LARHRA
UMR 5190
LABORATOIRE DE RECHERCHE HISTORIQUE RHÔNE-ALPES

Gérer ses données de recherche

Marie Puren

Université du Mans – 12 février 2020

CC BY

Introduction

« Une gestion des données efficace assure des bénéfices **tout au long et au-delà** de la durée de vie du projet de recherche [...] Pour le chercheur, la perception des données comme un instrument de recherche et un nouveau savoir, peut être à la source de **changements profonds**. Des données bien gérées mènent à des données de recherche de **meilleure qualité**, accroît leur **visibilité** mais aussi les bénéfices connexes liés à des **taux de citation améliorés**. »

How to develop RDM Services, Digital Curation Center

Les données de la recherche

- Encore difficile de donner une définition « canonique » des données de la recherche
- Pas de consensus
  - Peut varier en fonction de la discipline
  - Peut varier en fonction du ou des financeur(s)

Les données de la recherche

« 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. » *Principes et lignes directrices de l'OCDE pour l'accès aux données de la recherche financée sur fonds publics* (2006)

Les données de la recherche

- En fonction des disciplines ou des thèmes de recherche
  - Linguiste : des écrits, des discours, des enregistrements
  - Géologue : coupes et observations de terrain consignées sur un carnet, des résultats de carottage, des analyses d'échantillons, des données sismographiques

- Au sein d'une même discipline:
  - Médiéviste : sources archivistiques, épigraphiques, iconographiques, littéraires, voire même archéologiques
  - Contemporanéiste : transcription de documents d'archives, ou des entretiens enregistrés et/ou filmés.

- Pour résumer, données de recherche :
  - Pas seulement des données quantitatives
  - Pas seulement des séries homogènes, faciles à manipuler, échanger, compiler

Les données de la recherche

- Définition restrictive de l'OCDE car les chercheurs produisent beaucoup plus de données que celles qui sont strictement destinées à valider des résultats de recherche

- Des prises de note non exploitées, des résultats d'expérience non utilisés => utilisables par d'autres chercheurs

- « Enregistrement factuel » dans la définition proposée par l'OCDE : données en partie traitées ou dérivées car organisées, décrites, sélectionnées?

Les données de la recherche

Données : « Elément (fait, chiffre, etc.) qui est une information de base sur laquelle peuvent s'appuyer des décisions, des raisonnements, des recherches et qui est traité par l'humain avec ou sans l'aide de l'informatique. » Office québécois de la langue française, 2002.

Les données de la recherche

- Frontière entre données et publications parfois difficile à distinguer
- Blogs, tweets… Données de recherche?
  - Pas des matériaux bruts mais dépourvus de tout apparat critique et non validés par les pairs (contrairement aux publications)
- Données publiées dans des articles, monographies sous forme de résultats
- Place des carnets de laboratoire, des carnets de recherche, des projets, des rapports

=> Question pas encore tranchée!

Les données de la recherche

« Ces données sont des faits, des observations, ou des expériences sur lesquels on base un argument, une théorie ou un texte. Les données peuvent être numériques, descriptives ou visuelles. Les données peuvent être brutes ou analysées, expérimentales ou observationnelles. Les données incluent : carnets de laboratoire (ou cahiers de laboratoire) ; carnets de fouilles ; données de recherche primaires (incluant les données de recherche sur copie papier ou dans une forme lisible par l'ordinateur) ; des questionnaires ; cassettes audio ; cassettes vidéo ; modèles ; photographies ; réponses à des tests. »

"What is research data ?", Australian National Data Service

Les données de la recherche

Selon Thierry Fournier (pilote de la commission recherche de l'Association des directeurs et des personnels de direction des bibliothèques universitaires (ADBU),

ces données ne sont pas « nécessaires à la validation des résultats de la recherche », et se sont pas des données de la recherche

« Les données de la recherche, définition et enjeux », Arabesques, 2015

Les données de la recherche

« Les données de la recherche sont créées comme des résultats directs de “faire de la recherche”. En ce sens, cela exclut les supports destinés à l’enseignement et les documents administratifs (comme les fiches de poste, les mails ou les rapports financiers). Les données de la recherche peuvent être disponibles dans une infinie variété de formats. »

