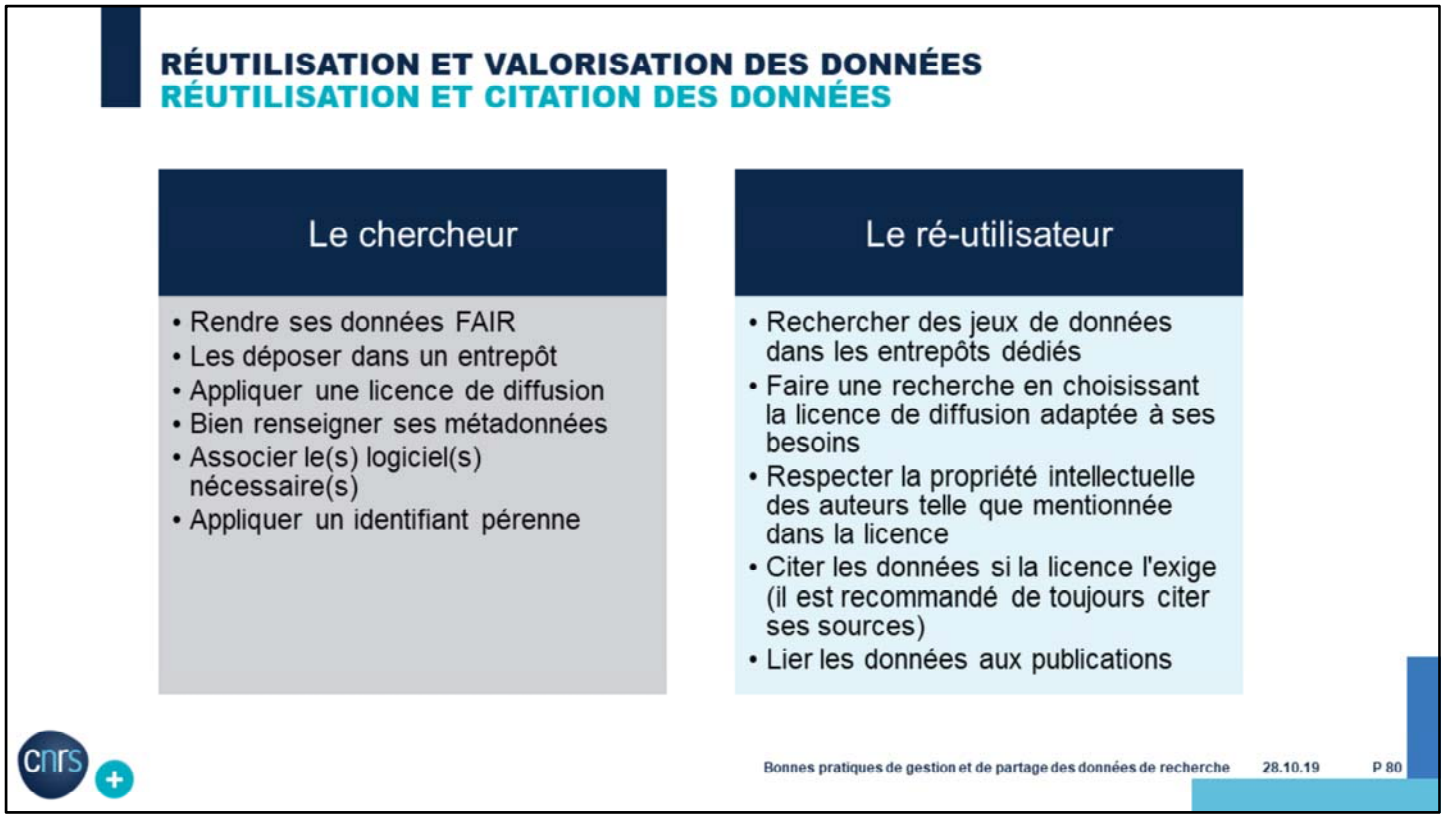
Ressources :

- Référentiel de gestion des archives de la recherche : https://www.archivistes.org/Referentiel-de-gestion-des-archives-de-la-recherche
- Le concept d'archivage numérique pérenne : https://www.cines.fr/archivage/un-concept-des-problematiques/le-concept-darchivage-numerique-perenne/
- Les formats de fichiers : https://www.cines.fr/archivage/des-expertises/les-formats-de-fichier/
- Software Heritage (archive de logiciels) : https://www.softwareheritage.org/?lang=fr


RÉUTILISATION ET VALORISATION DES DONNÉES
cnrs
Bonnes pratiques de gestion et de partage des données de recherche 28.10.19 P 78

CYCLE DE VIE DES DONNÉES DE RECHERCHE
RÉUTILISATION ET VALORISATION DES DONNÉES

D'après Research data lifecycle – UK Data Service
https://www.ukdataservice.ac.uk/manage-data/lifecycle



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

Ressources pour rechercher des jeux de données :
- Metadata Search : https://search.datacite.org/ui
- re3data : https://www.re3data.org/
- OAD : http://oad.simmons.edu/oadwiki/Data_repositories
- openDOAR : http://v2.sherpa.ac.uk/opendoar/

RÉUTILISATION ET VALORISATION DES DONNÉES
DATA JOURNAL – DATA PAPER

- Un **data paper** = publication qui décrit des jeux de données de recherche et les métadonnées associées
- C'est un article à part entière, suivant le même processus éditorial que les articles scientifiques classiques

- Deux possibilités de publication d'un data paper :
  - dans un **data journal** (revue dédiée à ce type de publication)
  - dans une revue classique

Structure d'un data paper :
Partie descriptive :
- Eléments communs aux articles classiques : titre, résumé, mots-clés…
- Eléments 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.

