

## Le partage des données vu par les chercheurs : une approche par la valeur

Article inédit, mis en ligne le 1er juillet 2021.

### Violaine Rebouillat

*Violaine Rebouillat est docteure en Sciences de l'information et de la communication. Elle est membre associée du laboratoire Dicen-IDF (EA 7339), où elle a soutenu sa thèse en 2019, et de l'unité de recherche ELICO (EA 4147). Elle travaille actuellement comme ATER au sein de l'équipe 71ème du département Informatique de l'Université Claude Bernard Lyon 1. Ses travaux analysent les pratiques et représentations des différents acteurs impliqués dans la production, l'utilisation et la gestion des données de la recherche. Ils posent la question des significations socio-économiques de logiques qui entrent en compte dans les dynamiques de l'ouverture. violaine.rebouillat@univ-lyon1.fr*

### Plan de l'article

Introduction

Les données de la recherche définies : objets pluridisciplinaires complexes

Les données de la recherche dans la communication scientifique

La question de la valeur adressée aux données de la recherche

Au cœur du laboratoire, à l'ombre des injonctions

Une valeur d'usage mesurée en unités comptables de l'évaluation

Valeur d'échange, ou les perspectives de reconfiguration de la valeur des données

Conclusion

Références bibliographiques

Remerciements

Annexes

### RÉSUMÉ

Le propos de cet article porte sur la compréhension des logiques qui interviennent dans la définition de la valeur des données de la recherche, celles-ci pouvant avoir une influence sur les critères déterminant leur motivation au partage. L'approche méthodologique repose sur une enquête qualitative, menée dans le cadre d'une recherche doctorale, qui a déployé 57 entretiens semi-directifs. Alors que les travaux menés autour des données sont focalisés sur les freins et motivations du partage, l'originalité de cette recherche consiste à identifier les différents prismes par lesquels la question de la valeur des données impacte la motivation et la décision de leur partage. L'analyse des résultats montre que, tous domaines confondus, la valeur des données reste encore cristallisée autour de la publication et de la reconnaissance symbolique du travail du chercheur. Les résultats permettent de comprendre que la question du partage est confrontée à un impensé : celui du cadre actuel de l'évaluation de la recherche, qui met l'article scientifique au cœur de son dispositif. Ce travail contribue donc à montrer que l'avenir du partage des données dépend des systèmes alternatifs futurs d'évaluation de la recherche, associés à la science ouverte.

## Mots clés

Données de la recherche, communication scientifique, science ouverte, cycle de la recherche, évaluation de la recherche, capital symbolique, reconnaissance sociale

## TITLE

Researchers' views on data sharing: a value-based approach

## Abstract

The aim of the paper is to provide an analysis of the logics involved in defining research data value, which can influence the criteria for sharing them. Methodological approach is based on a qualitative survey, conducted during a doctoral research project, which deployed 57 semi-directive interviews. While most scientific work on data focuses on the barriers and motivations for sharing, the originality of this research consists in identifying how data value impacts the decision of sharing them. Results show that, in all fields, data value still remains crystallized around the publication and symbolic recognition of the researcher's work. The study allows us to understand that the question of sharing is confronted with an "unthought": that of the current framework of research evaluation, which places scientific papers at the heart of its mechanism. This work thus contributes to show that the future of data sharing depends on alternative future systems of research evaluation, associated with open science.

## Keywords

Research data, scientific communication, open science, research cycle, evaluation of research, symbolic capital, social recognition

## TITULO

Compartir datos desde la perspectiva de los investigadores: un enfoque de valores

## Resumen

El objetivo de este artículo es proporcionar un análisis de la lógica que interviene en la definición del valor de los datos de investigación, que puede influir en los criterios para determinar su compartimiento. El enfoque metodológico se basa en una encuesta cualitativa, realizada como parte de una investigación doctoral, que desplegó 57 entrevistas semiestructuradas. Si bien la mayor parte de la labor científica sobre los datos se centra en los obstáculos y las motivaciones para compartirlos, la originalidad de esta investigación consiste en identificar los diferentes prismas a través de los cuales la cuestión del valor de los datos repercute en la decisión de compartirlos. El análisis de los resultados muestra que, en todos los campos tomados en conjunto, el valor de los datos sigue cristalizado en torno a la publicación y el reconocimiento simbólico del trabajo del investigador. Los resultados permiten comprender que la cuestión de la puesta en común se enfrenta a un impensable: el del marco actual de evaluación de la investigación, que sitúa el artículo científico en el centro de su mecanismo. Por lo tanto, esta labor ayuda a demostrar que el futuro del intercambio de datos depende de futuros sistemas alternativos de evaluación de la investigación, asociados con la ciencia abierta.

