

Stages ●●● Formations

Gérer, diffuser ses données de recherche : Introduction aux enjeux, méthodes, pratiques

Florence Thiault

URFIST de Rennes

(MAJ support M.L. Malingre/ F. Thiault, 2019)

Plan

- 1/ Introduction : Repères sur l'émergence d'une problématique
- 2/ Qu'entend-on par données de recherche ?
- 3/ Quelles questions juridiques posent les données de recherche ?
- 4/ Qu'est-ce qu'un PGD (Plan de Gestion de données) ? A quoi sert-il ?
- 5/ Stocker, sauvegarder, archiver ses données : quelles pratiques, quels enjeux, quelles solutions ?
- 6/ Diffuser, partager ses données : pourquoi, comment et où ? Que peut-on partager ?
- 7/ Ressources

1/ Introduction : Repères sur l'émergence d'une problématique

Repères sur l'émergence d'une problématique

- **Accroissement du volume des données de recherche exploitables**

Big Data : le déluge des données

Unité de mesure :
Le pétaoctet...



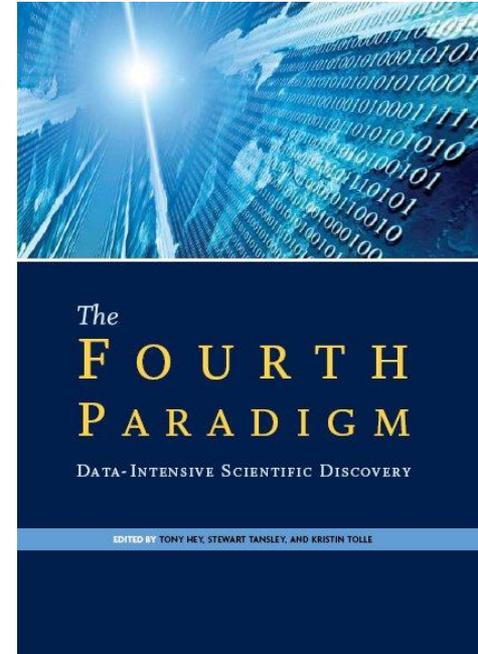
Comment gérer
ces données, les
stocker, les
partager ?



Repères sur l'émergence d'une problématique

• 4^e paradigme de la science

- 1^{er} paradigme (il y a 1000 ans) : **recherche empirique** pour décrire les phénomènes naturels
- 2^e paradigme (depuis 100 ans) : **recherche théorique**, à partir de modèles et de généralisations
- 3^e paradigme (depuis quelques décennies) : **recherche informatisée** basée sur des simulations de phénomènes complexes
- 4^e paradigme (aujourd'hui) : **exploration de données, intégration des théories, simulations et expérimentations**



Source : Jim Gray, cité par Joachim Schöpfel

- Escience / data-based science / data-intensive science / data driven science
- Pratique de la science passe par les données et les logiciels

« *La science du XXI^e siècle est la science de l'exploration des données* » (F. André)

Repère sur l'émergence d'une problématique

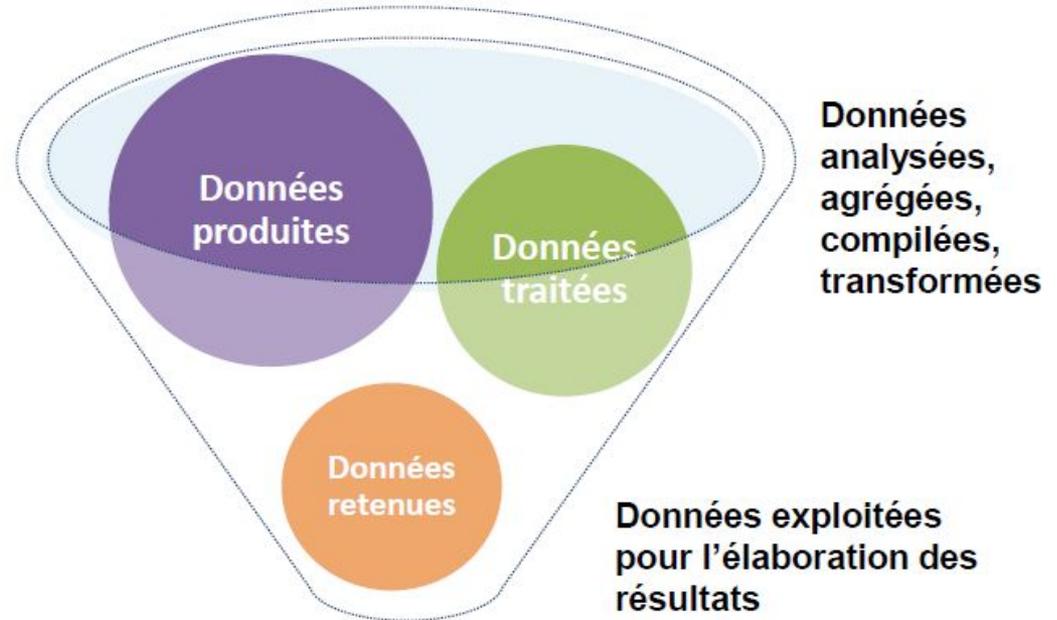
Que deviennent les données ?

INVISIBLE

Perte d'information importante au cours du processus de publication

VISIBLE

Données primaires
produites ou collectées
au cours de la recherche



Données publiées

Données intégrées à titre d'illustrations ou de résultats dans une publication sur support papier ou en ligne

Repères sur l'émergence d'une problématique

Enjeux liés à leur utilisation

- Rôle des données dans la validation de la recherche, rôle de la donnée comme guide de la recherche
- Maîtrise des données

Il importe de :

- Traiter ces masses d'information pour les analyser
- Réaliser une gestion des données permettant de les retrouver et de les rendre compréhensibles
- Développer des services d'accès
- Garantir la fiabilité et la traçabilité des données

Repères sur l'émergence d'une problématique

Enjeux liés à leur conservation

- Garantir de la perte de données ou de leur altération
- Protéger les données
- Permettre leur éventuelle réutilisation dans un autre contexte de recherche
- Conserver une trace de la recherche (patrimoine scientifique)

Il importe de :

- Définir les modalités de diffusion, de conservation, d'archivage des données (sélection et préparation des données, migration vers des formats de données sans perte de qualité, entrepôts...)

Repères sur l'émergence d'une problématique

Enjeux liés à leur partage et leur réutilisation

- Permettre et faciliter l'accès le plus large
- Eviter les phénomènes de privatisation et d'appropriation des données
- Améliorer la qualité, la visibilité, l'impact de la recherche
- Favoriser les collaborations scientifiques demandant le partage ou l'élaboration collective de données

Il importe de :

- Documenter les données et les décrire de manière très précise
- Définir des conditions d'accès, des droits de réutilisation
- Contrôler la compatibilité, l'interopérabilité des formats, la lisibilité
- Procéder à une Dé-contextualisation – re-contextualisation

**« Aussi ouvert que possible, aussi fermé que nécessaire »
Les données peuvent relever de droits multiples**

Repères sur l'émergence d'une problématique

- **Dans le contexte du mouvement global de l'accès aux données**

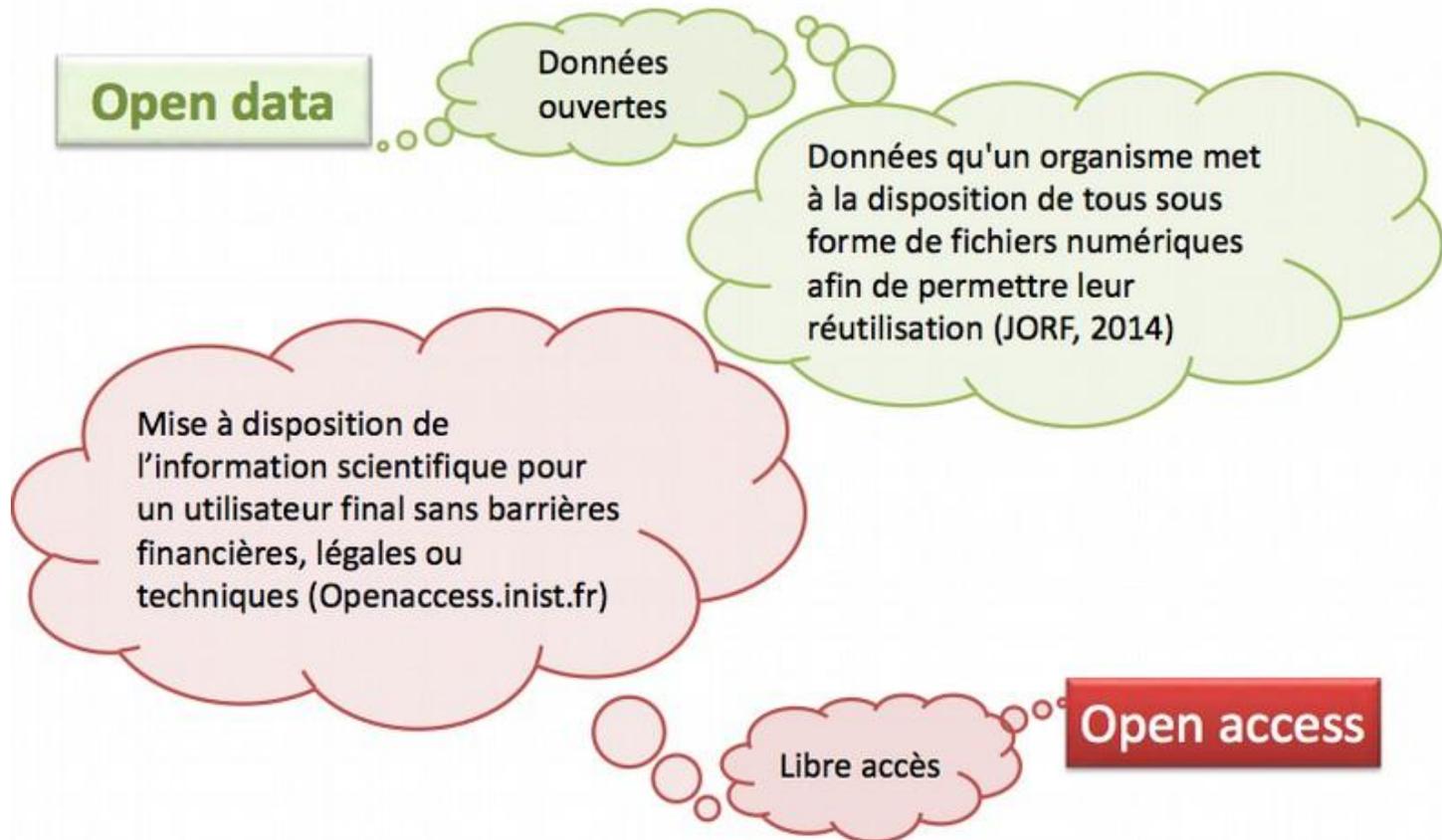
- Open Access : [Déclaration de Berlin](#) sur le libre accès (octobre 2003) : libre circulation des productions scientifiques et de leurs données.



« Les contributions au libre accès se composent de résultats originaux de recherches scientifiques, de données brutes et de métadonnées, de documents sources, de représentations numériques de documents picturaux et graphiques, de documents scientifiques multimédias. »

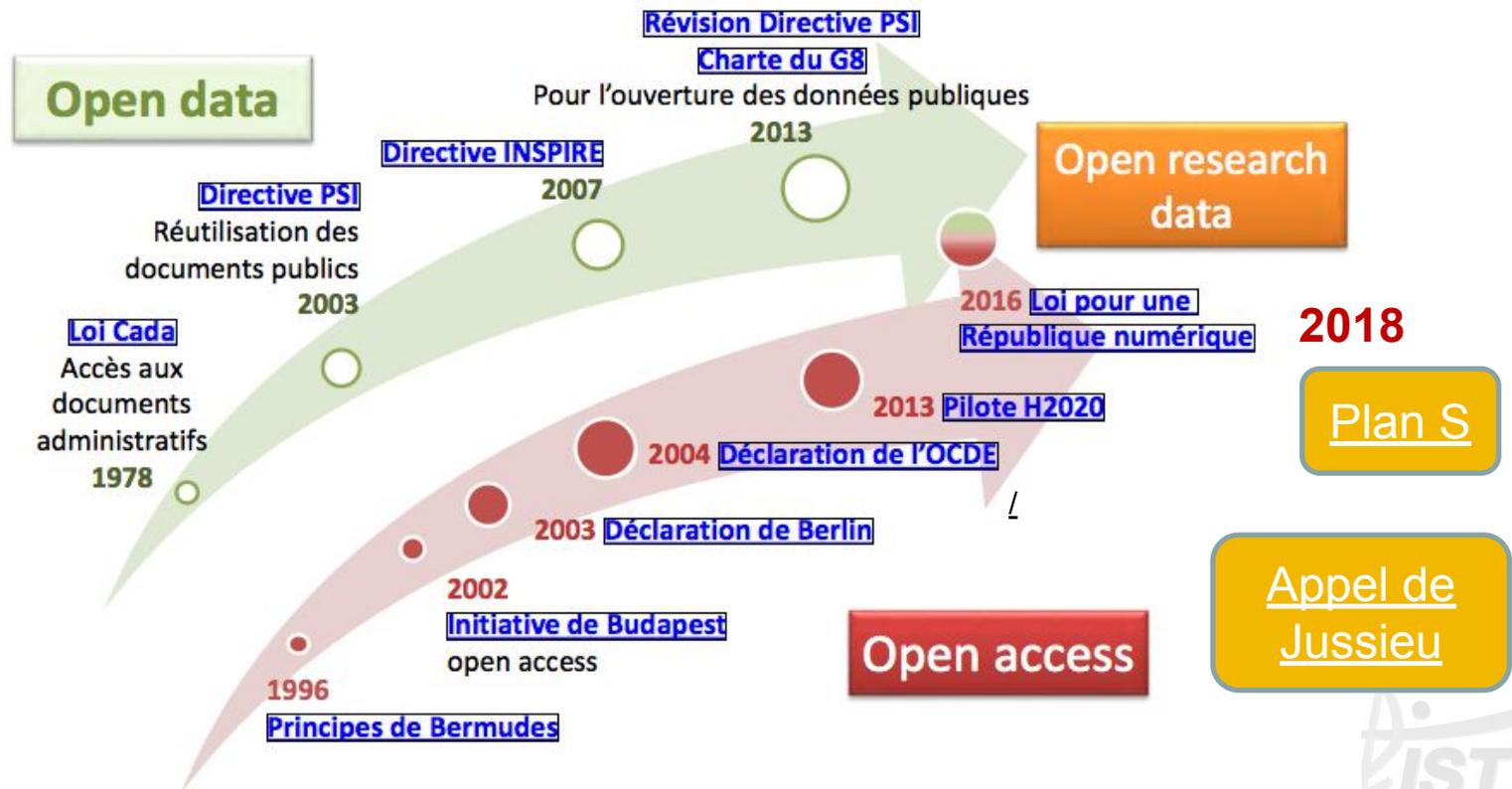
Repères sur l'émergence d'une problématique