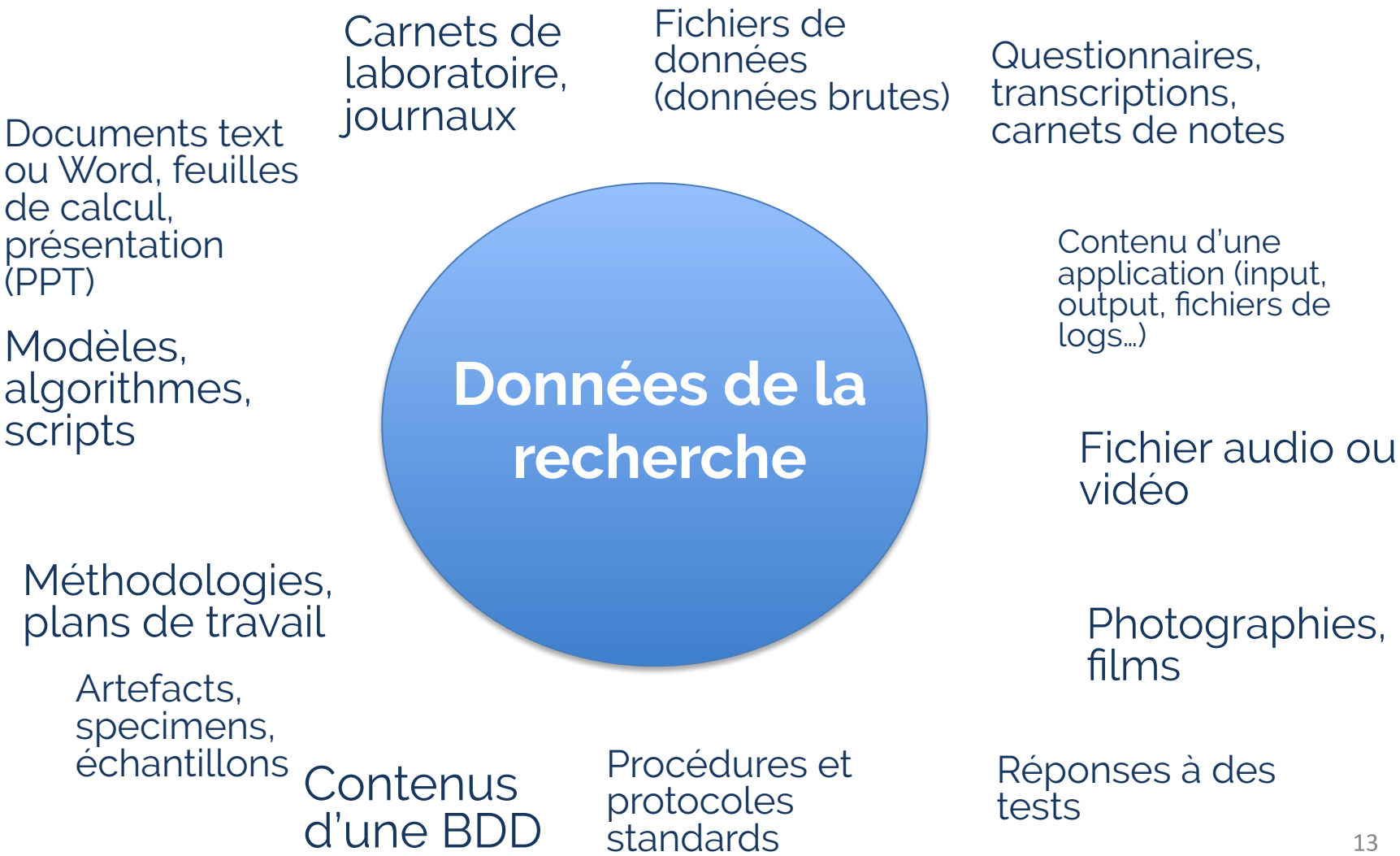
University of Bristol, “A brief guide to research data”