## Palabras clave

Datos de investigación, comunicación de la investigación, ciencia abierta, ciclo de investigación, evaluación de la investigación, capital simbólico, reconocimiento social

## INTRODUCTION

Parmi les questions qui agitent l'actualité de la pandémie de la Covid-19, celle de l'accès aux productions scientifiques a été largement relayée par les médias. Les éditeurs scientifiques ont ouvert l'accès aux articles relatifs à la Covid<sup>1</sup> et les serveurs de preprints (BioRxiv, MedRxiv...) ont accueilli des milliers d'articles, accentuant l'accroissement exponentiel de l'information scientifique. Au 16 juillet 2020, plus de 70.000 articles scientifiques étaient disponibles sur le virus SARS-CoV-2, dont plus de 9500 depuis des serveurs de preprints<sup>2</sup>.

Or, les données de la recherche ne semblent pas connaître le même le phénomène. Plusieurs constats convergent sur le fait que les données sur la Covid-19 n'ont pas été partagées, ou seulement à la marge. Une étude d'Homolak *et al.* (2020) révèle ainsi des problèmes de disponibilité des données associées aux articles publiés pendant le premier trimestre de la pandémie. Quant aux débuts difficiles du projet européen Discovery dans la recherche de traitements contre la Covid-19<sup>3</sup>, ils sont le témoin de la difficulté à fédérer les scientifiques autour de la réalisation d'essais cliniques de grande envergure. Les politiques d'ouverture des données se heurtent donc manifestement à un verrou. Ce phénomène est à interroger car les données participent tout autant aux avancées de la connaissance scientifique que les publications. C'est la raison pour laquelle elles sont aujourd'hui soumises aux politiques de science ouverte.

## LES DONNÉES DE LA RECHERCHE DÉFINIES : OBJETS PLURIDISCIPLINAIRES COMPLEXES

Une des toutes premières occurrences du terme « données de la recherche » provient d'une déclaration de l'OCDE, qui en 2004 engage trente États dans le développement de régimes d'accès ouverts aux données de la recherche, financée sur fonds publics (OCDE, 2004). Cette déclaration est complétée en 2007 par des principes et lignes directrices (OCDE, 2007), qui poseront la définition suivante :

*« 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 des résultats de recherche. Un ensemble de données de recherche constitue une représentation systématique et partielle du sujet faisant l'objet de la recherche. Ce terme ne s'applique pas aux éléments suivants : carnets de laboratoire, analyses préliminaires et projets de documents scientifiques, programmes de travaux futurs, examens par les pairs, communications personnelles avec des collègues et objets matériels (par exemple, les échantillons de laboratoire, les souches bactériennes et les animaux de laboratoire tels que les souris). L'accès à tous ces produits ou résultats de la recherche est régi par d'autres considérations que celles abordées ici. » (OCDE, 2007, p.18)*

L'expression « données de la recherche » est donc relativement nouvelle et fait émerger un objet jusqu'alors absent des discours des politiques publiques. En les nommant, elles mettent en visibilité des produits de la science, dans le but de les rendre accessibles.

.....

1. *Sharing research data and findings relevant to the novel coronavirus (COVID-19) outbreak* (2020, janvier 31). Wellcome Trust. <https://wellcome.org/coronavirus-covid-19/open-data>
2. *Biologie : La pandémie valorise les prépublications* (2020, juillet 16). CNRS. <https://www.cnrs.fr/fr/cnrsinfo/biologie-la-pandemie-valorise-les-prepublications>
3. Morin, H. (2020, mai 7). *Coronavirus : L'essai clinique Discovery englué faute de coopération européenne*. *Le Monde.fr*. [https://www.lemonde.fr/planete/article/2020/05/07/coronavirus-l-essai-clinique-discovery-englué-faute-de-cooperation-europeenne\\_6038909\\_3244.html](https://www.lemonde.fr/planete/article/2020/05/07/coronavirus-l-essai-clinique-discovery-englué-faute-de-cooperation-europeenne_6038909_3244.html)

Leur discours prend en compte l'ensemble des domaines disciplinaires, qu'elles considèrent tous comme concernés par la question. Or ce que recouvre le terme « données de recherche » n'est pas univoque. À titre illustratif, une enquête qualitative de F. Cabrera (2014, p. 60) montre l'hétérogénéité des objets concernés, à la seule échelle des Sciences humaines et sociales – rompant ainsi avec la possibilité d'en établir une liste exhaustive.

