

CommonData

Les données de la recherche, des communs scientifiques ?

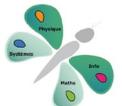
Coordinatrice : Agnès Robin, *Maître de conférences HDR en Droit privé à l'Université de Montpellier –*

Co-auteurs :

Philippe Amiel
Pierre-Yves Lacour
Francesca Frontini

Présentant :

Pierre-Yves Lacour



DU SCIENTIFIC DATA MANAGEMENT GESTION DES DONNÉES DE LA RECHERCHE

Responsable : Agnès Robin
Maître de conférences HDR, Université de Montpellier

Inscriptions :
<https://sdm.edu.umontpellier.fr/>



Actions de CommonData

- Cycle de **6 journées d'étude** autour du « cycle de vie des données »
 - Notion de donnée scientifique : visions pluridisciplinaires (oct. 2017)
 - Méthode, intégrité scientifique et big data (fév. 2018)
 - Stockage et archivage des données scientifiques : solutions techniques et garanties juridiques (juin 2018)
 - Partage et diffusion : disponibilité technique et juridique des données scientifiques (juill. 2018)
 - Corpus de données et auctorialité (oct. 2018)
 - Exploitation économique des données scientifiques (nov. 2018)
- **Etude des pratiques** des chercheurs du bassin régional : **enquête pluridisciplinaire** (1^{ère} en France avec les STM) > action de **recherche** et non de valorisation
- **Entretiens** avec les services de valorisation des établissements et organismes de recherche

Restitution des résultats

Objectifs

- **Documenter les pratiques : moyens et outils** (techniques ou contractuels) utilisés par les chercheurs et **comportements** des chercheurs tels qu'ils pouvaient le déclarer
- **Documenter les conceptions de la relation des chercheurs avec leurs données**
- **Documenter la compétence des chercheurs à l'égard des règles de droit : incorporation** des normes relatives aux données dans le processus de recherche

Méthode

Questionnaire auto-administré en ligne

Anonymat garanti aux répondants

Inclusion :

- Chercheurs et ingénieurs de recherche actifs de laboratoires académiques dans toutes disciplines (lettres et sciences humaines comprises).

Non inclusion :

- Chercheurs en biomédecine réalisant des « recherches impliquant la personne humaine » (RIPH) ;
- Chercheurs dont l'activité est sans manipulation de données.

Méthode

Termes :

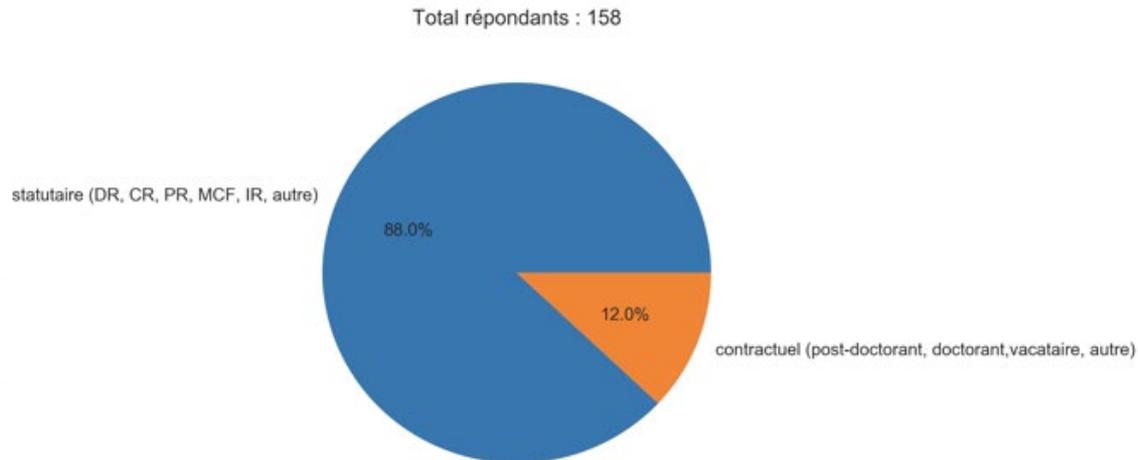
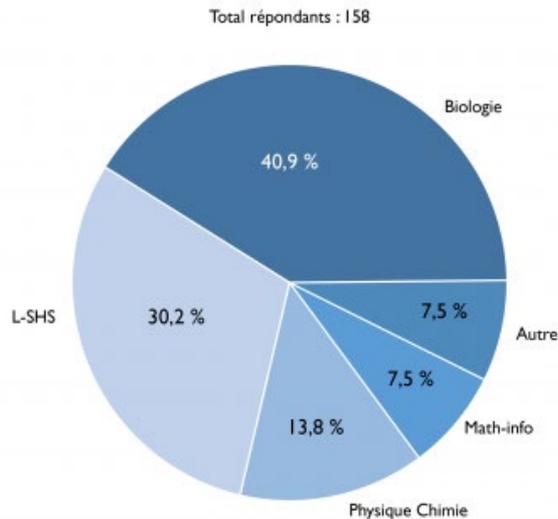
- Données de « votre dernière recherche » ou de la recherche la plus significative du point de vue de la gestion des données
- **Définition** des données scientifiques : *tous relevés, enregistrements factuels ou renseignements factuels, notamment : chiffres, textes, images et sons, quels qu'en soient le support (audiovisuel, archives, publications, base de données, plateformes numériques, etc.), utilisés pour produire des résultats de recherche*

