

Gérer les données de recherche : planifier, organiser, diffuser et archiver

Présentation au Pôle image, 07/03/2017



Magalie MOYSAN, Bureau des archives

Nathalie REYMONET, Direction d'Appui à la Recherche

Les données de recherche : définition

« Les données, ou unités d'information, qui sont créées au cours d'une recherche, subventionnée ou non, et qui sont organisées ou formatées de telle sorte qu'elles soient communicables, interprétables et adaptées à un traitement souvent informatisé ».

Université de Bristol

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

Organisation de Coopération et de Développement Économiques (OCDE)

« Les données de la recherche sont l'ensemble des informations et matériaux produits et reçus par des équipes de recherche et des chercheurs. Elles sont collectées et documentées à des fins de recherche scientifique. A ce titre, elles constituent une partie des archives de la recherche. »

Association des archivistes français. Section « Aurore »

*« 1) the data, including associated metadata (i.e. the metadata describing the research data deposited), needed to validate the results presented in scientific publications ;
2) other data (i.e. curated data not directly attributable to a publication, or raw data), including associated metadata.»*

Commission européenne pour Horizon 2020

« Les données de la recherche sont un ensemble d'informations factuelles enregistrées sur des supports, produites ou collectées, selon divers procédés au cours d'un processus de recherche »

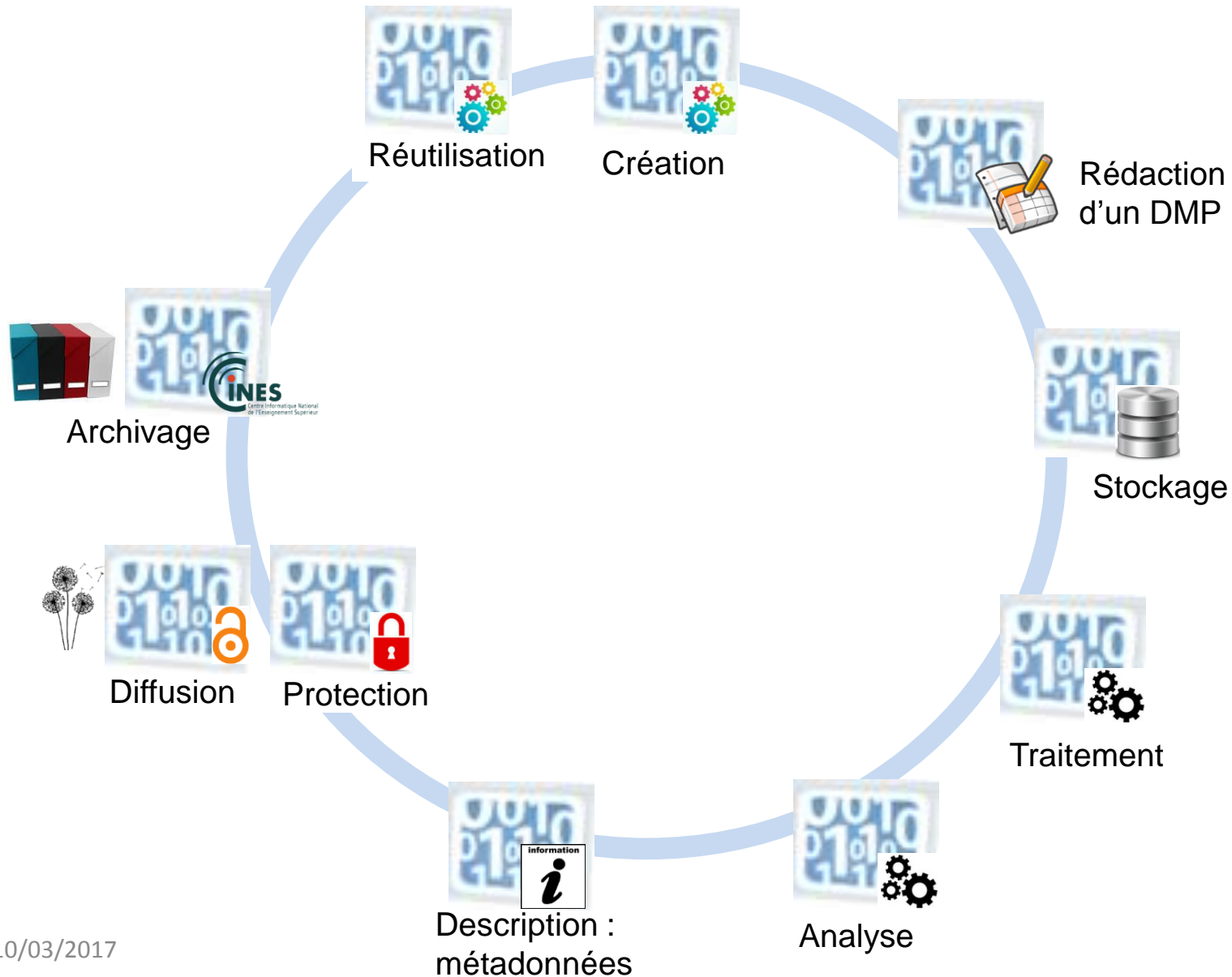
Définition du groupe de travail P5/P7

Le jeu de données

Selon le projet, un jeu de données peut recouvrir des réalités différentes. Un ou plusieurs jeu(x) de données peuvent être lié(s) au projet de recherche, et désigner :