Une autre manière de définir les données de la recherche consiste à les catégoriser. Différentes typologies existent, comme celle de la *National Science Foundation*, fondée sur la méthodologie de collecte des données (National Science Board, 2005). On peut également citer celle du système d'observation de la Terre de la NASA<sup>4</sup>, qui consiste à catégoriser les données en fonction de leur niveau de traitement (des données brutes aux données dérivées). Ou bien encore celle utilisée dans l'enquête de Prost et Schöpfel, distinguant données sources et données produites (Prost & Schöpfel, 2015).

Ces typologies s'inscrivent dans une perspective de traitement des données souvent opérationnelle qui vise à les saisir par leur plus petit dénominateur commun. Néanmoins, cette approche pragmatique de la catégorisation tend à masquer la diversité des données. Son apport est limité lorsque le but est de les appréhender dans leur complexité et leur ancrage social.

L'un des apports majeurs des travaux de C. L. Borgman, spécialiste des données de la recherche, est d'avoir montré qu'il n'est pas possible de définir les données dans l'absolu. Elle montre, à partir d'études de cas réalisées dans divers champs scientifiques, que la nature des données varie en fonction de la discipline, de l'objectif de la recherche, de la méthodologie et de l'instrumentation mobilisés (Borgman, 2015). Les données ne sont donc pas des objets purs ou naturels ayant une essence propre. Elles existent dans un contexte et prennent leur sens en fonction de ce contexte et des conditions de travail qui s'y déroulent.

S. Leonelli (2015), philosophe des sciences et spécialiste des données de la recherche, écrit qu'il n'existe pas de données en elles-mêmes. Elles ne sont pas définies selon leurs propriétés intrinsèques mais selon leur fonction au sein de processus de recherche particuliers. Leonelli propose ainsi de considérer les données comme des produits de la recherche collectés, enregistrés et diffusés dans la perspective d'être utilisés ensuite comme preuve d'une théorie scientifique au sujet d'un phénomène particulier. L'expertise du chercheur joue ainsi un rôle important dans la détermination de ce qui constitue une donnée. C'est le chercheur qui, par un arbitrage scientifique, érige une entité en donnée, comme étant la preuve du phénomène étudié. Les données apparaissent donc comme des objets « fabriqués » par les chercheurs, des artefacts, produits dans le temps de travail du chercheur, ancrés dans sa pratique et ses interactions avec sa communauté et ses pairs.

Les tentatives de définitions ci-dessus sont avancées par plusieurs disciplines, dont les Sciences de l'information et de la communication (SIC). Celles-ci ont la particularité d'accueillir les données de la recherche en reconnaissant leur complexité, prenant en compte l'ancrage social des outils numériques et des structures disciplinaires dans lesquelles elles sont produites. Cette approche, propre aux SIC (Miège, 2007), permet d'appréhender notre objet de recherche en intégrant cette complexité. Les SIC disposent, par ailleurs, d'outils conceptuels et théoriques pour les penser au sein du triptyque « donnée, information, connaissance ». M. J. Bates exprime clairement cette filiation en définissant la donnée comme une entité indivisible, n'acquérant d'utilité que par son association dans un contexte de production d'informations (Bates, 2005).

.....

4. « Data Processing Levels », [en ligne], consulté le 1er décembre 2020, <https://earthdata.nasa.gov/earth-science-data-systems-program/policies/data-information-policy/data-levels>

## LES DONNÉES DE LA RECHERCHE DÉFINIES : OBJETS PLURIDISCIPLINAIRES COMPLEXES

### Importance de l'article

Dans l'histoire de ses avancées, la sociologie des sciences a mis au centre de ses travaux la question de la reconnaissance. D'abord avec ceux de R. K. Merton autour des normes sociales de la science (CUDOS), qui accordent une importance cruciale à la notion de la structure sociale des sciences et de la place des pairs (Merton, 1973). Puis avec les travaux de P. Bourdieu (1975), autour de sa théorie du champ scientifique, fait de luttes pour la reconnaissance symbolique. Dans les années 1990, la mise en place d'un nouveau mode de gestion, qualifié de *National Science Foundation*, a conduit à une cristallisation de l'évaluation de la recherche sur la mesure quantitative de la productivité du chercheur et de l'impact de ses publications. Les analyses d'Y. Gingras montre comment, dans ce contexte, l'article est devenu une unité comptable participant à la marchandisation de la publication scientifique (Gingras, 2018).