Open data - Open access



Repères sur l'émergence d'une problématique

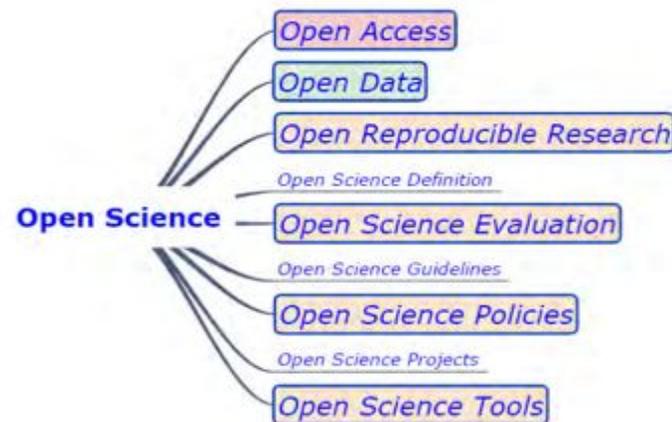
Open data, Open access, Open research data



Repères sur l'émergence d'une problématique

Vers l'Open Science

Mouvement pour rendre la recherche scientifique, les données accessibles à tous les niveaux de la société (traduit de FOSTER)



D'après « The taxonomy tree » <https://www.fosteropenscience.eu/foster-taxonomy/open-science>

« Qu'est-ce que la Science ouverte ?

L'Open Science est une nouvelle approche transversale de l'accès au travail scientifique, des visées et du partage des résultats de la science mais aussi une nouvelle façon de FAIRE de la science, en ouvrant les processus, les codes et les méthodes. »

DIST-CNRS (2016). Livre blanc – Une Science ouverte dans une République numérique

Plusieurs jalons pour l'open research data

Problématique de l'accès aux données de la recherche financée sur fonds publics

- **Le rapport de l'OCDE (2007)**

[Principes et lignes directrices pour l'accès aux données de la recherche financée sur fonds publics]

- « La valeur des données réside dans leur exploitation. L'accès total et ouvert aux données scientifiques devrait devenir la norme internationale pour l'échange des données scientifiques issues de la recherche financée sur fonds publics »

- **Programme de recherche européen Horizon 2020**

- Projet pilote lancé en 2013 sur les données de recherche
- Fait « du libre accès aux données de recherche la règle générale »

- **Programme cadre Horizon Europe pour 2021-2027 dans la continuité**

Les politiques actuelles

En France :

- **Loi pour une République numérique (octobre 2016)**
 - Article 30 : conditions de libre réutilisation des données de recherche

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

Les politiques actuelles

- **2018 : Plan national pour la Science Ouverte.**
 - Pour les données : principe d'ouverture par défaut pour toutes les données publiées dans le cadre d'appels à projet sur fonds publics
 - Lancement du CoSO, Comité pour la science ouverte

15 octobre



COMITÉ POUR LA SCIENCE OUVERTE

Lancement

DISCOURS DE FRÉDÉRIQUE VIDAL | LANCEMENT DU
PLAN NATIONAL POUR LA SCIENCE OUVERTE



4 juillet

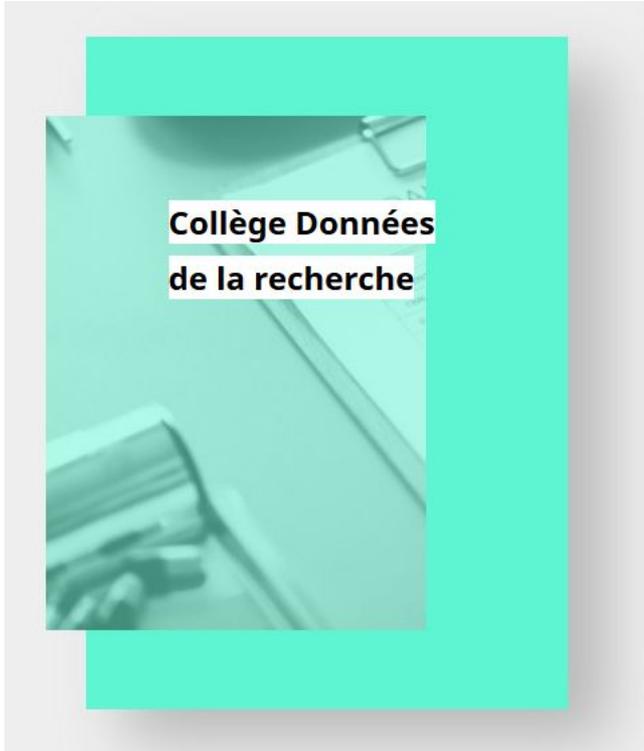
« La science est un bien commun que nous devons
partager le plus largement possible ».

Les politiques actuelles

- **Axe 2 du PNSO « Structurer et ouvrir les données de la recherche »**
- **Collège Données de la recherche du CoSO**

MESURES

- 4  Rendre obligatoire la diffusion ouverte des données de recherche issues de programmes financés par appels à projets sur fonds publics.
- 5  Créer la fonction d'administrateur des données et le réseau associé au sein des établissements.
- 6  Créer les conditions et promouvoir l'adoption d'une politique de données ouvertes associées aux articles publiés par les chercheurs.

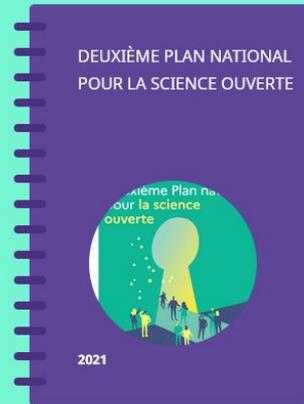


Collège Données
de la recherche

Les politiques actuelles

**Les nouvelles
mesures pour la
science ouverte
2021-2024**

DÉCOUVRIR LE 2E PLAN



PNSO 2

4 axes

- **PREMIER AXE : GÉNÉRALISER L'ACCÈS OUVERT AUX PUBLICATIONS**
- **DEUXIÈME AXE : STRUCTURER, PARTAGER ET OUVRIR LES DONNÉES DE LA RECHERCHE**
- **TROISIÈME AXE : OUVRIR ET PROMOUVOIR LES CODES SOURCES DE LA RECHERCHE**
- **QUATRIÈME AXE : TRANSFORMER LES PRATIQUES POUR FAIRE DE LA SCIENCE OUVERTE LE PRINCIPE PAR DÉFAUT**

PNSO 2

Les politiques actuelles

DEUXIÈME AXE : STRUCTURER, PARTAGER ET OUVRIR LES DONNÉES DE LA RECHERCHE



PNSO 1

4. **Rendre obligatoire la diffusion ouverte des données de recherche** issues de programmes financés par appels à projets sur fonds publics
5. **Créer la fonction d'administrateur des données** et le réseau associé au sein des établissements
6. **Créer les conditions et promouvoir l'adoption d'une politique de données ouvertes associées aux articles publiés par les chercheurs.**

PNSO 2

4. **Mettre en œuvre l'obligation de diffusion ouverte** ou de partage des données de recherche financées sur fonds publics
5. **Créer Recherche Data Gouv**, la plateforme nationale fédérée des données de la recherche
6. **Promouvoir l'adoption d'une politique de données** prenant en compte **l'ensemble du cycle des données** de la recherche avec pour ambition de rendre les données **Faciles à trouver, Accessibles, Interopérables et Réutilisables**
 - Développer et structurer les offres d'accompagnement et les outils à disposition des chercheurs
 - Reconnaître et amplifier la réutilisation des données de la recherche
 - Coordonner et promouvoir la politique d'ouverture des données

Les politiques actuelles

- **Plan d'action 2019 de l'ANR et ouverture des données de recherche**

« **Engagée dans une politique de science ouverte [...], l'ANR appelle les coordinateurs à considérer la question des données de recherche au moment du montage et tout au long du projet.** L'Agence demande à ce que les publications consécutives aux projets qu'elle finance, soient déposées en texte intégral dans une archive ouverte, et imposera de plus un plan de gestion des données (DMP) pour les projets financés à partir de 2019. »



Modèle de DMP / PGD ANR sur DMP Opidor

- **Feuille de route du CNRS**

Les politiques actuelles

• Au niveau européen



Lancement en novembre 2018 par la Commission européenne, <https://www.eosc-portal.eu/>

Offrir un environnement sûr avec des services ouverts et transparents pour stocker, gérer, analyser et réutiliser les données de recherche

GO FAIR

et international

RDA
(Research Data
Alliance, 2013)

Ouvrir progressivement les données de la recherche existantes au sein des institutions scientifiques et académiques, <https://www.go-fair.org/>

Créer des passerelles pour permettre un partage ouvert des données, <https://www.rd-alliance.org/>
RDA France

2/ Qu'entend-on par données de recherche ?

Définition et typologie des données

Qu'entend-on par données de recherche ?

- « Pour commencer, pourriez-vous définir ce que veut dire données de la recherche ? » [[Schöpfel, Kergosien, Prost, 2017](#)]
 - Difficulté d'une définition
 - **Un premier constat : leur grande diversité**
 - dans leur mode d'obtention, leur nature physique, leur forme, le traitement propre à les rendre lisibles, leur support...
 - Quel(s) dénominateur(s) communs à :

Un questionnaire d'enquête sociologique

QUESTIONS	
1-	A B C D
2-	A B C D
3-	A B C D
4-	A B C D
5-	A B C D
6-	A B C D

Un document d'archives



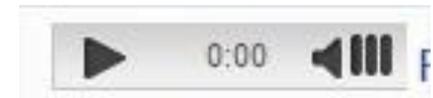
Un rapport de fouilles



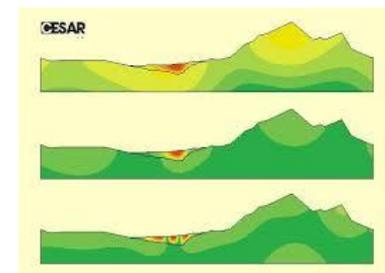
Des séries statistiques



L'enregistrement d'un dialecte

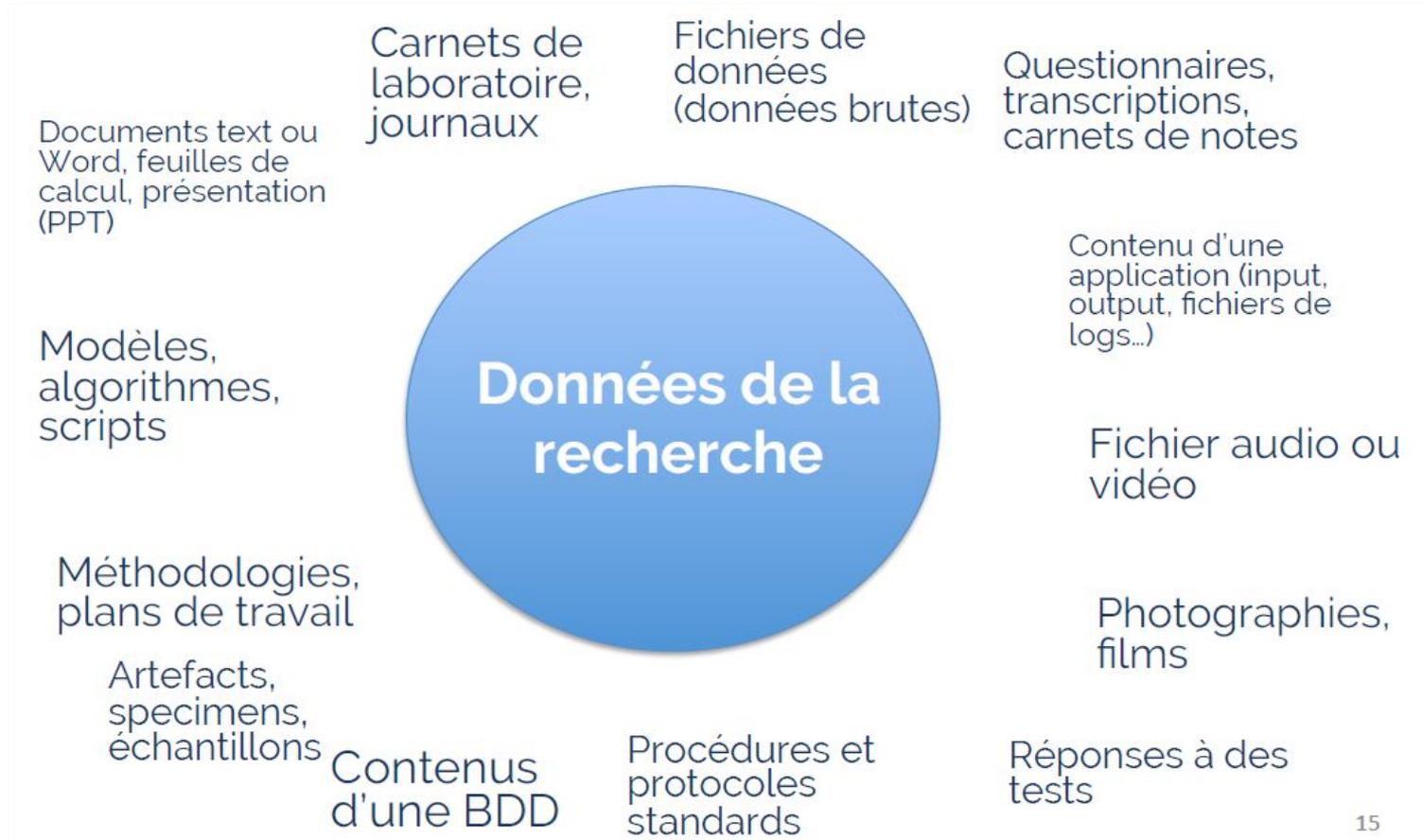


Un modèle de simulations sismiques



Qu'entend-on par données de recherche ?