Réalisation

- Enquête déroulée du 31 mai au 30 juin 2018 (4,3 semaines)
- 258 réponses enregistrées ; 159 questionnaires complets de chercheurs répondant aux critères d'inclusion.

→ **159 questionnaires exploités**

Population



- Des laboratoires et des financements publics, avec des partenaires privés (37,1 %) dans certaines disciplines (biologie et physique-chimie)

Données

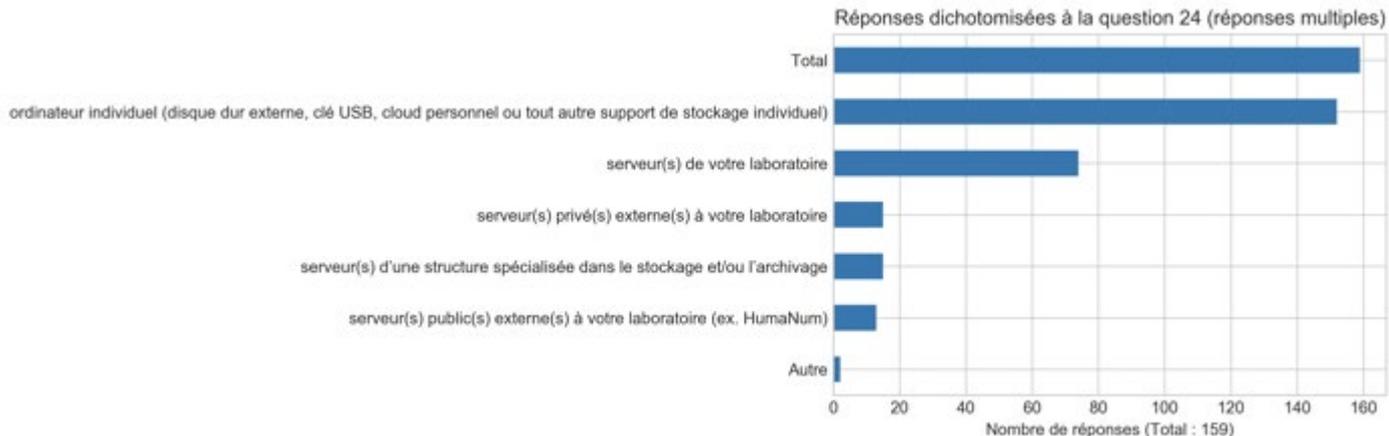
	Effectif (n)	Proportion (%)
(1) utilisation d'un appareillage (type capteur, GPS, autre)	80	51 %
(2) consultation d'autres bases de données	58	37 %
(3) génération des données par ordinateur (analyse computationnelle)	57	36 %
(4) observation sur site (hors internet)	54	34 %
(5) lecture de documentation (autre que archives)	46	29 %
(6) réalisation de questionnaires/entretiens/enquêtes	35	22 %
(7) consultation de fonds d'archives (départementales, nationales, etc.)	19	12 %
(8) Autre	15	9 %
(9) réalisation d'exams cliniques	7	4 %
Total	158	100 %