Les données de la recherche

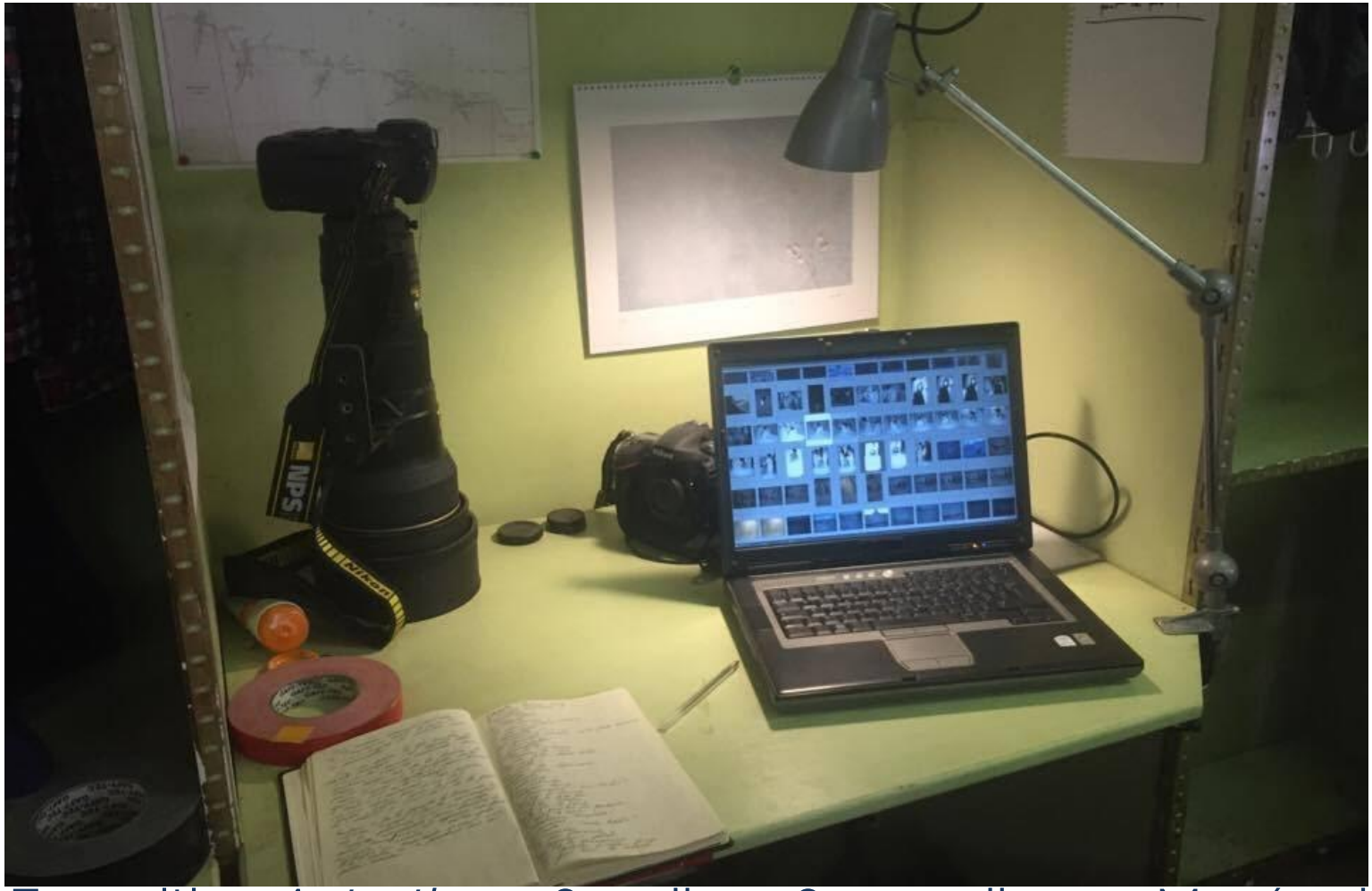
Les données de la recherche peuvent être :

- Observationnelles
- Expérimentales
- Générées au moyen de simulation
- Dérivées ou compilées (comme avec la fouille de textes et de données ou *text and data mining*)

Les données de la recherche



Les données de la recherche
Lors d'une expédition polaire…



Exposition *Antartica*, 26 avril 2016-17 avril 2017, Musée des confluences, Lyon

Les données de la recherche

- Jeu de données ou dataset : “agrégation, sous une forme lisible, de données brutes ou dérivées présentant une certaine ‘unité’, rassemblées pour former un ensemble cohérent”. Rémi Gaillard, De l’Open data à l’Open research data : quelle(s) politique(s) pour les données de recherche?
  - Série de données compilées dans plusieurs documents différents, ou série de données compilées dans un document unique - par exemple, dans un fichier CSV
  - en fonction de la discipline, mais aussi des projets et des objectifs de la recherche, “l’échelle à laquelle ces données assemblées acquièrent leur unité pour former un ‘jeu’”” va beaucoup varier. (*Ibid.*)