- **Diversité des données de la recherche**



Qu'entend-on par données de recherche ?



Diversité des données de la recherche

Selon le projet, les données de la recherche peuvent être :

La définition de l'OCDE :

« Les données de la recherche sont définies comme des enregistrements factuels (chiffres, textes, images et sons), qui sont utilisés comme sources principales pour la recherche scientifique et sont généralement reconnus par la communauté scientifique comme nécessaires pour valider les résultats de la recherche »

([OECD Principles and Guidelines for Access to Research Data from Public Funding, 2007](#))

1

produites : ce sont les données créées, élaborées, générées lors d'activités de recherche (observations, mesures...)

2

collectées : ce sont des données déjà existantes (corpus, archives...) qui sont utilisées pour le projet. Les données utilisées peuvent avoir été recueillies initialement dans un autre contexte que celui de la recherche mais elles sont utilisées comme données de recherche dans le cadre du projet.

3

produites et collectées



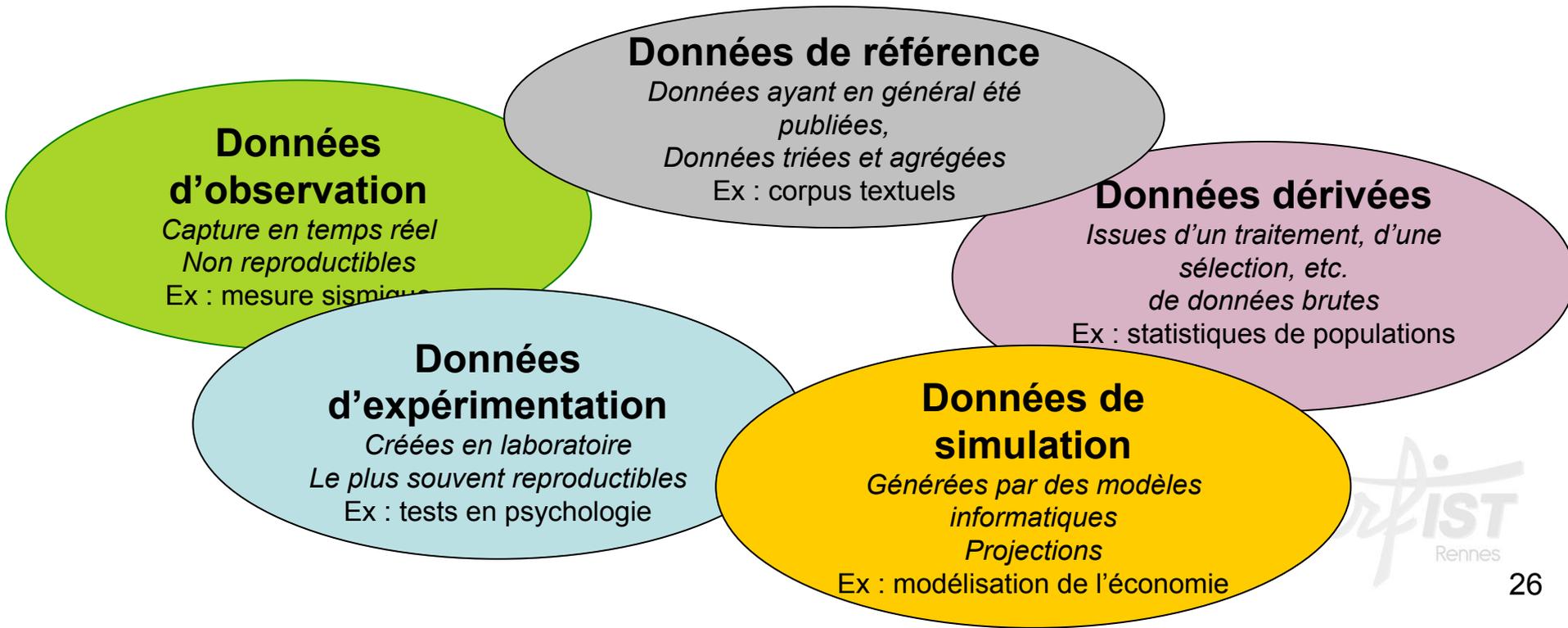
Qu'entend-on par données de recherche ? Typologie

• Approches classificatoires

- Selon leur méthode d'obtention

(André, Francis. Déluge des données de la recherche ? Petit manuel d'immersion. Curation, infrastructures et partage. *Big Data*, 2014)

= Selon les « méthodes et instruments qui les ont générées »



Qu'entend-on par données de recherche ? Typologie

Selon le stade de leur cycle de vie

- Données préliminaires ou préparatoires
- Données « brutes » (acquises lors du processus de recherche, « neutres »)
- Données traitées et analysées

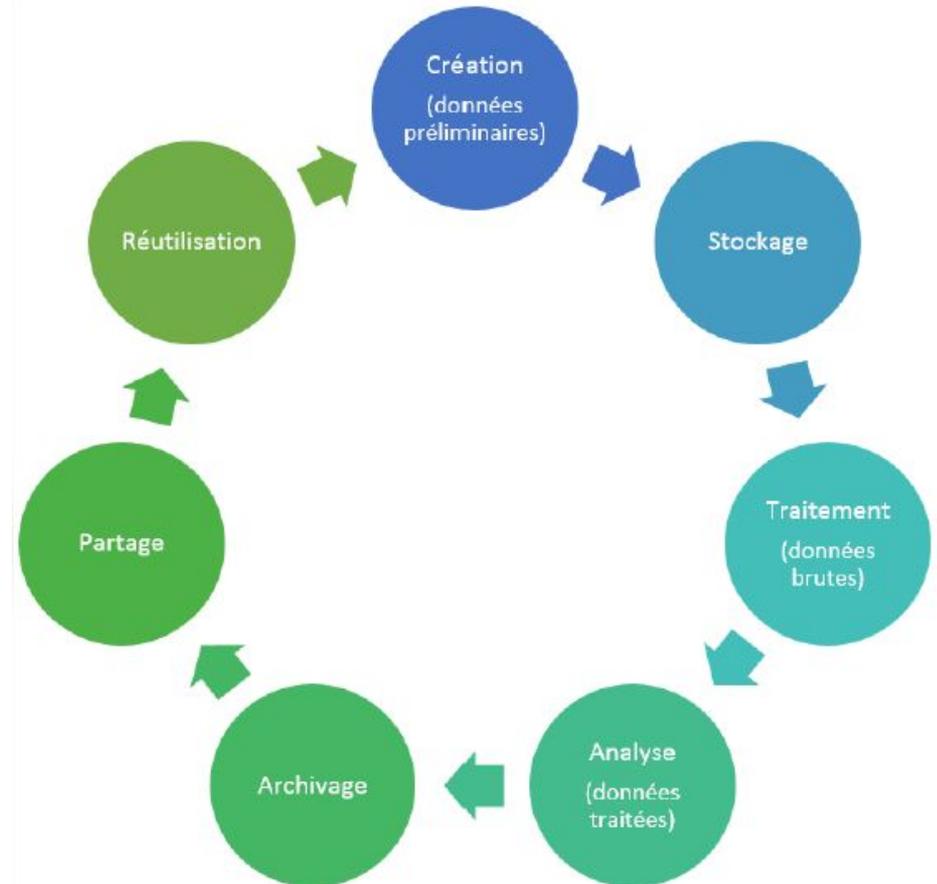


Figure 3 : Le cycle de vie des données de recherche³³

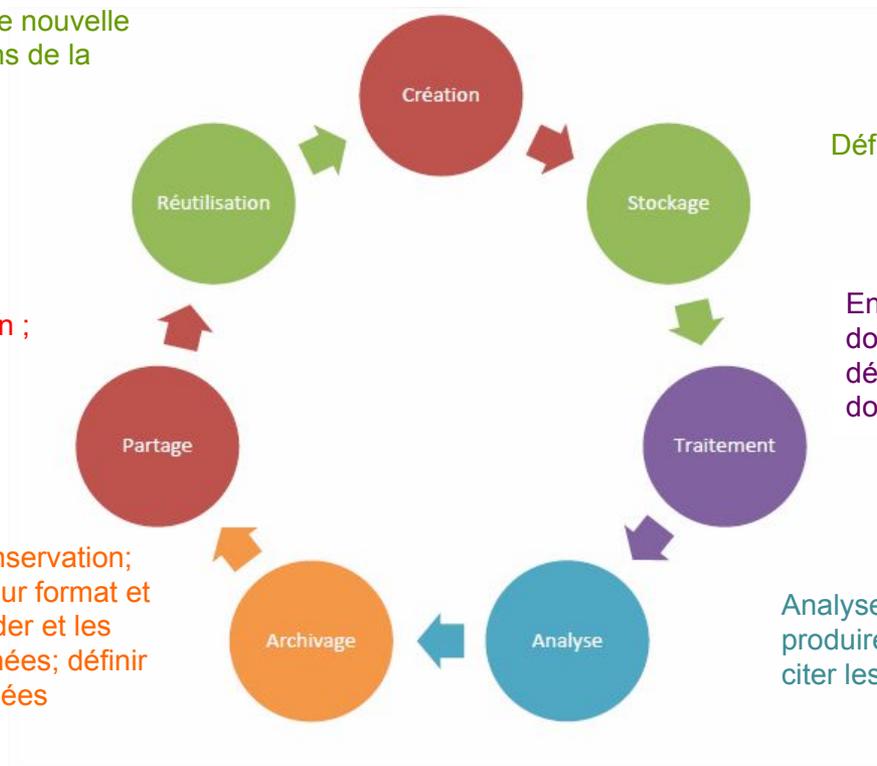
(Pain, Marielou. « Les données de la recherche et leurs entrepôts, de la documentation à la réutilisation : étude de cas pour l'archive HAL », 2016 et Schöpfel, Kergosien, Prost, 2017)

Qu'entend-on par données de recherche ? Typologie

Le cycle de vie des données

Collecter / créer les données; capturer les données avec les métadonnées; acquérir des données existantes d'un tiers.

Réaliser des analyses secondaires, un suivi de recherche, une nouvelle recherche, des évaluations de la recherche...



Définir les droits d'auteur et d'utilisation; contrôler les métadonnées et la documentation ; faire des copies de sauvegarde; publier et partager les données.

Préparer les données pour la conservation; migrer les données vers le meilleur format et un support adapté; les sauvegarder et les décrire dans un entrepôt de données; définir les accès et promouvoir les données archivées

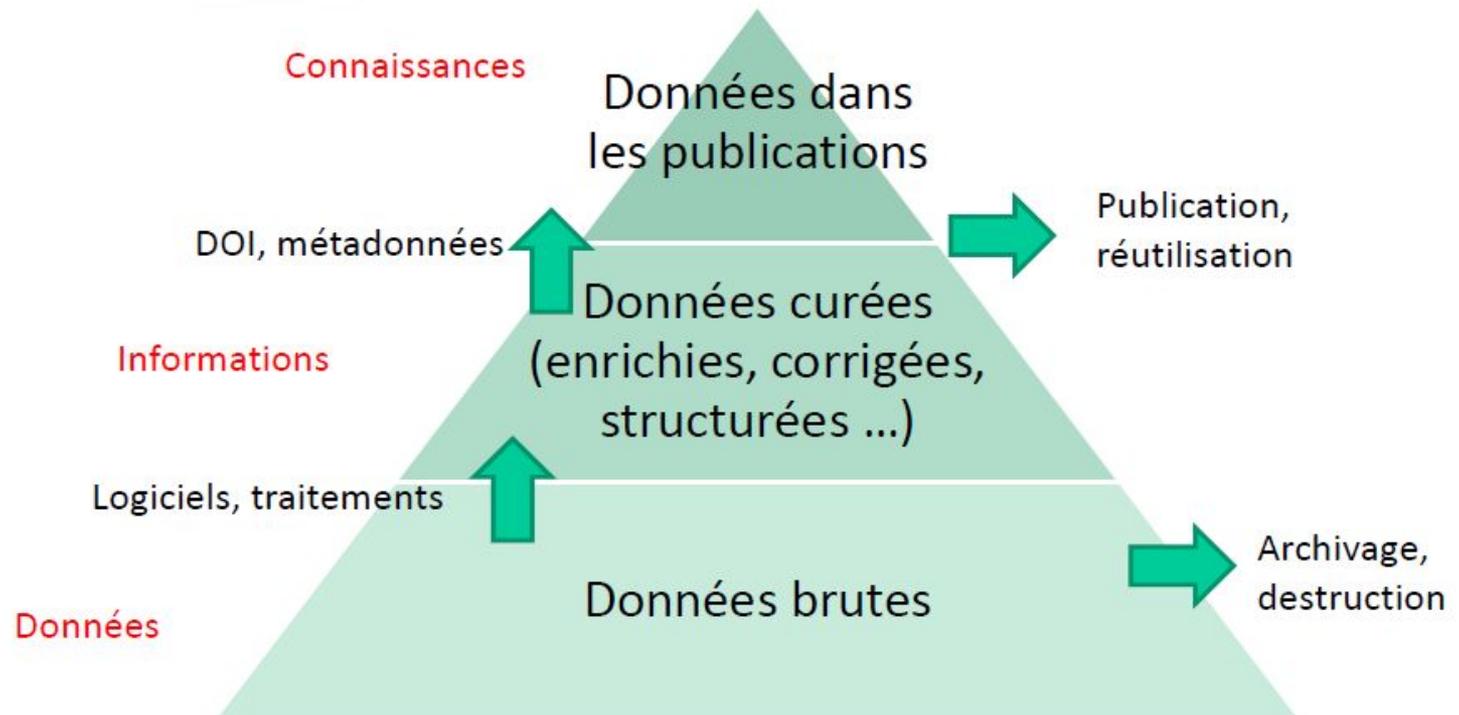
Analyser et interpréter les données ; produire les résultats de recherche ; citer les sources de données.