- 73,4 % des répondants ont déclaré avoir travaillé à la collecte avec au moins une autre personne

Réutilisation

	À être conservées ou archivées pour des recherches futures	À être utilisées immédiatement pour votre recherche	Les deux
Autre	0	7	5
Biologie	0	38	27
L-SHS	0	20	28
Math-info	2	6	4
Physique Chimie	2	11	9

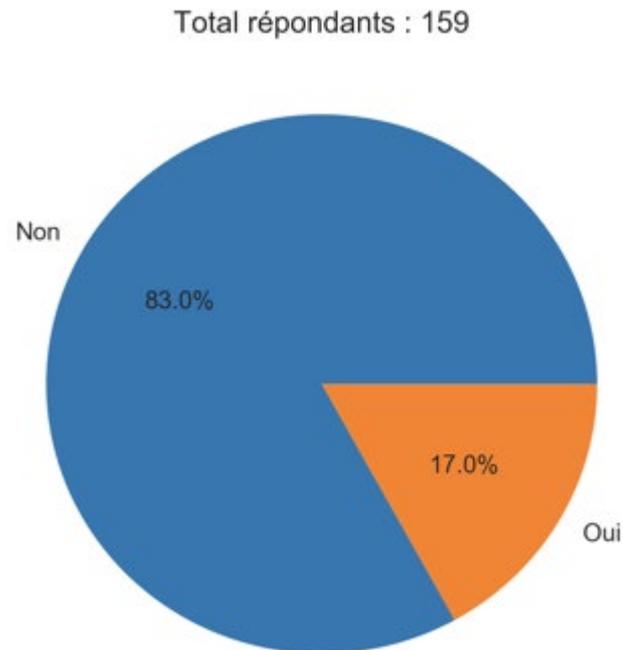
Gestion, stockage



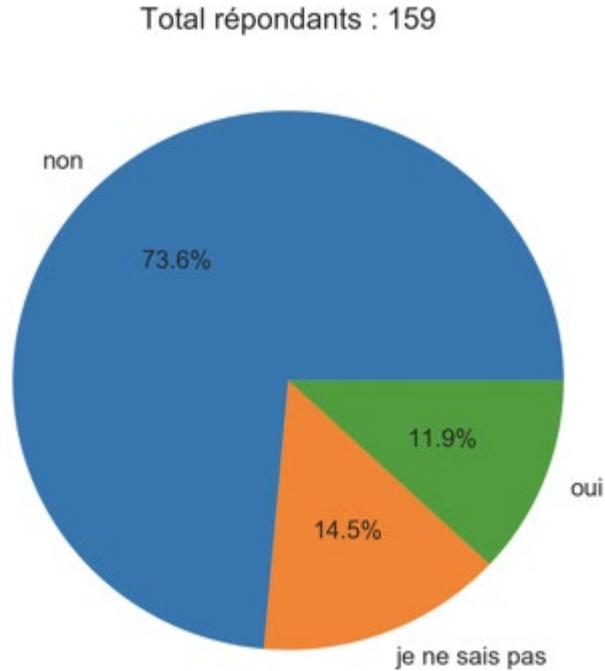
- Stockage pour la plupart individuel, avec un faible recours aux procédures standardisées et aux services de stockage/archivage institutionnels.
- En pratique, seulement 28 % des répondants ont fait appel à des dispositifs d'entrepôts de données, à des serveurs publics d'archivage et conservation ou à des serveurs ou entrepôts privés.