Une nouvelle demande

- Demande : données de recherche = accessibles

- Plus conservées dans des données personnels ou au fond d'un tiroir

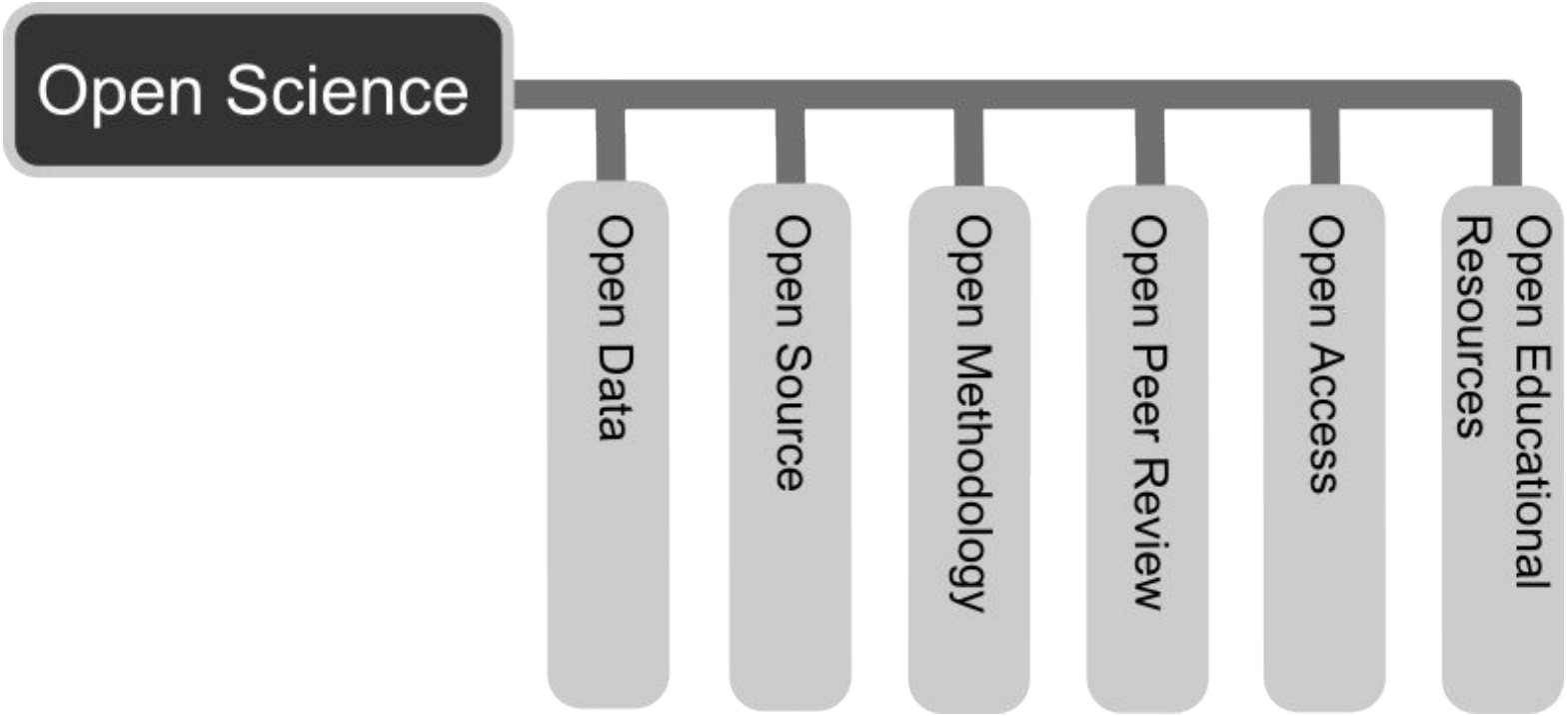
Un nouveau modèle d'ouverture pour les données de recherche

- La Science Ouverte : « mouvement destiné à promouvoir plus de partage et de transparence » dans la science

- Promeut un nouveau modèle d'ouverture pour les données de la recherche.

- Principaux aspects:
  - La disponibilité et l'accès
  - La réutilisation et la redistribution
  - La participation de tous

Un nouveau modèle d'ouverture pour les données de recherche



Par Andreas E. Neuhold, travail personnel- basé sur "The taxonomy tree", FOSTER (Facilitate Open Science Training for European Research)

Pour plus d'infos sur la Science ouverte: Michael Nielsen, *Reinventing Discovery: The New Era of Networked Science,* Princeton University Press, 2011.

Un nouveau modèle d'ouverture pour les données de recherche

- **2003** : Déclaration fondatrice de Berlin en faveur de l'Open Access : les données de la recherche sont assimilées aux publications.