Définir un espace de stockage des données

Entrer, numériser, transcrire, traduire les données ; vérifier, valider nettoyer, anonymiser ; dériver des données ; décrire et documenter les données ; gérer et sauvegarder les données.

Qu'entend-on par données de recherche ? Typologie

Les états de la donnée



3/ Quelles questions juridiques posent les données de recherche ?

Quelles questions juridiques posent les données de recherche ?

Loi pour une république numérique, 2016

- **Article 30** : intègre la problématique des données de recherche
- Principe **d'ouverture par défaut**, mise à disposition et libre réutilisation **sous conditions**
 - Libre réutilisation totale sauf si les données « *sont protégées par un droit spécifique ou une réglementation particulière.* »
 - Licence correspondante : **CC0**
- **Lionel Maurel** : Quel statut pour les données de recherche après la loi numérique ?
- **Article 38 sur le TDM** : modifiant le Code de la propriété intellectuelle
 - « *Un décret fixe les conditions dans lesquelles l'exploration des textes et des données est mise en oeuvre, ainsi que les modalités de conservation et de communication des fichiers produits au terme des activités de recherche pour lesquelles elles ont été produites ; ces fichiers constituent des données de la recherche* »

Quelles questions juridiques posent les données de recherche ?

Les données peuvent relever de droits multiples pour la collecte.... Et pour la publication

- Propriété intellectuelle (sur l'ensemble ou sur la donnée individuelle)
- Droit des archives
- Droit des bases de données
- Droit des données publiques
- Protection des données personnelles
- Loi Jardé (recherches impliquant la personne humaine)
- Code de la santé publique
- Secret statistique
- Conditions contractuelles (CGU, licences...)
- Cas du droit des bases de données / Loi Lemaire
 - Notion de donnée publique
 - « Droit *sui generis* des producteurs de bases de données ne peut faire obstacle à la réutilisation du contenu des bases de données que les administrations publient. »
(<https://siafdroit.hypotheses.org/659>)

Quelles questions juridiques posent les données de recherche ?

Puis-je diffuser mes données de recherche ?

Veillez répondre aux questions suivantes :	Ce que nous vous conseillons :
<p>1- Le jeu de données contient-il des données personnelles (qui identifient directement ou indirectement des individus) ?</p> <p><input type="radio"/> Oui</p> <p><input type="radio"/> Oui, mais les formalités CNIL sont respectées, ET les individus concernés ont donné leur accord écrit à la diffusion OU les données ont été anonymisées (non conservation des noms, des identifiants et des pseudonymes, cf CNIL et CNRS).</p> <p><input type="radio"/> Non</p> <p><input checked="" type="radio"/> Vous n'avez pas encore répondu</p> <p>2- Avez-vous utilisé d'autres données que celles que vous avez produites vous-mêmes ?</p> <p><input type="radio"/> Oui</p> <p><input type="radio"/> Oui, mais j'ai bien vérifié la licence d'utilisation de la base de données et j'ai le droit de les extraire et de les réutiliser.</p> <p><input type="radio"/> Non</p> <p><input checked="" type="radio"/> Vous n'avez pas encore répondu</p> <p>3- Est-ce que certaines de ces données, produites dans le cadre de l'établissement public, contiennent des données protégées par le droit d'auteur (textes, images, schémas, ...) ?</p> <p><input type="radio"/> Oui</p> <p><input type="radio"/> Oui, mais j'ai obtenu l'accord écrit de l'auteur.</p> <p><input type="radio"/> Non</p> <p><input checked="" type="radio"/> Vous n'avez pas encore répondu</p> <p>4- Est-ce que ces données ont été produites dans le cadre d'un consortium avec le secteur privé ?</p> <p><input type="radio"/> Oui</p> <p><input type="radio"/> Oui, mais elles peuvent être diffusées librement</p> <p><input type="radio"/> Non</p> <p><input checked="" type="radio"/> Vous n'avez pas encore répondu</p> <p>5- Votre Unité est-elle protégée ou en Zone à Régime Restrictif (ZRR) ?</p> <p><input type="radio"/> Oui</p>	<p>Vous devez encore répondre à 8 questions dans la partie de gauche</p> <p>Exporter en PDF</p>

Quelles questions juridiques posent les données de recherche ?

Puis-je diffuser mes données de recherche ?

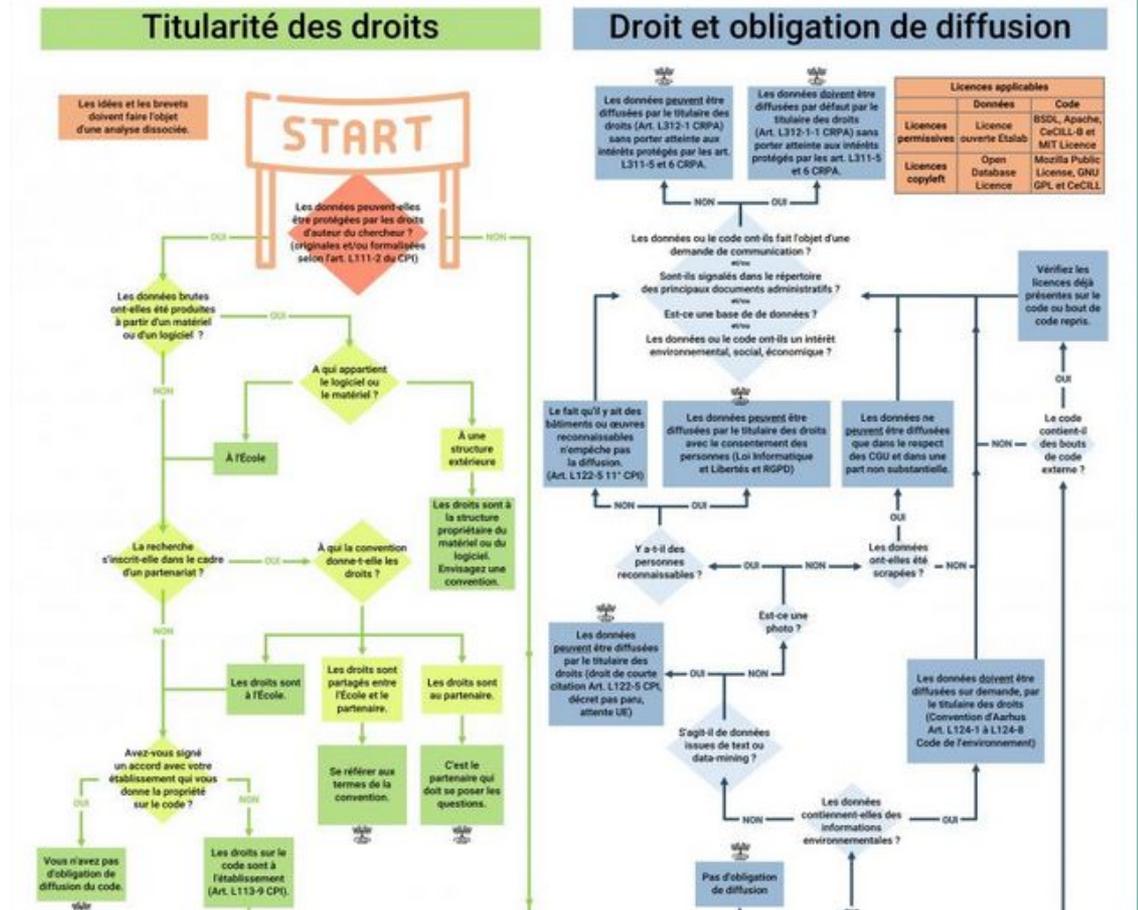
- Vérifier la **titularité des droits**

- En fonction de la titularité des droits, respecter les **droits et obligations de diffusion**

□ Voir [le logigramme](#) de l'Ecole des Ponts Paristech

Qui a les droits ? Qui a les droits ?
Qui a les droits de faire quoi ?

Frédérique Bordignon, Romain Boistel, Delphine Du Pasquier
Ecole des Ponts ParisTech – 2018
Après consultation du cabinet d'avocats August-Debouzy

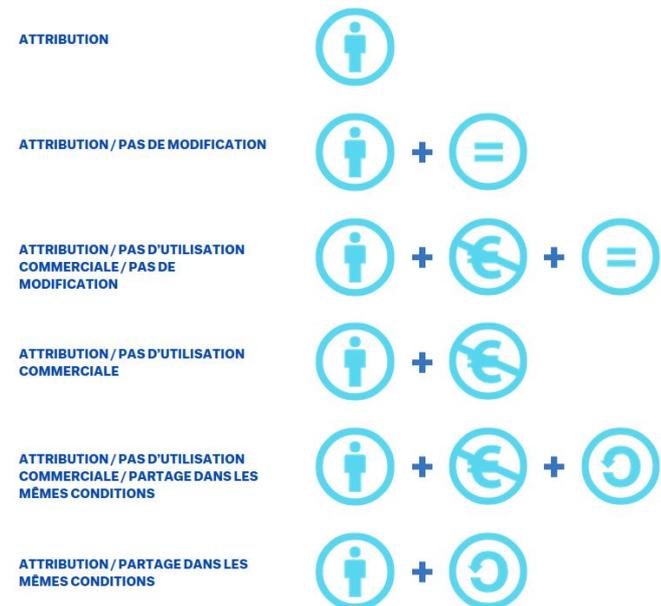
Quelles questions juridiques posent les données de recherche ?

Attribuer une licence à son jeu de données

- Licences de l'Open Knowledge Foundation

- **ODC-by** : Open Database Commons : partage, réutilisation et adaptation sous condition d'attribution
- **ODC-ODbL** : Open Database License : partage, réutilisation et adaptation sous condition d'attribution et de licence à l'identique
- **PDDL** : Public Domain dedication and License : domaine public

- Licences Creative Commons (version 4.0)



Licence ouverte Etalab

- Publication des données publiques

Questions juridiques :

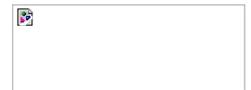
La question de la protection des données personnelles

- **Principe 1 la Finalité** : définir les objectifs du fichier
- **Principe 2 La pertinence** : vérifier la pertinence des données
- **Principe 3 La conservation** : limiter la conservation des données
- **Principe 4 Les droits** : respecter les droits des personnes
- **Principe 5 La sécurité** : sécuriser les données

- Un contact dans les universités :
 - le **DPO** Délégué à la protection des données (DPD) - Data Protection Officer (**DPO**)

Source :

<https://www.cnil.fr/fr/comprendre-vos-obligations/les-principes-cles>



- **Le RGPD** (Règlement général sur la protection des données)
 - Applicable dans les pays de l'Union européenne depuis mai 2018

Quelles questions juridiques posent les données de recherche ?

Préconisations / données personnelles

- Réfléchir à la collecte de ses données
- Demander une autorisation si nécessaire
- Informer les personnes
- Recueillir le consentement si nécessaire en prévoyant tous les usages et en garder une trace
- Prévoir un recours (contact)
- Organiser la sécurité et la confidentialité
- Anonymiser
- Réfléchir à la durée de conservation

Quelles questions juridiques posent les données de recherche ?

Ethique de la recherche et données

- Autorisations et protocoles spécifiques à certaines disciplines et certains objets de recherche
- Respect de la propriété intellectuelle
- Documenter
 - Fiabilité et recueil loyal des données
 - Assurer l'archivage et la traçabilité des données brutes
 - [Charte Ethique & Big data](#)

Pratiquer une recherche intègre et responsable

Comité d'éthique du CNRS (COMETS),
2017

4/ Qu'est-ce qu'un PGD (Plan de Gestion de données) et à quoi sert-il ?

Qu'est ce qu'un PGD et à quoi sert-il ?

• Le Plan de gestion de données (Data Management Plan)

- Un document rendu obligatoire par certains financeurs
- Mais avant tout un document utile au projet de recherche

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

Source : [Doranum](#)

C'est un document **normalisé et évolutif**, qui doit être **mis à jour** tout au long du projet et qui documente les données et indique les **conditions de partage**.

□ **obligation** (initiative pilote H2020) ou **incitation**

Qu'est ce qu'un PGD et à quoi sert-il ?

Document technique stratégique / plan d'action

Document évolutif qui explicite les principes de gestion de vos données dans leurs différentes étapes :

- tout au long de leur cycle de vie : collecte/création, traitement, documentation, stockage, archivage, conservation et partage



Vidéo
Doranum
DMP

Qu'est ce qu'un PGD et à quoi sert-il ?

- Informations administratives du projet
- Identification des jeux de données
- Description des jeux de données
 - Types de données : origine, méthodologie, description formelle et technique, volume...
- Standards et métadonnées
- Partage des données
 - Modalités de dépôt ; Modalités de partage ; Licences
- Archivage des données



A quelles questions répond un PGD ?



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

Ressources

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

Responsabilités dans le projet

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

Collecte des données

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

Documentation des données

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

Accès et partage des données

What's the plan?

Dans le cadre d'un projet de recherche ou non...