Dans le système actuel de la recherche, l'article fait ainsi converger trois valeurs : une valeur scientifique, une valeur symbolique et une valeur marchande. Il n'est donc pas anodin que la publication conditionne l'évolution de la carrière du chercheur et concentre ses efforts et ses attentions. Ce phénomène apparaît clairement dans les stratégies de carrière des jeunes chercheurs, qui choisissent de concentrer leurs efforts sur la publication d'articles dans des revues prestigieuses (Nicholas *et al.*, 2017).

### La valeur des données dans le champ scientifique

La question de la valeur des données dans le champ scientifique est encore peu développée dans la littérature. Cela tient à la jeunesse des données en tant qu'objet de recherche, lesquelles sont davantage travaillées par des approches empiriques que conceptuelles et théoriques.

Les travaux de B. Latour nous invitent à considérer ce qu'il appelle des « inscriptions » (Latour & Woolgar, 1979), des objets intermédiaires des savoirs produits, en préalable à la publication. Même si Latour ne s'est pas intéressé aux données de la recherche en tant que telles, son approche nous permet de les inclure dans cette catégorie. Cette approche rejoint des travaux plus récents (Fecher *et al.*, 2015), portant explicitement sur les données du point de vue de leur valeur. Leur intérêt est d'aborder le système d'échange de l'information scientifique comme une *Reputation Economy* (Fecher *et al.*, 2015). Par *Reputation Economy*, ils entendent la reconnaissance symbolique des pairs et de la communauté. La production scientifique n'est partagée que si elle permet au chercheur d'accumuler cette réputation. Celle-ci étant principalement liée à ses publications (articles de revues, monographies, actes de colloques...), le chercheur ne trouvera donc que peu d'intérêt à diffuser d'autres produits de son travail – les données scientifiques en particulier. Selon Fecher *et al.*, les données de la recherche ne possèdent par conséquent qu'une faible valeur d'échange (Fecher *et al.*, 2017). L'approche de la reconnaissance permet ainsi de donner du sens au manque d'intérêt que peuvent avoir les chercheurs à partager leurs données.

Plusieurs études permettent de valider cette thèse. Elles ont montré que l'ouverture des données était un sujet connu des chercheurs, mais qu'elle était évacuée par des préoccupations de publication (Da Costa & Lima Leite, 2019; Tenopir *et al.*, 2011; Wallis *et al.*, 2013).

## La valeur des données vue par les politiques publiques

Depuis le milieu des années 2000, les données occupent une part grandissante dans le champ de la communication scientifique. Comme le formule S. Leonelli :

« *Les données tendent à être conceptualisées de plus en plus comme des produits de la recherche détenant une valeur en elles-mêmes, et de moins en moins comme des composants du processus de recherche n'ayant pas de valeur inhérente* » (Leonelli, 2013, p.7).

En prônant l'ouverture des données, les politiques publiques véhiculent l'idée selon laquelle ces dernières constituent des entités à fort potentiel de réutilisation. Elles avancent deux catégories d'arguments pour justifier le partage des données. La première catégorie s'appuie sur une conception cumulative du savoir : le partage des données contribuerait au renforcement de la transparence de la recherche, à la réutilisation des résultats par la communauté scientifique et l'enseignement, mais aussi par les entreprises et les citoyens, et au développement de nouvelles formes de recherche fondées sur l'usage de données massives. La seconde catégorie d'arguments est d'ordre économique : le partage des données favoriserait l'innovation et la croissance économique et contribuerait au retour sur investissement des sommes engagées par les instances publiques dans le financement de la recherche (CE, 2012; MESRI, 2018; OCDE, 2007). La valeur que les politiques d'ouverture confèrent aux données est ainsi à la fois une valeur normative, liée à l'intégrité scientifique et à la science comme bien commun (Merton, 1973), et une valeur économique fondée sur une vision néolibérale du concept d'économie de la connaissance (Foray & Lundvall, 1997).