- i) un lot techniquement homogène,*
- ii) un lot intellectuellement cohérent même si celui-ci est composé de lots techniquement hétérogènes.*

Le cycle de vie des données



Gestion, partage et archivage des données

Gestion des
données
Data management



Mise en place d'architectures, de réglementations, de pratiques et de procédures pour gérer correctement les besoins de la recherche sur tout au long du cycle de vie des données

Partage des
données
Open data



Diffusion de manière structurée selon une méthode et une *licence ouverte* des données pour leur libre accès et leur réutilisation par tous, sans restriction technique, juridique ou financière.

Archivage
*Long term
preservation*



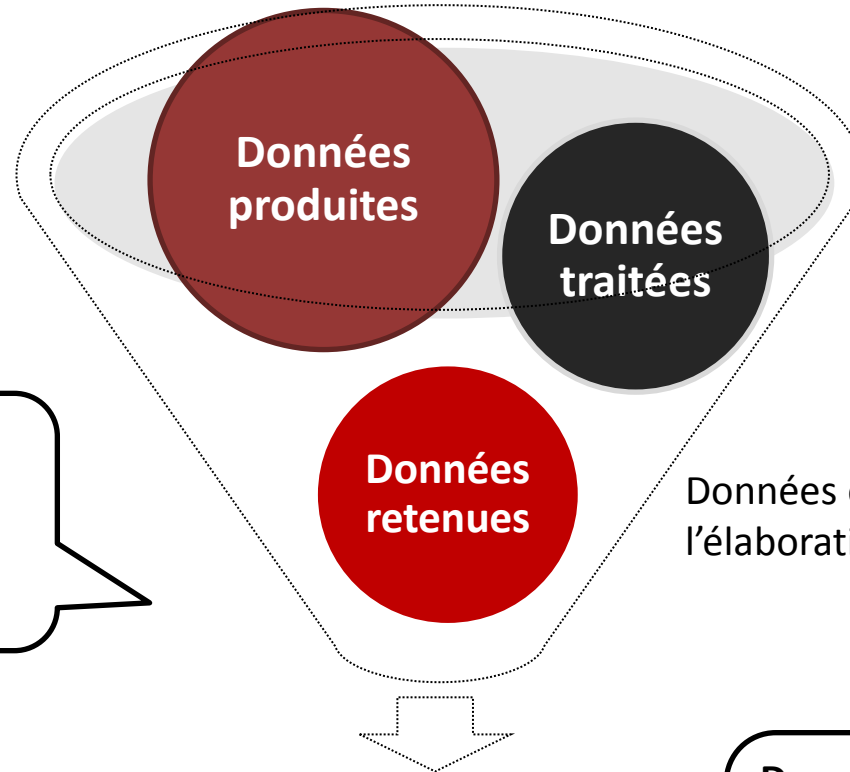
Conservation à long terme des données qui présentent un intérêt scientifique, juridique ou historique.

Améliorer la **visibilité** des résultats

Données primaires
produites ou collectées
au cours de la recherche

Données analysées,
agrégées, compilées,
transformées

INVISIBLES



Données exploitées pour
l'élaboration des résultats

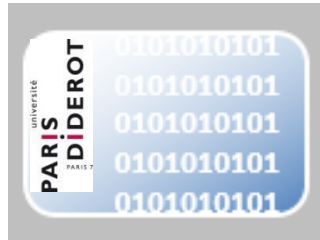
Perte d'information
importante au cours
du processus de
publication

VISIBLES

Données publiées

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

Enjeux de la gestion des données

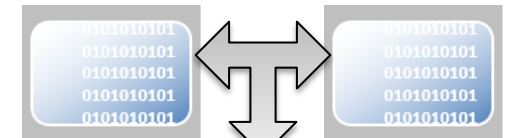


anticiper les besoins
de gestion, protection
et partage

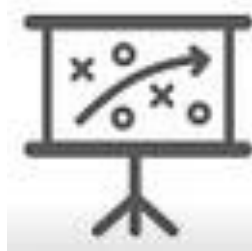


Nouvelles sources
pour la recherche

DATA



Fouille de données
/ data mining

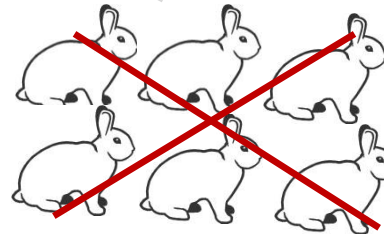


Valoriser la production
scientifique de
l'établissement

Fiabiliser les résultats
de recherche



Economies dans les
coûts de production



Conditions de
production non
reproductibles



Mieux diffuser
les résultats

Un contexte d'ouverture



Dans le cadre d'Horizon 2020, l'***Open research data pilot***, demande au chercheur :

- 1) produire *Data Management Plan*
- 2) déposer leurs données et métadonnées dans un entrepôt de données.