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

Sauvegarde des données

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

Propriété intellectuelle

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

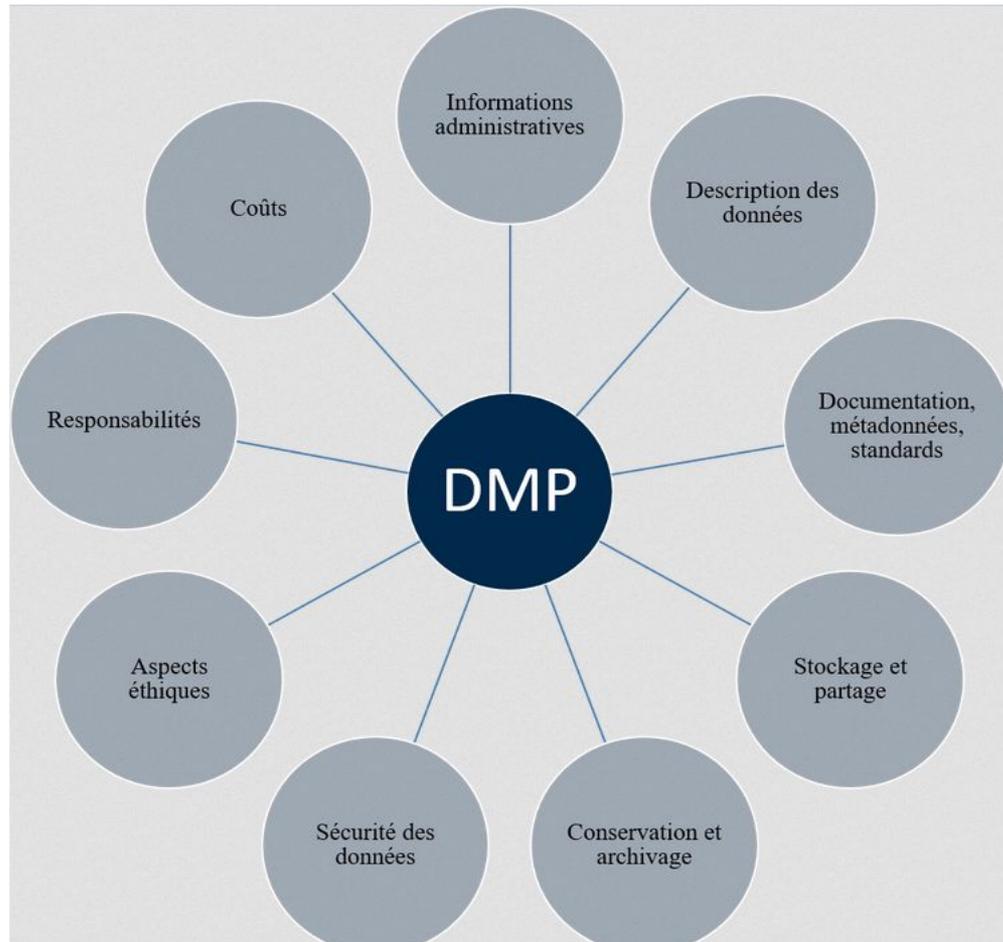
Ethique

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

Archivage et préservation des données

Qu'est ce qu'un PGD et à quoi sert-il ?

- **Les éléments du PGD / DMP**



Il existe différents modèles de PGD, qui comportent cependant les mêmes grandes rubriques.

Un exemple : le modèle [DMP H2020](#) établi par la Commission européenne

Source : [Doranum](#)

Qu'est ce qu'un PGD et à quoi sert-il ?

• La description des données

- fichier texte rassemblant les informations principales sur la recherche
- métadonnées (données sur des données) : de description, de gestion, de préservation et réutilisation

Les métadonnées permettent de répondre aux questions :

- **QUI** a créé les données ?
- **QUE** contiennent-elles ?
- **OÙ** sont-elles ?
- **QUAND** ont-elles été créées ?
- **COMMENT** ont-elles été créées ?
- **POURQUOI** ont-elles été créées ?

- > Permet de retrouver et réutiliser facilement ses données
- > Permet de connaître, évaluer et réutiliser les données des autres

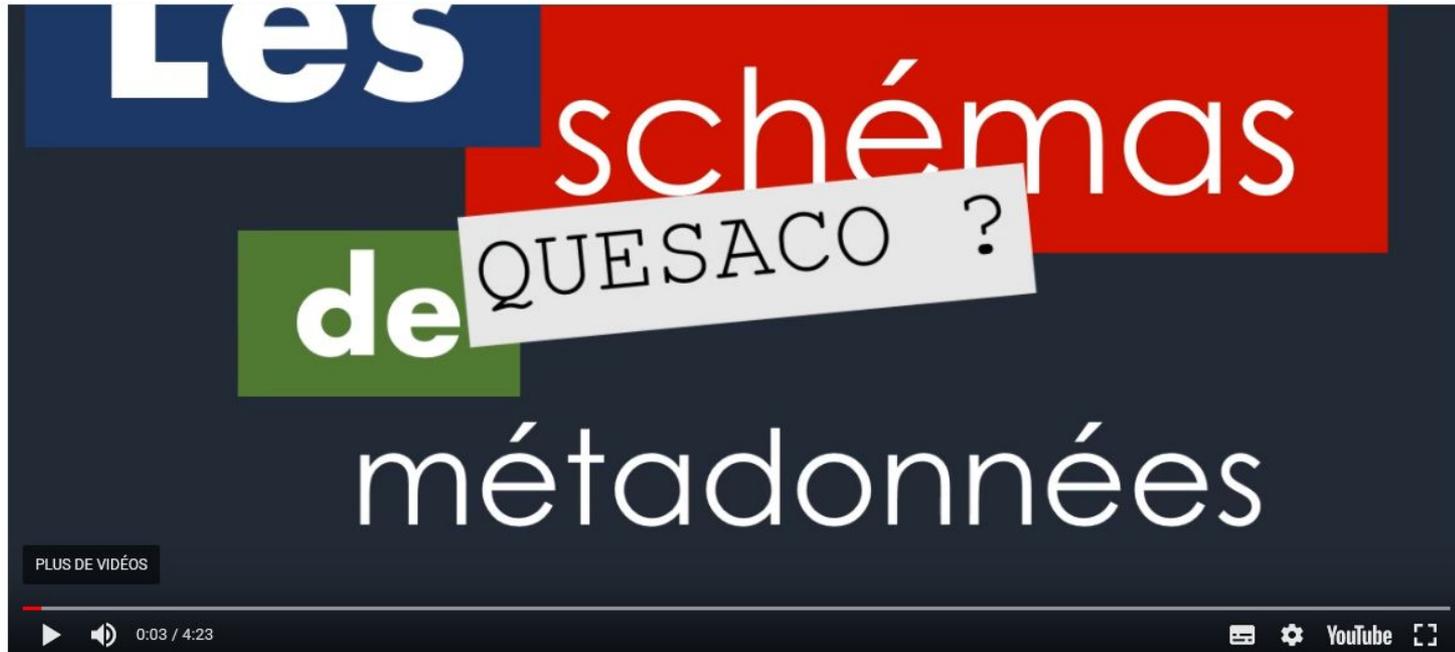
Qu'est ce qu'un PGD et à quoi sert-il ?

• Les standards de métadonnées

Modèles qui préconisent une manière uniforme de décrire les données (avec un vocabulaire contrôlé), selon :

- Le type de ressource (vidéo, image...)
- Le champ disciplinaire (représentation fine de données spécifiques)
- La communauté
- L'entrepôt choisi (certains disposent déjà de standards)

Vidéo
Dorandum



Qu'est ce qu'un PGD et à quoi sert-il ?

Standards généralistes (interdisciplinaires)

Ex : [Dublin Core](#) (norme internationale depuis 2001, utilisée par l'entrepôt de données Nakala), repose sur 15 éléments qui peuvent être étendus

- [DataCite Schema](#)

Standards disciplinaires

- Ex : [Data Documentation Initiative \(DDI\)](#) > Sciences sociales

Listes de standards

- [DCC \(Digital Curation Center\) disciplinary Metadata](#)

Élément simple	Définitions
Title ①	nom de la ressource
Subject	thème du contenu de la ressource
Description	résumé, table des matières...
Creator ②	auteur principal de la ressource
Publisher	entité responsable de la diffusion de la ressource
Contributor	co-auteurs associés à l'élaboration de la ressource
Date	date de création ou mise à disposition
Type ③	nature du contenu : image, son texte...
Format	format ou taille de la ressource
Identifiant	référence univoque, DOI, URL, ISSN...
Source	référence à une ressource à partir de laquelle la ressource actuelle a été dérivée ou créée
Language	langue originale de la ressource
Relation	référence à une ressource apparentée
Coverage	périmètre spatial et temporel
Rights	informations sur les droits associés à la ressource

Standard interdisciplinaire : Dublin Core

Qu'est ce qu'un PGD et à quoi sert-il ?

<https://opidor.fr>



CNRS | DIST | INIST | BSN 10

Recherche sur le site

Outils et services pour accompagner la gestion des données de la recherche
→ [Voir la suite](#)

Les services OPIDoR



Accès direct aux services

DMP OPIDoR
Cat OPIDoR

Coordonnées

Inist-CNRS
2, allée du Parc de Brabois



Qu'est ce qu'un PGD et à quoi sert-il ?



Quelques points d'attention :

- Politiques d'établissement / de laboratoire à respecter
- Responsabilité de la gestion des données
- Propriété des données
- Questions éthiques
- Sur l'organisation des données : quelle démarche qualité / vérification
- Stockage
- Question d'accès et de sécurité des données
- Ressources nécessaires (personnel, argent, matériel)

**5/ Stocker, sauvegarder, archiver ses données :
quelles pratiques, quels enjeux, quelles
solutions ?**

Stocker, sauvegarder, archiver ses données

4 objectifs :
rendre les documents

- **accessibles,**
- en préserver
l'intégrité,
- **l'intelligibilité**
- **la lisibilité**

CONNAÎTRE LES RISQUES



L'OBSOLESCENCE TECHNOLOGIQUE

Depuis plus de 50 ans, l'**évolution rapide des matériels** informatiques (nouvelle génération tous les 5 ans en moyenne) engendre une **obsolescence des logiciels** qui doivent sans cesse rester compatibles avec les nouveaux matériels.



LA DISPARITION DES FORMATS

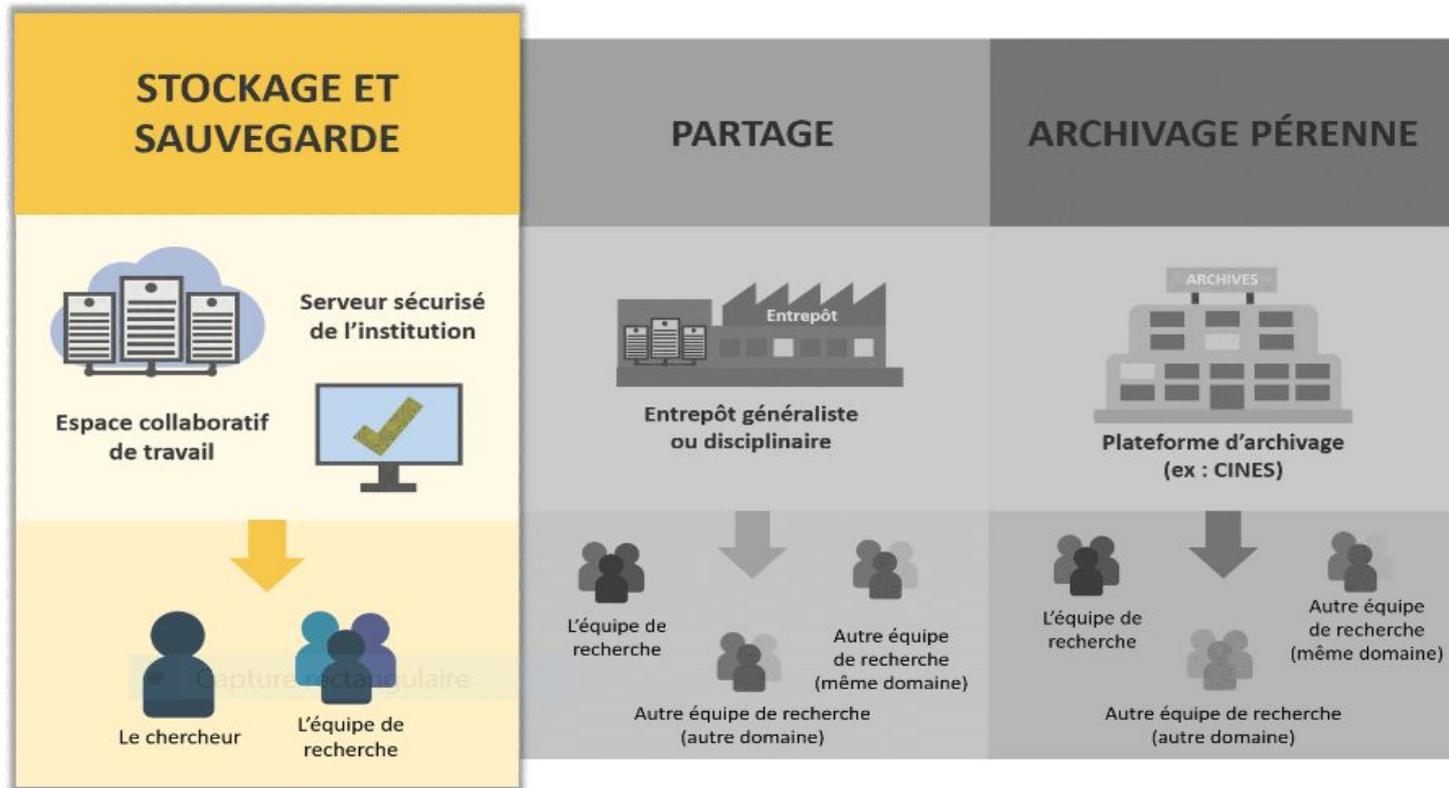
On observe en permanence la disparition de formats devenus obsolètes et l'apparition de nouveaux formats d'enregistrement des données.



LE MANQUE DE DOCUMENTATION

Afin de préserver l'intelligibilité des données, il est indispensable de les accompagner de toutes les métadonnées nécessaires à la bonne compréhension du contexte de leur production.