Standards et interopérabilité

- Les deux tiers des répondants **n'ont mobilisé aucun standard, norme ou procédure** d'organisation de leurs données.
- Ces standards ou procédures étaient, dans la plupart des cas, définis selon des **usages disciplinaires**, par le chercheur lui-même pour servir dans le cadre de sa recherche en cours.
- **Moins d'un cinquième des répondants à utilisé un protocole d'interopérabilité** (utilisation de vocabulaire contrôlé, d'ontologies de référence ou de standard de métadonnées)

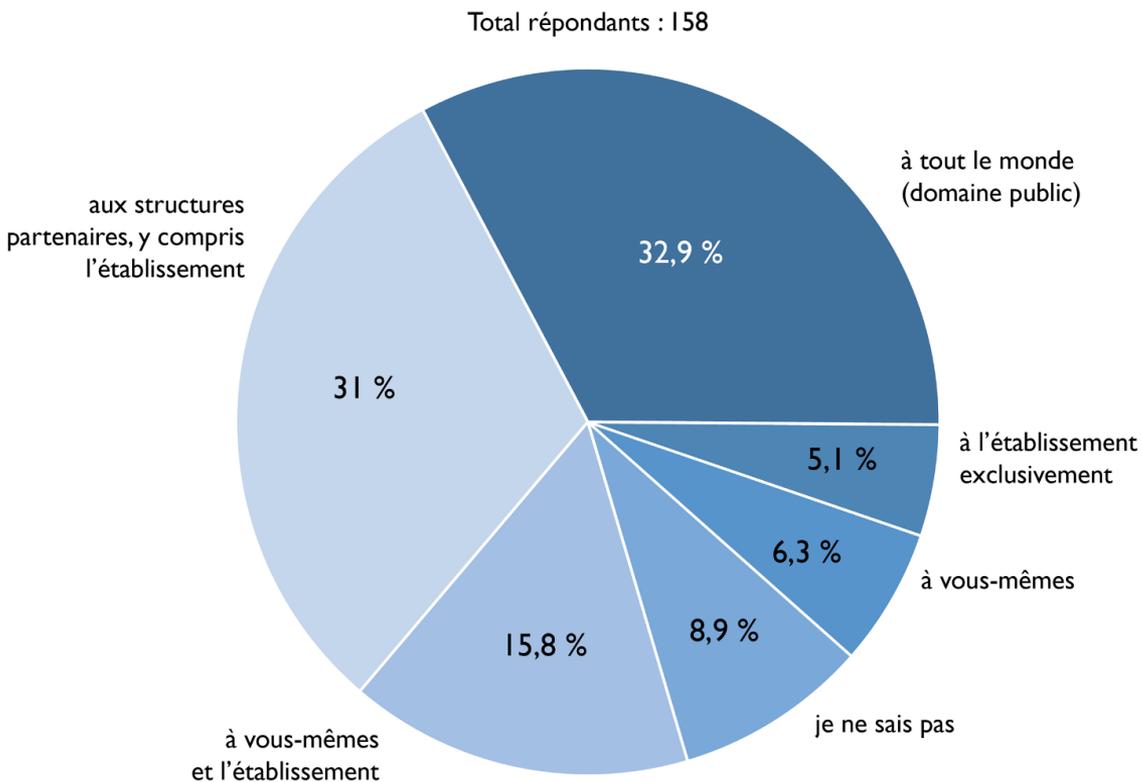


Exploitation économique ?



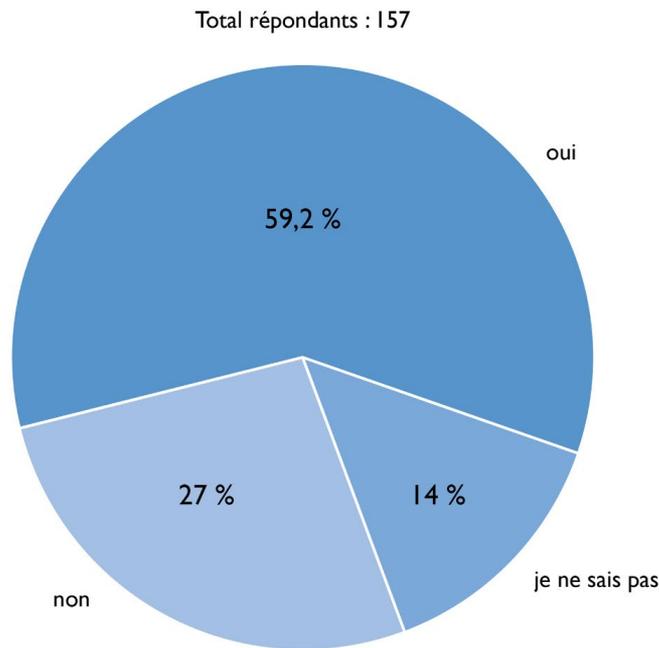
- **L'exploitation économique** des données est **minoritaire** dans notre échantillon

A qui appartiennent les données?