La ***Loi pour une République numérique*** stipule que lorsqu'une recherche est financée pour moitié au moins sur fonds publics:

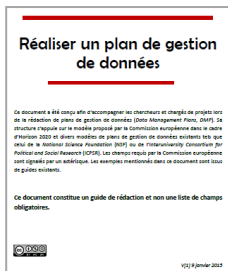
- 1) Le chercheur doit publier ses articles en *Open Access* dans un délai de 12 mois pour les SHS
- 2) La réutilisation des données est libre si ne sont pas protégées par un droit spécifique ou une réglementation particulière

Planifier la gestion des données : *le data management plan*

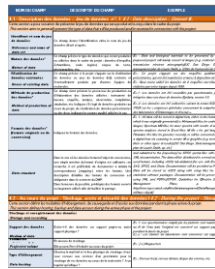


Document formel précisant la manière dont les données seront produites, traitées, décrites, partagées ou protégées et conservées au cours et à l'issue du projet.

- **décrire le cycle de vie** des données produites ou collectées au cours du projet de recherche.
- **anticiper les questions de gestion** qui surviennent au cours d'une recherche et les conditions d'une diffusion et d'une conservation futures des données.



[Télécharger le guide](#)



[Télécharger le template](#)



[S'inscrire](#)



[S'inscrire](#)

Données en appui des publications ou OK pour diffusion.

Autres données produites au cours du projet. Pas de diffusion.

Dépôt en archives ou entrepôts ouverts

Archivage auprès d'un organisme agréé

Structure d'un *data management plan*

- **Section 1** : Informations relatives au projet
- **Section 2** : Responsabilité des données
- **Section 3** : Ressources nécessaires

- **Section 4** : Jeux de données
 - 4.1 Description du jeu de données
 - 4.2 Stockage, accès et sécurité des données – *au cours du projet*
 - 4.3 Métadonnées : documentation et organisation des données
 - 4.4 Dissémination des jeux de données – *à l'issue du projet*

- **Section 5** : **Sélection et archivage** – que les données soient diffusées ou non

Valable pour l'ensemble du projet

Propre à chaque jeu de données sauf si valable à l'échelle du projet

Valable pour l'ensemble du projet

Organiser ses documents

➤ Bien nommer les document

Service_sujet_typologie_annee_mois_jour_version

BA_formation_archivage_diaporama_2017_03_03_V2F

➤ Etablir un plan de classement

Nom	Modifié le	Type	Taille
ASSOCIATIONS	25/07/2014 11:52	Dossier de fichiers	
BROCHURES	12/02/2015 17:20	Dossier de fichiers	
DEMEMAGEMENT	10/02/2015 14:32	Dossier de fichiers	
DEPARTEMENTS	19/08/2014 10:45	Dossier de fichiers	
DONNEES DE RECHERCHE	03/11/2015 13:48	Dossier de fichiers	
ECOLE INGENIEUR	04/04/2013 17:51	Dossier de fichiers	
ECOLES DOCTORALES	25/07/2014 11:30	Dossier de fichiers	
ENSEIGNANTS CHERCHEURS	20/07/2015 10:40	Dossier de fichiers	
FONDS CONSERVES AU RECTORAT	04/05/2015 09:42	Dossier de fichiers	
FORMATIONS	14/10/2015 10:17	Dossier de fichiers	
INSTITUT JACQUES MONOD	21/05/2015 10:28	Dossier de fichiers	
IUH	04/05/2015 09:20	Dossier de fichiers	
IUT	15/07/2014 19:02	Dossier de fichiers	
LABORATOIRES	10/11/2015 09:44	Dossier de fichiers	
PROCEDURES QUALITE	07/10/2015 10:21	Dossier de fichiers	
RECOLEMENT 2015	30/10/2015 12:24	Dossier de fichiers	
SCRIPT	07/09/2015 13:33	Dossier de fichiers	
SELECTION	20/10/2015 10:54	Dossier de fichiers	
UFR	10/09/2015 14:52	Dossier de fichiers	

Diffuser ses données

re3data.org
REGISTRY OF RESEARCH DATA REPOSITORIES

Entrepôts
disciplinaires



Financeurs



Établissements de
recherche



Éditeurs



Entrepôts de
communautés



Privilégier les entrepôts institutionnels ou disciplinaires dont le modèle économique est connu et présentant des garanties en matière d'indexation, de signalement et de conservation des contenus.