Stocker, sauvegarder, archiver ses données



• L'organisation des données

- Nommage
- Hiérarchisation
- versionning

Stocker, sauvegarder, archiver ses données

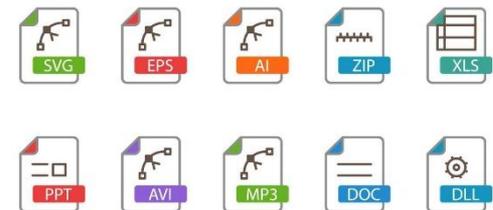
Formats de fichiers

- Privilégier les formats d'échange / les formats ouverts / les formats « standards » dans la discipline
- Pour les images : formats sans compression ou compression sans perte

Excel (.xls, .xlsx)	Comma Separated Values (.csv)	
Word (.doc, .docx)	Texte brut (.txt), or si besoin de formatage, PDF/A (.pdf)	
PowerPoint (.ppt, .pptx)	PDF/A (.pdf)	    
Photoshop (.psd)	TIFF (.tif, .tiff)	
Quicktime (.mov)	MPEG-4 (.mp4)	    

<https://dmptool.org>

[Quiz sur les formats](#)



Stocker, sauvegarder, archiver ses données

Nommage des fichiers

• Utilité

- Se repérer entre les types de fichiers / les versions
- Regrouper les fichiers similaires et classer automatiquement ses fichiers de façon pertinente dans les dossiers
- Faciliter l'archivage et la préservation à long terme des fichiers



20150405_CR



20160310_CR



20160515_CR

• Quelques règles

- Définir un plan de nommage et une arborescence de fichiers
- Noms courts et significatifs
- Pas de caractères spéciaux ni accentués : lettres et chiffres seulement
- Bannir les espaces, vive les _ (tiret du 8) !
- Dates au format américain (AAAAMMJJ)
- Nombres sur 2 unités (ex : 01, 09...)
- Numéros ou codes de versions

Sources :

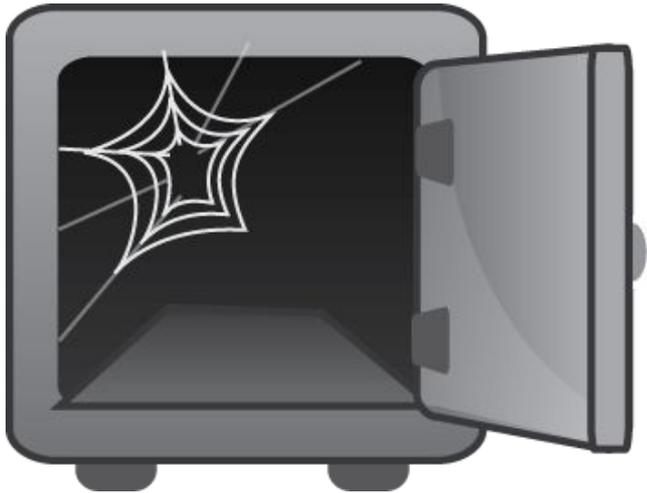
Doranum
Recommandations
Université
d'Edinburgh



Stocker, sauvegarder, archiver ses données

Sauvegarder, c'est dupliquer

- Sur plusieurs supports
- A plusieurs endroits
- Faire évoluer les formats et les supports dans le temps

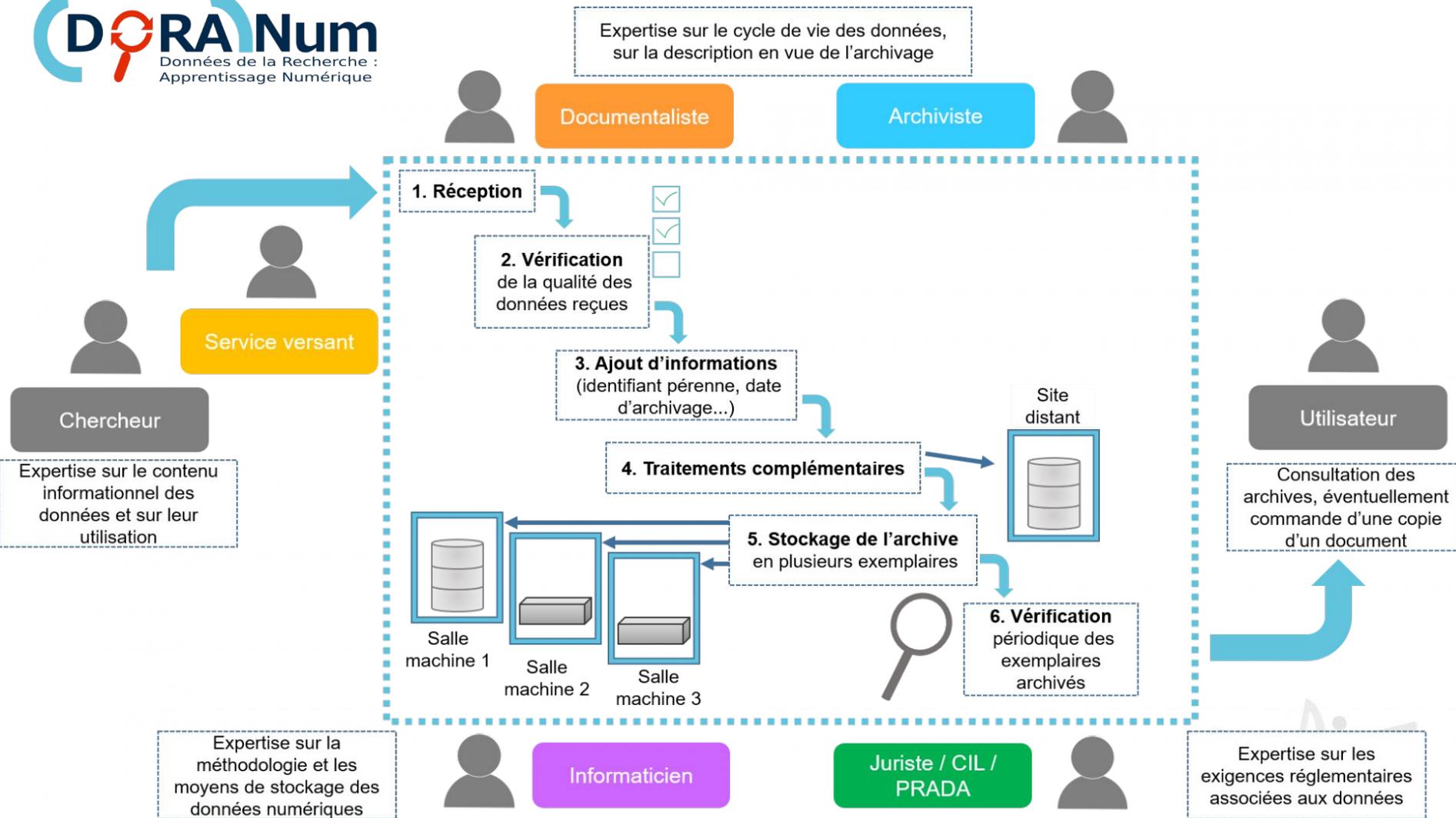


Stockage des données en ligne

- Sauvegarde ou travail dans le Cloud : peut poser des problèmes de sécurité / confidentialité
 - Piratage
 - Localisation des serveurs et lois nationales
 - Sensibilité de certaines recherche
- Attention aux CGU
- Privilégier serveurs institutionnels

Stocker, sauvegarder, archiver ses données

Les étapes et les acteurs du versement d'archives



6/ Diffuser, partager ses données : pourquoi, comment et où ? Que peut-on partager ?

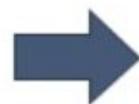
Diffuser, partager ses données

Pourquoi partager les données ? Entre injonctions et bénéfices

 RESEARCH & INNOVATION
Participant Portal H2020 Online Manual
Pilote données ouvertes par défaut



Incitation
ou
obligation
de partage



Reproductibilité
Preuve
Résultats accessibles à tous public
Préservation
Description et visibilité

Augmenter ces citations
Le partage des données
augmente l'attractivité
de l'article jusqu'à 69%

Étendre son réseau
nouvelle opportunité
pour
échanger/collaborer

**Attirer sur son profil
chercheur**
La demande pour réutiliser
les données est en forte
hausse

**Renforcer ses possibilités
de recevoir des
financements**
Compatibilité avec les
exigences de s financeurs

Source : J.C.
Desconnets,
P. Aventurier,
S. Banon.
Entrepôt de
données
ouvertes IRD

Diffuser, partager ses données

Enjeux pour le chercheur

- Transparence, validité des résultats.
- Visibilité et citabilité
- Collaborations scientifiques potentielles (interdisciplinarité, international...)
- Préservation des données
- Amélioration qualité des données
- Réponse exigences des financeurs
- Gain de temps

Enjeux pour la communauté

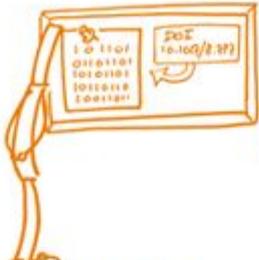
- Promouvoir la transparence des recherches et les collaborations
- Réduire les coûts et maximiser le retour sur investissement
- Promouvoir l'innovation
- Mettre à disposition des ressources pour l'ensemble de la société (société civile, pays moins favorisés)



Diffuser, partager ses données

Comment partager ?

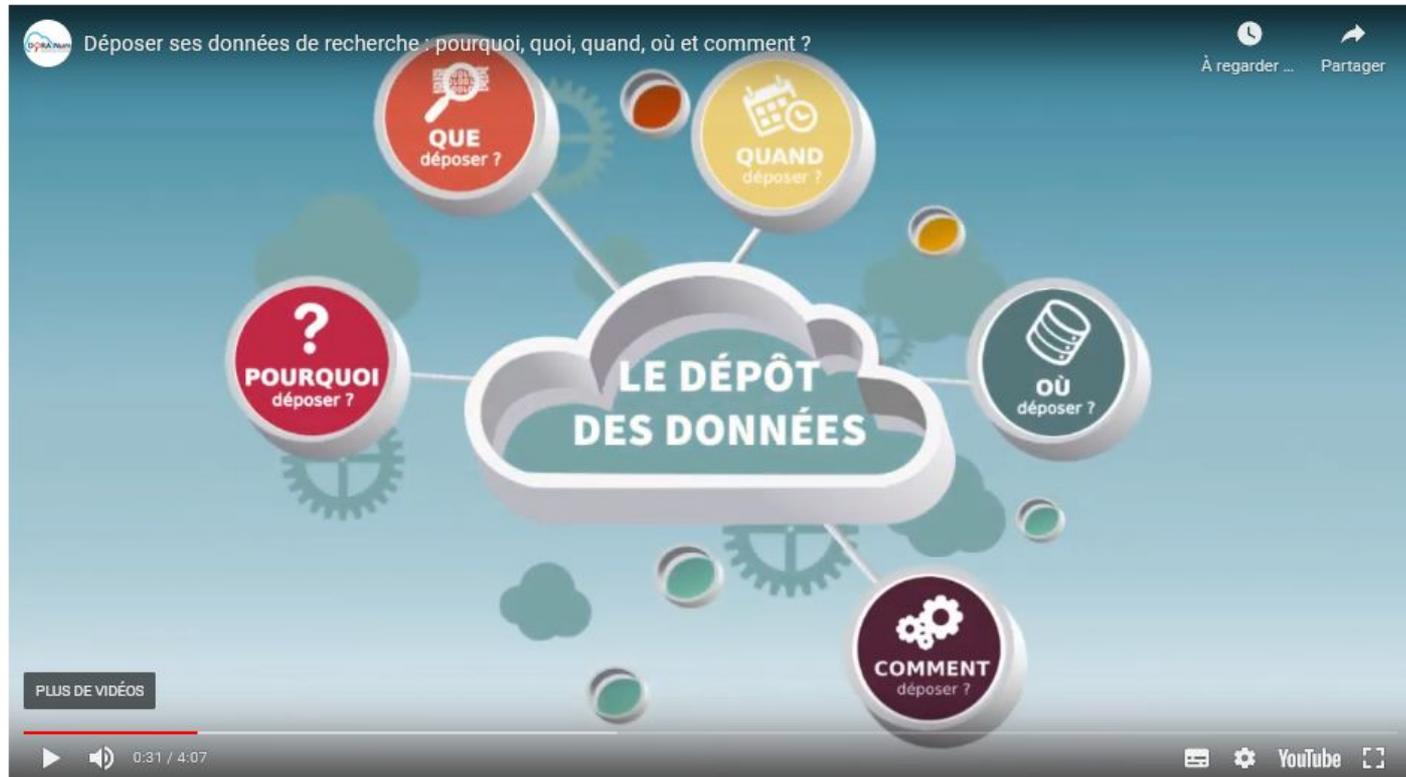
Pratiques diverses suivant les disciplines / Respect du droit

 <p>FINDABLE</p>	 <p>ACCESSIBLE</p>	 <p>INTEROPERABLE</p>	 <p>REUSABLE</p>
<ul style="list-style-type: none"> identifiant pérenne métadonnées enrichies données cherchables et trouvables en ligne 	<ul style="list-style-type: none"> données récupérables en ligne par des protocoles standardisés restrictions si nécessaire 	<ul style="list-style-type: none"> formats et standards communs vocabulaires contrôlés 	<ul style="list-style-type: none"> données documentées (pex. avec un fichier README) licence et information de provenance claires
<p>Si le jeu de données ne peut pas être mis à disposition librement, il est recommandé qu'au minimum sa description soit librement accessible en ligne.</p>	<p>Dans un repository qui stocke les données de manière sécurisée</p>	<p>Pour être compatibles et combinables avec d'autres jeux de données</p>	<p>Afin que les données soient interprétées correctement</p>



Principes FAIR

Diffuser, partager ses données



Source : Viet Jeannaud. Déposer ses données de recherche : pourquoi, quoi, quand, où et comment ? Doranum, Mars 2018.

Diffuser, partager ses données

Données communicables

Points de vigilance / conditions

- Traitement de données à caractère personnel (confidentialité) : consentement des personnes, anonymisation avant partage
- Respect des droits de propriété intellectuelle dont les données relèvent
- Avoir accord des détenteurs des droits
- Définir des modalités de réutilisation > licences

Toutes les données

Toutes les données sont susceptibles d'être déposées. Elles peuvent avoir un intérêt pour chaque utilisateur.



Données brutes

Données recueillies, non organisées, non mises en forme. Peuvent convenir à un public susceptible de les traiter.



Données traitées

Données calculées, corrigées, réorganisées... Peuvent convenir à un public susceptible de les analyser.



Données analysées

Données telles qu'apparaissant dans les publications. Peuvent convenir à un public susceptible de les comprendre.



=> Les données concernées par le dépôt dépendent de votre situation !

=> Les données personnelles, relevant de la propriété intellectuelle et industrielle doivent faire l'objet d'une attention particulière.