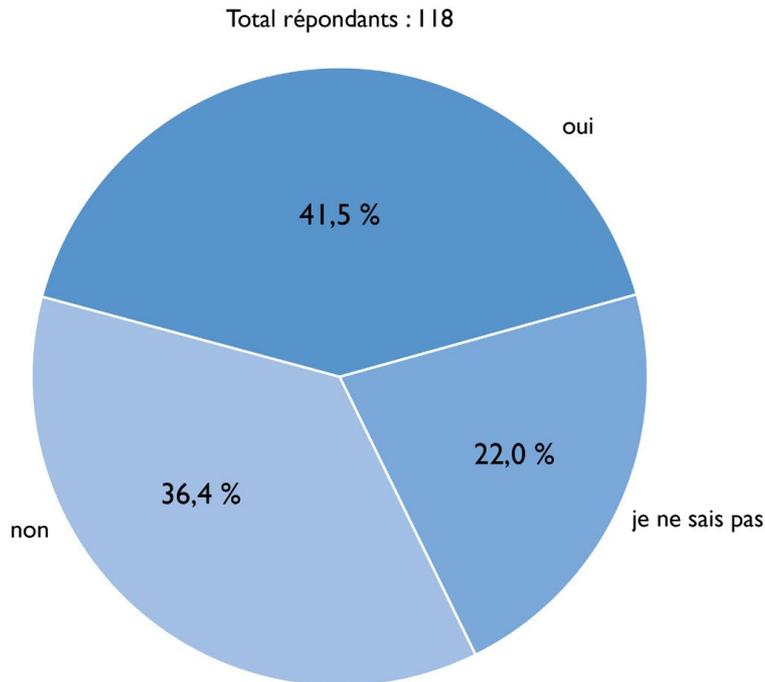
- Un tiers (33 %) des chercheurs interrogés ont répondu « **à tout le monde** », les données étant alors vues comme relevant du domaine public (notion mal distinguée de celle de « commun »).
- **L'origine des crédits** (financement de labo vs projet) dévolus à la dernière recherche influe sur les réponses.

Si vous quittiez votre unité, emporteriez-vous vos données avec vous ?



- Le **sentiment de propriété** des données variait selon les disciplines : 77 % des chercheurs en physique-chimie et le même pourcentage en L-SHS, contre 53 % en biologie, étaient prêts à s'approprier les données en les emportant avec eux
- même si 74 % de ces 93 « propriétaires » en laisseraient une **copie à leur laboratoire d'origine**

Partage



Les données de votre dernière recherche seront-elles diffusées (hors publication scientifique) ?

- À cette question qui est aujourd'hui au cœur du dispositif légal de science ouverte, les chercheurs répondent de **manière assez diversifiée**. Souvent recherches sont menées sans qu'ait été au préalable envisagée la diffusion des données en dehors des publications.
- Plus précisément, sur la chronologie de la diffusion, les chercheurs déclarent très majoritairement qu'ils diffuseront les données **d'abord à destination des chercheurs et des partenaires, puis du public**.

Acculturation à la norme juridique ?

- La **stratégie** de gestion des données reste le plus souvent **impensée** lorsque les données n'ont pas vocation à faire l'objet d'une valorisation économique.
- Les réponses montrent que la catégorie du **domaine public**, confondue avec celle de *propriété publique*, est une catégorie résiduelle non précisément identifiée.
- Le sentiment de propriété à l'égard des données, essentiellement lié à l'auctorialité (paternité) ou au travail fourni (effort), entre en **contradiction avec les réalités juridiques et institutionnelles**.

Auteurs : Ph. Amiel, F. Frontini, P.-Y. Lacour, A. Robin

article : <https://journals.openedition.org/cdst/2061>

site : commondatafr.wordpress.com

twitter : #scicommondata