Ressources :
- Exemple de data paper : Tracking vegetation phenology across diverse North American biomes using PhenoCam imagery
- Exemple de data journal : Journal of Physical and Chemical Reference Data (AIP Publishing)
- Exemples de revues publiant des data papers : https://coop-ist.cirad.fr/aide-a-la-publication/rediger/data-paper/5-liens-utiles-exemples-et-guides
- DoRANUm : https://doranum.fr/data-paper-data-journal/

RÉUTILISATION ET VALORISATION DES DONNÉES
EXPOSITION ET VISUALISATION DES DONNÉES

- Il peut s'avérer utile, surtout dans le cas de données nombreuses et complexes, d'exposer ses données **sous forme visuelle** (cartographies, graphiques, etc.) via une plateforme, **en complément** du dépôt dans un entrepôt

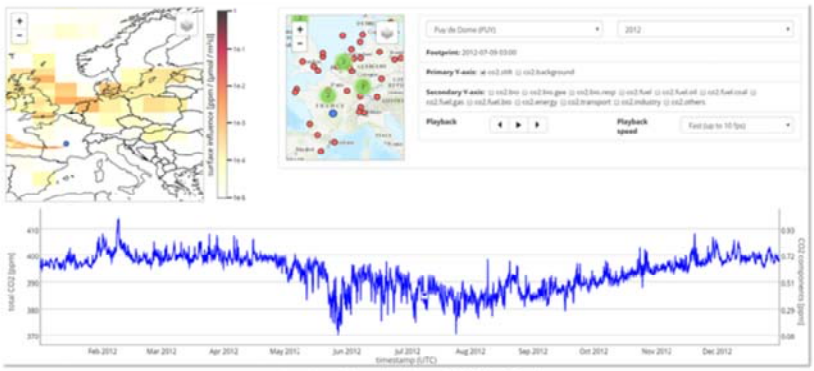


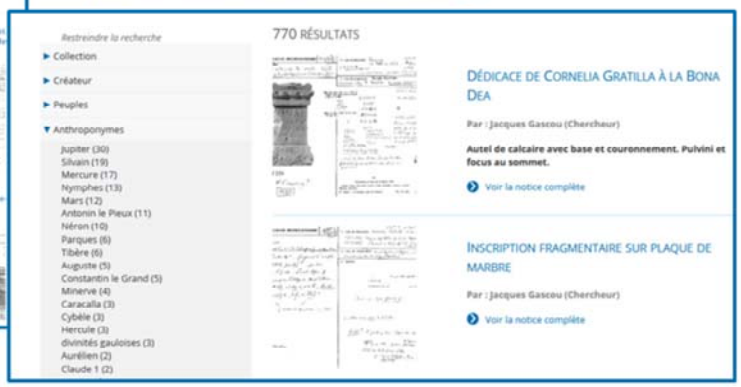
Image issue d'une recherche sur ICOS Carbon Portal

Exemple : Les données disponibles sur le portail ICOS Carbon Portal proviennent de séries chronologiques de valeurs sur des centaines de paramètres. On peut voir par exemple à l'aide d'outils de visualisation l'évolution des concentrations de CO2 sur une année, couplée à l'origine de la masse d'air. Chose qui serait très difficile à appréhender sans passer par la visualisation de données.

RÉUTILISATION ET VALORISATION DES DONNÉES
EXPOSITION ET VISUALISATION DES DONNÉES

- L'outil Omeka est un logiciel libre couramment utilisé pour l'exposition et la visualisation des données.

Exemple : Bibliothèque numérique pour la documentation archéologique du Centre Camille Julian, réalisée avec Omeka. Permet de naviguer dans des corpus et ressources en archéologie.



Exemple : Corea
Http://ccj-corea.cnrs.fr/

Ressources pédagogiques sur DoRANum : https://doranum.fr/acces-visualisation/

A RETENIR

A RETENIR

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

1. Stockage sécurisé durant le projet	2. Sauvegarde pour le partage	3. Archivage pérenne
Serveur sécurisé de l'institution Espace collaboratif de travail	Entrepôt Entrepôt généraliste ou disciplinaire	ARCHIVES Plateforme d'archivage (ex : CINES)