Données non communicables

- Données confidentielles (informations commerciales, industrielles...)
- Secret défense et sécurité publique

Données de la recherche = *open data*

Avec des restrictions

– respecter des interdictions de diffusion

protection du potentiel scientifique et technique de la nation (PPSN), Zone à régime restrictif (ZRR)

secret défense , secret médical, secret des affaires

– respecter les droits des tiers

respect de la vie privée, données à caractère personnel

droits de propriété intellectuelle : droits d'auteur, des bases de données



Et des obligations

rendre accessibles certaines données

Données géographiques, environnementales (INSPIRE), projets H2020 (obligations contractuelles)

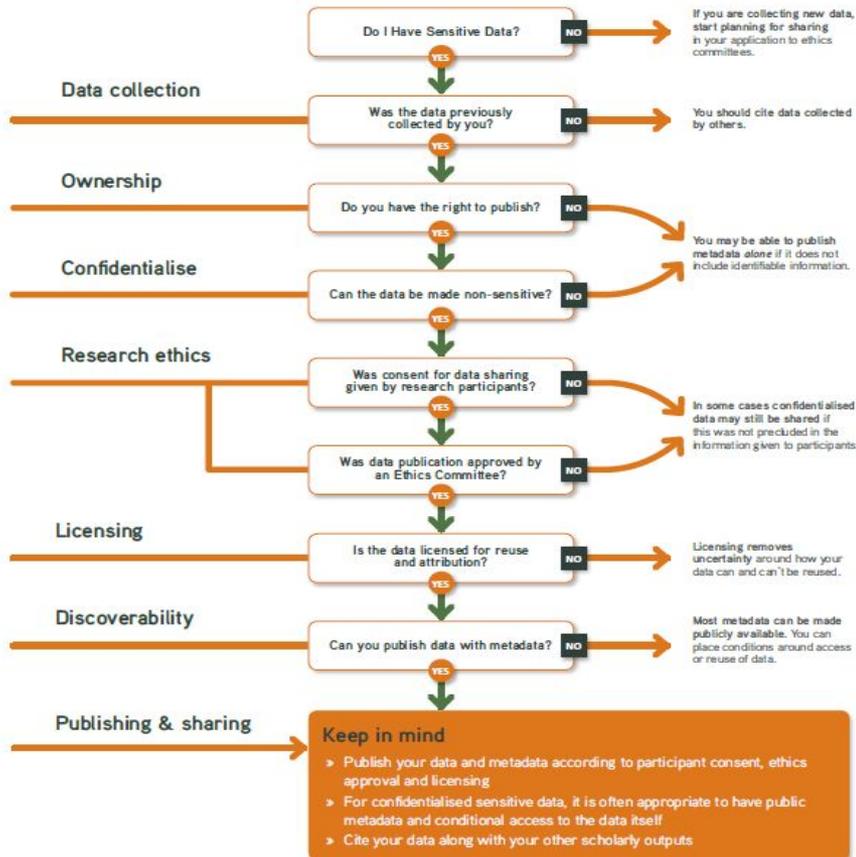
Diffuser, partager ses données

PUBLISHING AND SHARING SENSITIVE DATA

When and how to publish sensitive data as openly and ethically as possible
For more information see: ands.org.au/sensitivedata



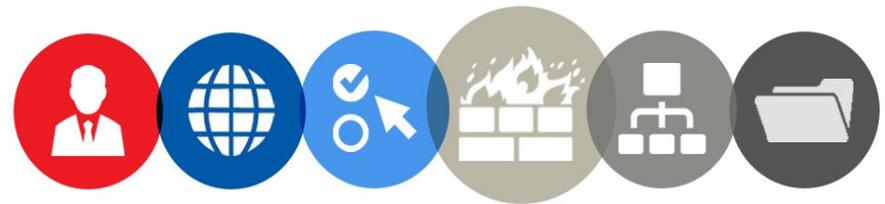
Sensitive data identifies individuals, species, objects or locations, and carries a risk of causing discrimination, harm or unwanted attention



Partager ses données sensibles

- Identifier les données sensibles
- Traiter, anonymiser, supprimer...
- A défaut de partager : signaler ?

[Guide ANDS sur les données sensibles \(australien\)](#) / [Arbre de décision](#)



Où partager ?

- **Différentes modalités**

- Entrepôts de données
- Application Web avec services d'identification, de stockage et de partage de données de recherche structurées (Dataverse).



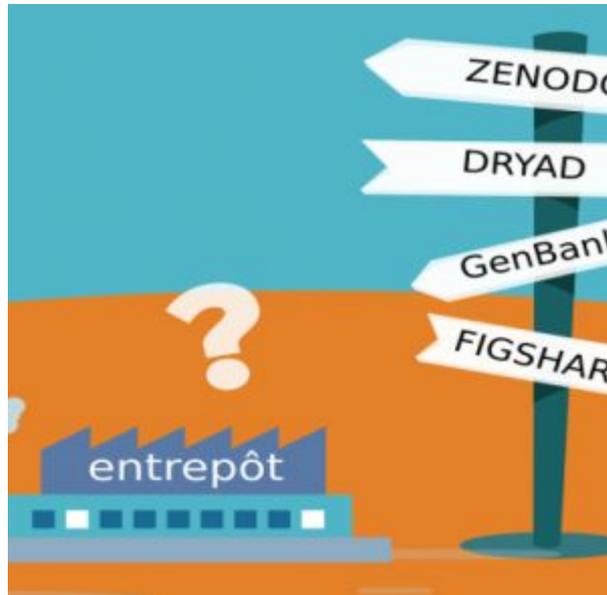
« Un *entrepôt de données* est une archive numérique qui collecte et diffuse des jeux de données et leurs *métadonnées*. Un grand nombre d'entrepôts de données acceptent également des publications et permet de lier les publications afférentes. »

Source : [Openaire](#)

- Espaces de dépôt, de stockage et de diffusion des données de recherche

Diffuser, partager ses données

Types d'entrepôts



Source : [Doranum](#)

Entrepôts propres à un éditeur : [GigaDB](#)
(GigaScience)

Entrepôts d'un champ disciplinaire : [SEANOE](#)
(sciences de la mer), [Réseau Quételet](#) (sciences sociales) (PROGEDO Diffusion), [Nakala](#) (SHS)

Entrepôts d'une institution : [Edinburgh Datashare](#), [Harvard dataverse](#)

Entrepôts généralistes/ multidisciplinaires :
Dataverse Project, [Zenodo](#)

Sur les données :

- Entrepôts acceptant tous les jeux de données, même non liés à une publication : [Zenodo](#), [Figshare](#)
- Entrepôts acceptant publications scientifiques et jeux de données : [Zenodo](#), [DRYAD](#)

Diffuser, partager ses données

Choisir un entrepôt ?

Un entrepôt est recommandé :

- Par votre institution, votre financeur...

Pas d'entrepôt recommandé :

- Identifier des entrepôts via des **répertoires** :
 - ✓ [CatOPIDoR](#)
 - ✓ [RE3data.org](#)
 - ✓ [OAD : Data Repositories](#)

Chercher un **entrepôt certifié**

Trouver un entrepôt de données généraliste

- **Reconnaissance par la communauté scientifique**
 - ✓ Ex : [GenBank](#) (genetic sequence database)



Entrepôts de

confiance : Certification des entrepôts selon certains critères (formats des données, qualité des métadonnées..)

- Certification niveau de base [Data Seal of Approval](#)
- Audit et certification des entrepôts numériques au niveau européen [TrustedDigitalReposistory.eu](#)
- Trouver des entrepôts de confiance avec [Coretrustseal certified data repositories](#)

Diffuser, partager ses données

Critères de choix

Contraintes pour le chercheur/institution	Contraintes imposées par l'entrepôt
Exigences de l'organisation, du financeur, de l'éditeur Communauté d'appartenance Réputation de l'entrepôt Accès contrôlé : restreint ou embargo Localisation du serveur Facilité au dépôt Coût ...	Discipline(s), thématique Types de données acceptés Formats acceptés Volumétrie Coût Licences

1. is a reputable repository available?
2. will it take the data you want to deposit?
3. will it be safe in legal terms?
4. will the repository sustain the data value?
5. will it support analysis and track data usage?

[Whyte \(2015\). Where to keep research data: DCC checklist for evaluating data repositories v.1](#)

+ principes FAIR
+ réutilisation, citation, analyse
+ préservation +/- long terme

Diffuser, partager ses données

Un exemple de jeu de données partagées sur Zenodo :

Titre: Dataset for Interspeech 2018 submission, Singing voice phoneme segmentation by jointly learning syllable and phoneme onset positions

Type de document: Dataset

Accès: Open Access

Auteurs: Rong Gong; Xavier Serra

Description: This dataset contains the materials for training, testing the joint and HSMM models mentioned in the paper "Singing voice phoneme segmentation by jointly learning syllable and phoneme onset positions". This is dataset is made for the paper review process and may not be its final version.

Date de publication: February 27, 2018

identifiant: DOI: 10.5281/zenodo.1185124

Mots clés: jingju singing voice, interspeech 2018, phoneme segmentation, dataset, praat, textgrid, audio

Financement: European Commission: COMPMUSIC - Computational models for the discovery of the world's music (267583)

Licence: Creative Commons Attribution-NonCommercial 4.0

Tableau des fichiers

Name	Size
interspeech2018.zip	641.6 MB
md5:07ca96cfc56f46cd3253330979b5c61d	

Tableau des fichiers dans le dossier

File Name	Size
20171211SongRuoXuan	
_DS_Store	10.2 kB
daxp-Fei_shi_wo-Hua_tian_cuo-dxjky	
_DS_Store	6.1 kB
lyrics.txt	617 Bytes
student01.TextGrid	167.7 kB
student02.TextGrid	173.7 kB
student03.TextGrid	167.5 kB
student04.TextGrid	166.3 kB
teacher.TextGrid	288.8 kB
daxp-Meng_ting_e-Mu_gui_ying_gua_shuai-dxjky	
_DS_Store	6.1 kB
lyrics.txt	243 Bytes
student01.TextGrid	68.4 kB
student02.TextGrid	66.6 kB
student03.TextGrid	66.4 kB

Diffuser, partager ses données

- Les archives ouvertes

- Peuvent développer des services de dépôt et d'hébergement de données de la recherche
- HAL : données de recherche et données supplémentaires sous forme d'**annexes** (à l'origine MediHAL)

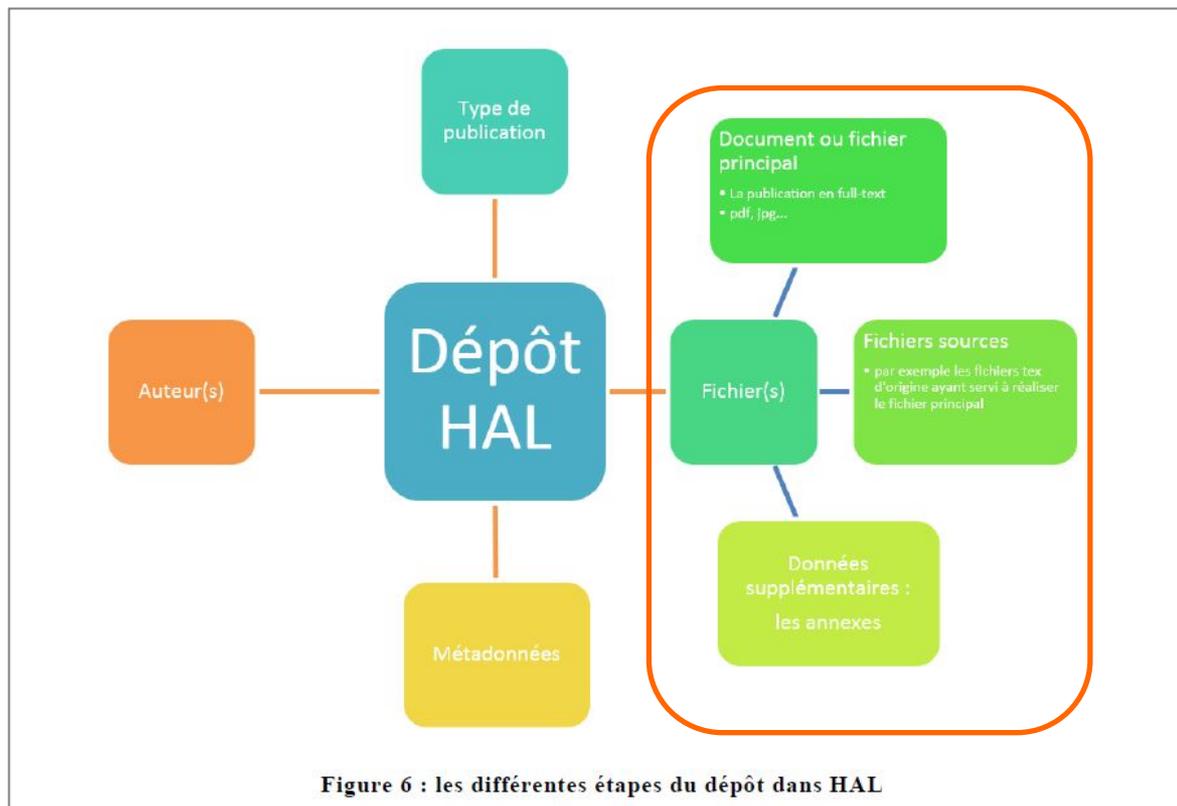


Figure 6 : les différentes étapes du dépôt dans HAL

Source : Pain, M.
Les données de la
recherche et leurs
entrepôts, de la
documentation à la
réutilisation : étude de cas
pour l'archive HAL, 2016

Diffuser, partager ses données

HAL et les données de recherche

Hal propose un type de document « Données de la recherche ».