Diffuser ses données : publier un *data paper* (1)

- article scientifique, décrit des données
- exposer des données, en répondant à plusieurs besoins :
 - ✓ répondre aux financeurs & communauté (visibilité)
 - ✓ donner un accès aux données et les rendre intelligibles
 - ✓ fournir une référence citable (*peer-reviewed*)
 - ✓ permettre la génération de citations par des services tel que le *Data Citation Index*
 - ✓ reconnaître le travail réalisé par l'équipe de recherche qui a produit les données décrites

Diffuser ses données : publier un *data paper* (2)

- Instructions aux auteurs
 - thèmes / langage(s) / longueur / style de références
 - liste d'entrepôts recommandés
- Auteurs, titre, résumé, mots-clé, références
- Texte :
 - Contexte de la recherche
 - Méthode : échantillon, matériel, procédures, contrôle qualité, questions éthiques
 - Description du/des jeu(x) de données
 - Bibliographie
- Article : texte + un lien vers le jeu de données décrit + matériaux supplémentaires (illustrations, tableaux, vidéos, sons, etc.)

Renseigner un DMP permet d'extraire plus facilement les info nécessaires à la rédaction d'un *data paper* !

Archiver ses données

Quel intérêt?



Conserver des preuves

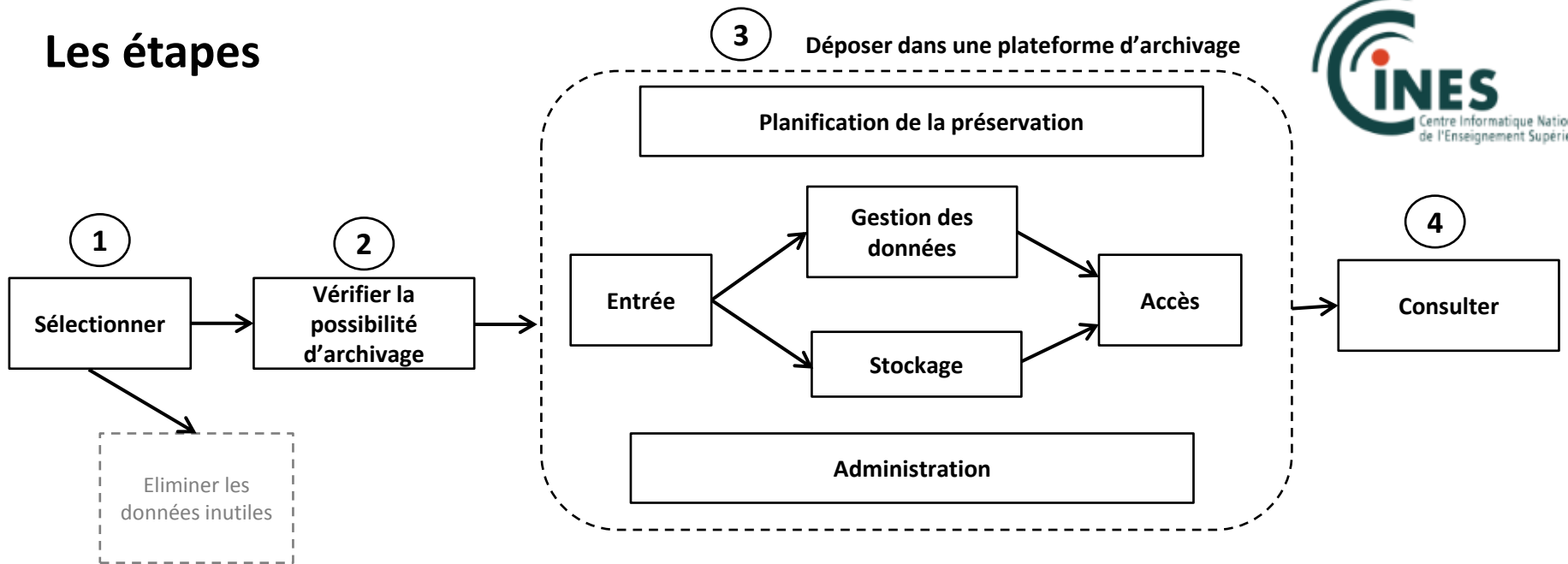


Assurer la continuité des activités de recherche



Conserver le patrimoine scientifique

Les étapes



Merci pour votre attention

Magalie Moysan – magalie.moysan@univ-paris-diderot.fr

Nathalie Reymonet – Nathalie.reymonet@univ-paris-diderot.fr