Sauvegardes

Métadonnées

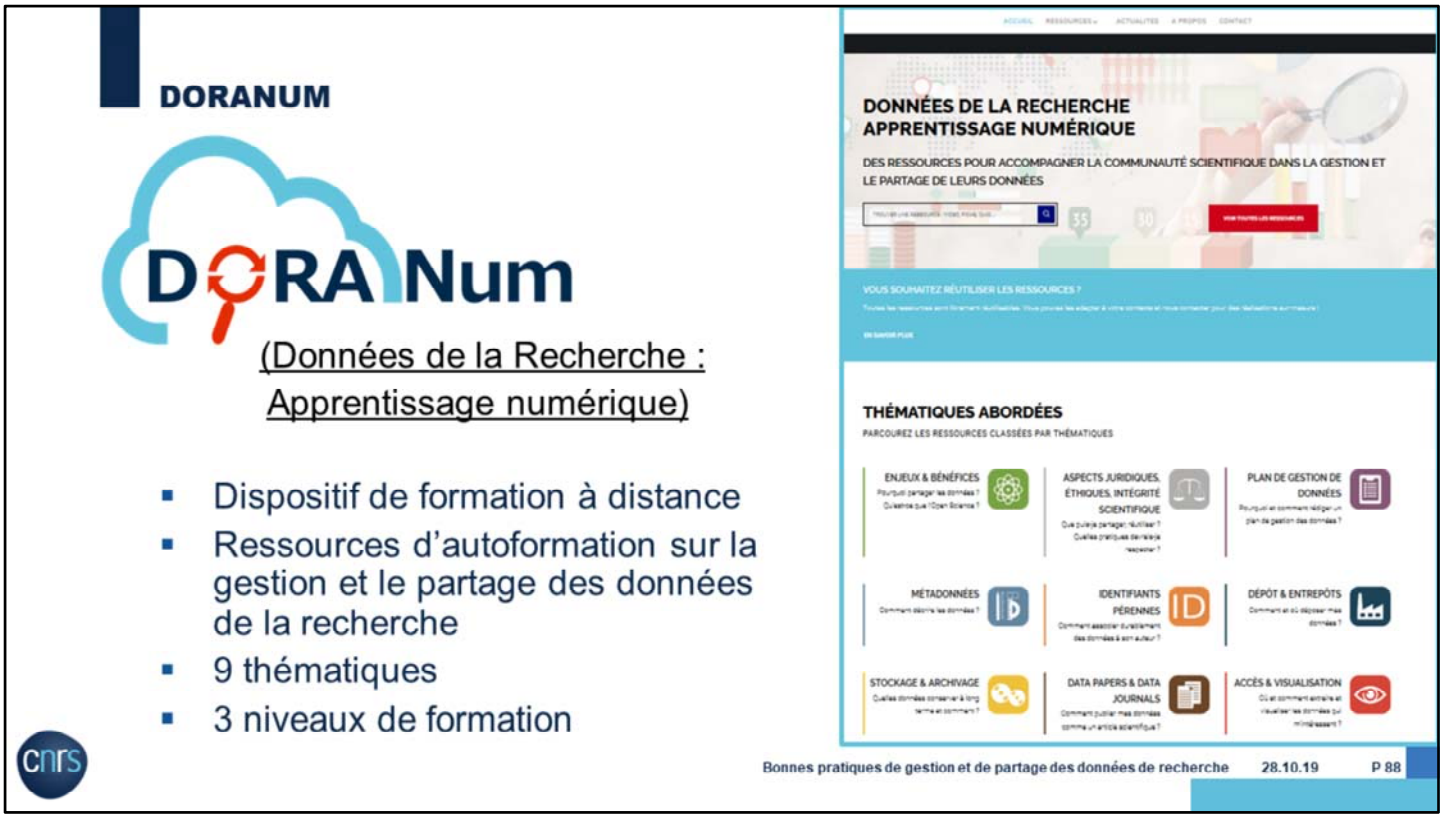
Identifiants pérennes

Licences

Findable Accessible Interoperable Reusable

Le DMP n'est pas un document qui oblige à l'ouverture des données mais qui accompagne et renseigne les données de la recherche à toutes les étapes du cycle de vie des données.
Dans le DMP, vous devez donc renseigner tout ce qui sera fait, comment cela sera fait, où, avec quels moyens etc. Pour cela, le chercheur peut se faire accompagner.
En cas d'empêchement, vous devrez en expliciter les raisons.
C'est le principe « aussi ouvert que possible, aussi fermé que nécessaire ».
Le DMP est surtout un outil de gestion qui permet d'anticiper et évaluer chaque aspect en accord avec les principes FAIR.

Pour aller plus loin...



https://doranum.fr

Ressources d'autoformation sur la gestion et le partage des données de la recherche : fiches synthétiques, minutes, vidéos, interviews, infographies, cours, présentations, modules de formation, tutoriels, guides, quiz, glossaire, outils et services…

3 niveaux de formation : en bref, l'essentiel, pour aller plus loin

9 thématiques :
- enjeux et bénéfices,
- aspects juridique, éthiques, intégrité scientifique,
- Plan de gestion de données,
- Métadonnées,
- Identifiants pérennes,
- Dépôt et entrepôts,
- Stockage et archivage,
- Data papers et data journals,
- Accès et visualisation.

WEBOGRAPHIE

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