Dans ce cas, le fichier principal déposé = la donnée ou le jeu de données

Données de recherche	41122	4	98	41224
Image	38475		4	38479
Vidéo	2358			2358
Carte	178			178
Logiciel	22	4	94	120
Son	89			89

(18/04/2019)

Données de recherche (56162)

(24/11/2021)

Image (51663) Vidéo (3264) Son (225) Carte (315) Logiciel (695)



Des identifiants pérennes pour les données de recherche

- A quoi cela sert-il ?
 - À identifier les auteurs, les contributeurs et à attribuer les données de manière unique à chaque fois
 - De **manière pérenne** quel que soit le trajet de la ressource ou de la personne
 - Par des identifiants pérennes, non-proprétaires, ouverts et interopérables
 - Rend les données **accessibles et citables**, permet de les lier aux publications, augmente la **visibilité**.
- Qu'est-ce qu'un identifiant pérenne ?
 - C'est un code unique, invariable, qui est associé de manière permanente à une personne (> chercheur) ou à un objet (> donnée)

Diffuser, partager ses données

• Les différents types d'identifiants

- **Le DOI** : Digital Object Identifier, identifiant numérique d'objet
 - = chaîne de caractères univoque et pérenne pour identifier et retrouver les ressources
 - Association avec des métadonnées
 - Syntaxe du DOI



Publication date:
December 1, 2017

DOI:
DOI 10.5281/zenodo.1184901

Keyword(s):
R GeoServer API web-service

Communities:
Zenodo

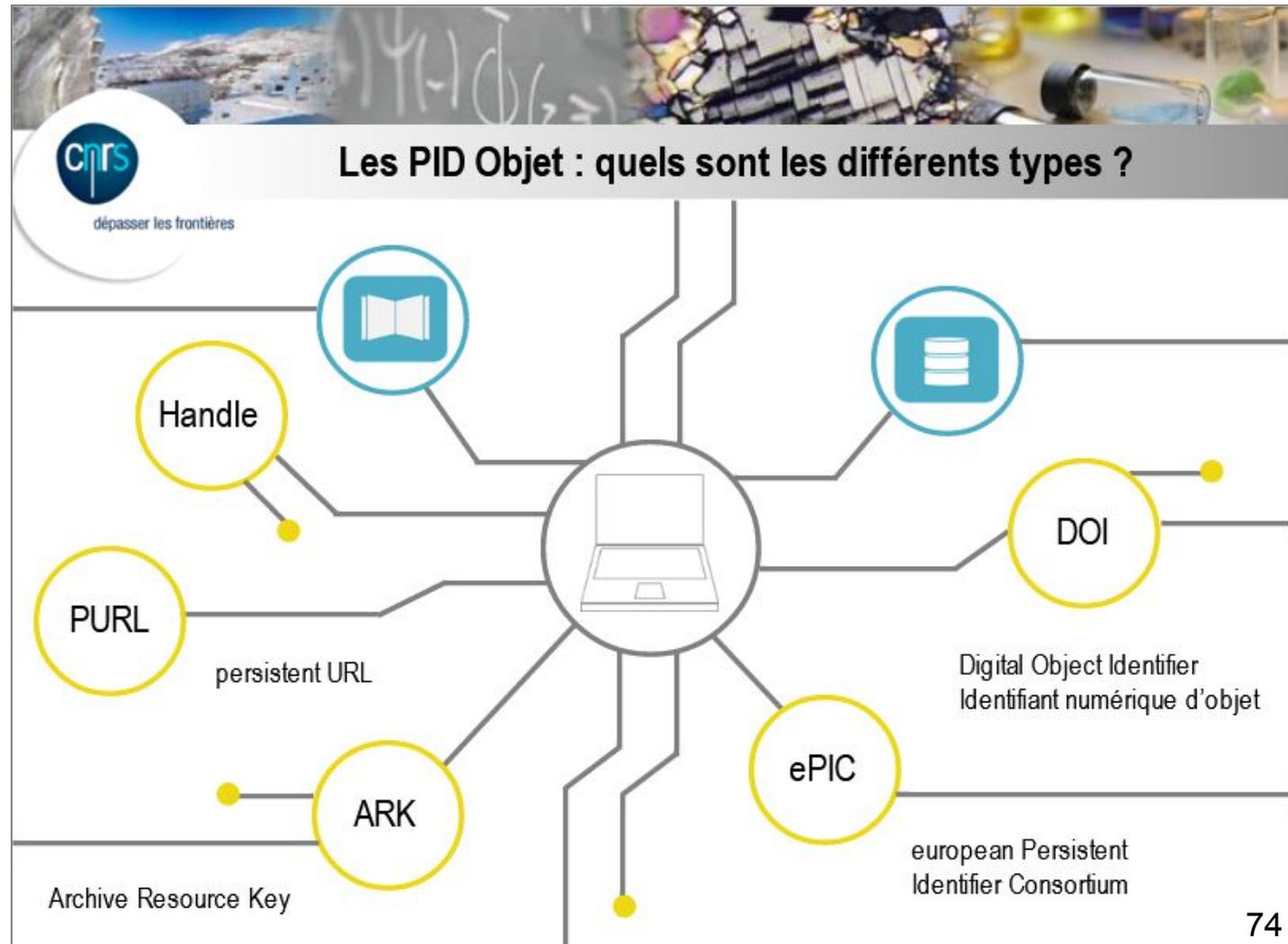
License (for files):
MIT License

PID OPIDOR : attribuer un identifiant pérenne aux données de la recherche



Diffuser, partager ses données

- **Autres systèmes d'identifiants pérennes**



Diffuser, partager ses données

- **Comment attribuer un identifiant pérenne ? Le cas du DOI**
 - Abonnement à une agence d'enregistrement de DOI
 - Crossref, DataCite, mEDRA
 - Enregistrer un DOI pour un jeu de données via [DataCite](#)
 - [L'INIST / CNRS](#) est membre pour la France de DataCite et a de ce fait le statut d'agence DOI > adhésion des établissements à DataCite via L'INIST pour l'attribution de DOI.
 - Service Metadata Store : interface pour enregistrer les métadonnées et le DOI d'une ressource



Diffuser, partager ses données

La publication des données

• Deux possibilités :

- « Supplementary data » : publier les données dans des fichiers supplémentaires associés à l'article

Supplementary Data



2013 Supplementary Data

January 2013, Volume 28, Number 1



Paleobiology and paleoecology of the early Turonian (Late Cretaceous) ammonite *Pseudaspidoceras flexuosum*

Christina Ifrim

[Supplementary Data](#)

[Abstract](#)

9

February 2013, Volume 28, Number 2



Preservational and morphological variability of assemblages of agglutinated eukaryotes in Cryogenian cap carbonates of northern Namibia

Lilly A. Dalton, Tanja Bosak, Francis A. Macdonald, Daniel J.G. Lahr, and Sara B. Pruss

[Supplementary Data](#)

[Abstract](#)

67

Vinegaroon (Arachnida: Thelyphonida: Thelyphonidae) trackway production and morphology: Implications for media and moisture control on trackway morphology and a proposal for a novel system of interpreting arthropod trace fossils

Joshua D. Schmerge, David J. Riese, and Stephen T. Hasiotis

Supplementary Data Key

Windows Media Video (.wmv files) QuickTime Movie files (.mov)

Supplementary Data 1	Supplementary
Supplementary Data 2	Supplementary
Supplementary Data 3	Supplementary
Supplementary Data 4	Supplementary
Supplementary Data 5	Supplementary
Supplementary Data 6	Supplementary
Supplementary Data 7	Supplementary
Supplementary Data 8	Supplementary

Diffuser, partager ses données

• Deuxième possibilité :

- « Data papers » : publier un data paper dans une revue scientifique (data journal ou revue scientifique traditionnelle > stratégie des éditeurs)

Le *data paper* est une publication qui décrit un jeu de données scientifiques brutes (*data, dataset*), notamment à l'aide d'informations précises, appelées métadonnées (*metadata*)

Source : [CoopIST / cirad](#)

Accès aux données :

- lien vers l'entrepôt de données
- fichiers annexés



ELSEVIER

Data in Brief

Volume 1, December 2014, Pages 19-21
open access

Data Article

Data from a proteomic analysis of colonic fibroblast secretomes

Sun-Xia Chen, Xiao-En Xu, Xiao-Qing Wang, Shu-Jian Cui, Lei-Lei Xu, Ying-Hua Qian Zhang, Jie Qiao, Peng-Yuan Yang, Feng Liu

Show more

<https://doi.org/10.1016/j.dib.2014.08.003>

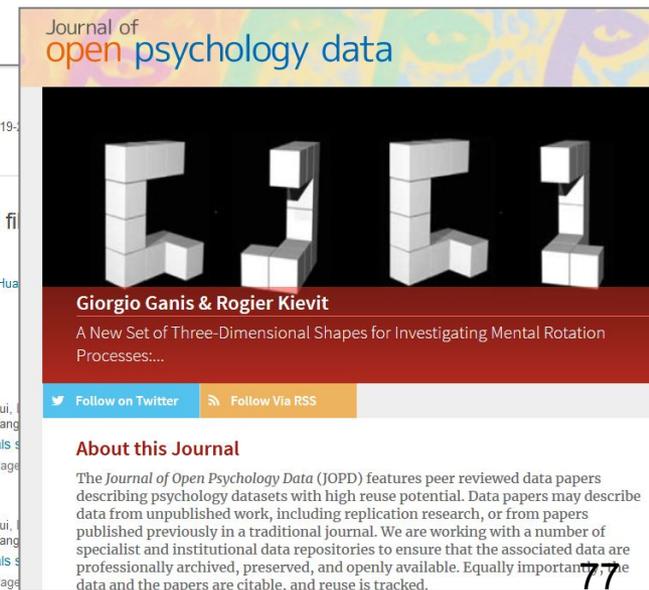
Under a Creative Commons license

Refers to

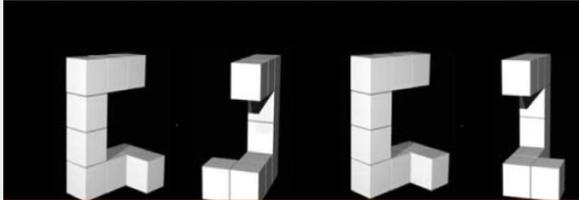
Sun-Xia Chen, Xiao-En Xu, Xiao-Qing Wang, Shu-Jian Cui, Qian Zhang, Hai-Bo Yan, Qian Zhang, Jie Qiao, Peng-Yuan Yang
Identification of colonic fibroblast secretomes reveals s
Journal of Proteomics, Volume 110, 14 October 2014, Page
Download PDF

Referred to by

Sun-Xia Chen, Xiao-En Xu, Xiao-Qing Wang, Shu-Jian Cui, Qian Zhang, Hai-Bo Yan, Qian Zhang, Jie Qiao, Peng-Yuan Yang
Identification of colonic fibroblast secretomes reveals s
Journal of Proteomics, Volume 110, 14 October 2014, Page
Download PDF



Journal of
open psychology data



Giorgio Ganis & Rogier Kievit

A New Set of Three-Dimensional Shapes for Investigating Mental Rotation Processes:...

Follow on Twitter Follow Via RSS

About this Journal

The *Journal of Open Psychology Data* (JOPD) features peer reviewed data papers describing psychology datasets with high reuse potential. Data papers may describe data from unpublished work, including replication research, or from papers published previously in a traditional journal. We are working with a number of specialist and institutional data repositories to ensure that the associated data are professionally archived, preserved, and openly available. Equally important, the data and the papers are citable, and reuse is tracked.

77

Data paper / data journal

- moyen de valoriser ses données comme publications citables
- évaluation par les pairs
- Complémentaire au dépôt des données
- forte incitation lors de la publication de l'article de recherche dans certaines disciplines

Nature – Instructions aux auteurs



Caractéristiques :

- Des revues sur le modèle auteur-payeur
- Détenues par des éditeurs monopolistiques
- Plutôt orientés STM
- En SHS : Research Data Journal for the Humanities and Social Sciences (Brill)
- Propriété des données
- Des tentatives à suivre : revue RFSIC

Diffuser, partager ses données

- Vous avez fait un PGD ? Vous avez bien avancé votre data paper !

Informations requises dans un Data Paper	Informations du PGD
Disponibilité des jeux de données	Partage / modalités prévues
Conflits d'intérêt	Description / origine des données
Etendue géographique et chronologique des données	Description / type de données
Format	Description / format
Licence	Partage / droit et licences
Remerciement / attribution des crédits	Métadonnées / responsables
Qualité	Description / origine, contrôle qualité
Provenance	Description / origine
Projet	Description / origine
Réutilisation	Partage

DONNÉES DE LA RECHERCHE APPRENTISSAGE NUMÉRIQUE

DES RESSOURCES POUR ACCOMPAGNER LA COMMUNAUTÉ SCIENTIFIQUE DANS LA GESTION ET LE PARTAGE DE LEURS DONNÉES

THÉMATIQUES ABORDÉES

PARCOUREZ LES RESSOURCES CLASSÉES PAR THÉMATIQUES

<https://dorandum.fr/>

ENJEUX & BÉNÉFICES

Pourquoi partager les données ?
Qu'est-ce que l'Open Science ?



ASPECTS JURIDIQUES, ÉTHIQUES, INTÉGRITÉ SCIENTIFIQUE

Que puis-je partager, réutiliser ?
Quelles pratiques devrais-je
respecter ?



PLAN DE GESTION DE DONNÉES

Pourquoi et comment rédiger un
plan de gestion des données ?



MÉTADONNÉES

Comment décrire les données ?



IDENTIFIANTS PÉRENNES

Comment associer durablement
des données à son auteur ?



DÉPÔT & ENTREPÔTS

Comment et où déposer mes
données ?



STOCKAGE & ARCHIVAGE

Quelles données conserver à long
terme et comment ?



DATA PAPERS & DATA JOURNALS

Comment publier mes données
comme un article scientifique ?



ACCÈS & VISUALISATION

Où et comment extraire et
visualiser les données qui
m'intéressent ?



Quelques ressources

- OPIDoR, <https://opidor.fr/>
- Doranum, Données de la recherche - Apprentissage numérique, <https://doranum.fr/>
- Questions juridiques liées aux données de recherche : interview de Lionel Maurel, 2019, <https://doranum.fr/aspects-juridiques-ethiques/questions-juridiques-liees-aux-donnees-de-la-recherche/>
- Loi pour une République numérique, https://www.legifrance.gouv.fr/affichTexte.do;jsessionid=D646DA1665089C6A554FD6C1CF0983C2.tplqfr42s_2?cidTexte=JORFTEXT000033202746&categorieLien=id
- ENPC (Ecole des Ponts Paristech), Données de la recherche 2, Contexte juridique, <https://espacechercheurs.enpc.fr/donnees-recherche-contexte-juridique>
- CoSO, Guide pratique pour une harmonisation internationale de la gestion des données de recherche, <https://www.ouvrirlascience.fr/guide-pratique-pour-une-harmonisation-internationale-de-la-gestion-des-donnees-de-recherche/> + les autres [documents du Collège Données](#).
- A. Serres, M.-L. Malingre, M. Mignon, C. Pierre, D. Collet. Données de la recherche en SHS. Pratiques, représentations et attentes des chercheurs : une enquête à l'Université Rennes 2, 2017, <https://hal.archives-ouvertes.fr/hal-01635186v2>
- CoSO, pour une politique des données de la recherche, guide stratégique, 2019, <https://www.ouvrirlascience.fr/pour-une-politique-des-donnees-de-la-recherche-guide-strategique-a-lusage-des-etablissements/>