- **2013** : Open Research Data Pilot : mise en place d'une politique européenne en matière de libre-accès aux données

Un nouveau modèle d'ouverture pour les données de recherche

- **2013** : Open Research Data Pilot : mise en place d'une politique européenne en matière de libre-accès aux données

- **2019** : Appel générique 2019 de l'ANR : obligation de fournir un plan de gestion des données

Partager ? Oui, mais…

« Partager ses données, c'est un peu comme aller chez le dentiste. Nous pouvons tous être d'accord que c'est une bonne chose à faire et que c'est lié, de manière intrinsèque, à de bonnes pratiques scientifiques. En réalité, toutefois, les chercheurs tendent à envisager le partage de leurs données avec un mélange de peur, de dédain et d'effroi. »

Felicia LeClere, "Too Many Researchers Are Reluctant to Share Their Data", *The Chronicle of Higher Education*, 2010.

Partager ? Oui, mais…

« Déposer ou ne pas déposer, telle est la question. »
Roche DG, Lanfear R, Binning SA, Haff TM, Schwanz LE, et al. (2014) "Troubleshooting Public Data Archiving: Suggestions to Increase Participation", *PLoS Biol* 12(1): e1001779. doi:10.1371/journal.pbio.1001779

Encore des réticences – notamment en SHS

- Hélène Prost, Joachim Schöpfel. *Les données de la recherche en SHS. Une enquête à l'Université de Lille 3.* : Rapport final. [Rapport de recherche] Lille 3. 2015. ⟨hal-01198379⟩

- Alexandre Serres, Marie-Laure Malingre, Morgane Mignon, Cécile Pierre, Didier Collet. *Données de la recherche en SHS. Pratiques, représentations et attentes des chercheurs : une enquête à l'Université Rennes 2 : Rapport* ; Annexe 1 : Résultats de l'enquête statistique ; Annexe 2 : Croisements statistiques ; Annexe 3 : Extraits des entretiens ; Synthèse des résultats. . [Rapport de recherche] Université Rennes 2. 2017, 159 p., 47 p., 114 p., 26 p., 23 p. ⟨hal-01635186v1⟩

Des bénéfices directs pour les chercheurs

Erin C. McKiernan, Philip E. Bourne, C. Titus Brown, Stuart Buck, Amye Kenall, Jennifer Lin, Damon McDougall, et al, “How open science helps researchers succeed.” *ELife* 5. eLife Sciences Publications Limited, 2016, 372–82, doi: https://doi.org/10.7554/eLife.16800.

Des bénéfices directs pour les chercheurs

1. Satisfaire les exigences de leurs financeurs et celles de certaines revues (qui demandent que les données analysées dans les articles publiés soient archivées)

2. Augmenter l'impact de leur recherche et leur visibilité
   - en étant crédité pour leurs résultats,
   - en améliorant leur taux de citation,
   - en étant mieux connu (et reconnu) par leurs pairs.

3. Gagner du temps en étant plus efficace. Plus besoin de répondre aux requêtes de vos collègues qui veulent avoir accès à vos données.

Des bénéfices directs pour les chercheurs

4. Préserver leurs données
   - en les déposant dans un entrepôt de données,
   - en rendant accessibles les données non publiées grâce à des liens citables. Par exemple, les interventions à des colloques.

5. Améliorer la qualité de leurs données en s'assurant qu'elles seront bien gérées en vue de leur dépôt dans des entrepôt de données.
6.
7. Améliorer la transparence des recherches scientifiques
   - en permettant l'examen des résultats scientifiques,
   - et par conséquent, en réduisant la fraude.

Gérer et partager des données

Comment puis-je faire?

Gestion des données de la recherche
Définition

"La gestion des données fait partie intégrante du processus de recherche."

Université de Leicester, "When do you need to think about RDM"

Pendant un projet de recherche et après l'achèvement du projet:

- Rassembler,
- Organiser,
- Gérer,
- Stocker,
- Sauvegarder,
- Préserver,
- Partager vos données.

Gestion des données de la recherche
Définition

La gestion des données implique généralement:

- De créer un **Plan de Gestion des Données** (DMP), soumis avec la demande de financement, afin d'expliquer comment les données seront gérées pendant et après l'achèvement du projet ;

- De stocker des données de recherche durant le projet et de les partager avec les collègues autorisés ;

- A la fin du projet, de déposer ces données dans un entrepôt pérenne, et de les rendre disponibles pendant plusieurs années.