Pour autant, lorsqu'on se penche sur les travaux ayant investi les pratiques de partage des données, la littérature scientifique montre à quel point les chercheurs qui adhèrent à la valeur du partage des savoirs sont, dans les faits, peu motivés à partager leurs données (Tenopir *et al.*, 2011 ; Andreoli-Versbach & Mueller-Langer, 2014; Fecher *et al.*, 2015; Serres *et al.*, 2017). L'un des intérêts de l'étude de C. Tenopir (menée auprès de 1329 chercheurs) est aussi d'avoir souligné un paradoxe : les chercheurs conviennent à 60% que le manque d'accès aux données constitue un obstacle majeur au progrès de la science, mais déclarent pour 46% d'entre eux ne pas mettre leurs données à disposition.

Comme le montrent Levin et Leonelli avec l'exemple de la biologie (Levin & Leonelli, 2017), les chercheurs sont pris en tension entre un idéal de la science, représenté par le CUDOS de R. K. Merton (1973), consistant à partager les connaissances nouvelles, et l'absence d'intérêt que cela représente pour eux de rendre les données publiques, étant donné le peu de reconnaissance qu'ils en retirent et la réalité de la lutte du champ scientifique que décrit P. Bourdieu (1975).

## LA QUESTION DE LA VALEUR ADRESSÉE AUX DONNÉES DE LA RECHERCHE

Notre question de recherche se positionne donc sur les mécanismes de valeur qui entrent en compte dans les motivations ou résistances au partage des données scientifiques. En donnant la parole aux chercheurs, nous appréhendons les données dans le contexte dans lequel elles sont produites (ou coproduites). L'objectif est de comprendre les raisons pour lesquelles les chercheurs adhèrent ou résistent au principe de partage. Nous explorons comment ils définissent eux-mêmes leurs données, dans le contexte quotidien de leur travail. En filiation des travaux de W. D. Garvey (1979), analysant les évolutions de la communication scientifique, nous prenons également en compte l'ancrage social des pratiques et donc du rôle des interactions des chercheurs avec leurs collègues, étudiants (Master), doctorants, post-doctorants et pairs (au niveau national et international).

Nous avons fait le choix d'une méthode qualitative, dont la dimension interprétative rend compte de la complexité de « l'objet » données et de la question de leur partage. Nous avons mené des entretiens semi-directifs auprès de 57 chercheurs du domaine des Sciences, de la Santé et des Sciences humaines et sociales, réalisés dans le cadre d'une recherche doctorale (Rebouillat, 2019a). Les entretiens ont eu lieu entre novembre 2017 et mai 2019 dans les bureaux des laboratoires de l'Université de Strasbourg ([annexe 1](#)). D'une durée moyenne de 1h30, ils ont donné lieu à une retranscription systématique.

L'analyse des entretiens nous permet de situer les données dans les différentes étapes de conception, production, diffusion et valorisation de l'information scientifique. Dès lors, s'inscrit en creux des entretiens le statut de la valeur des données. Leur analyse a reposé sur les thématiques transversales autour desquelles se sont structurées les perceptions que les chercheurs avaient de la valeur de leurs données. La dimension pluridisciplinaire du panel a permis d'analyser les entretiens dans la perspective de dégager les différentes logiques de valeur entrant en compte dans la production, l'analyse et la valorisation des données de recherche. Nous livrerons par ce biais une meilleure compréhension de l'apparente contradiction des chercheurs, dont le discours, favorable à l'ouverture des données, n'est pas ou peu corroboré par leurs pratiques. Il s'agit de justifier cette apparente contradiction, en mobilisant le prisme de la valeur des données. Comment la question de la valeur des données émerge dans les discours des chercheurs ? Comment est-elle caractérisée à partir de ses pratiques de recherche et de communication scientifique ?

Cette problématique, positionnée en amont de l'activité du chercheur, à l'étape de la conception, couvre une zone d'ombre peu traitée dans la littérature scientifique. Son originalité réside dans le fait de donner la voix aux chercheurs, pour explorer les logiques structurelles qui font du partage des données une pratique rarement spontanée chez eux.

L'article relève des Sciences de l'information et de la communication. Il contribue au champ de recherche de l'*Open Science* et des données de la recherche en France (Chartron & Schöpfel, 2017) et à l'international (Borgman, 2015; Tenopir *et al.*, 2019), mais aussi à celui de l'*Open Data* et du *Big Data* sur la compréhension du phénomène de *dataification* et de ses implications sociales (Broudoux & Chartron, 2015; Bullich & Clavier, 2018). Il s'inscrit comme une contribution à la compréhension de l'enjeu des données dans les mutations actuelles et de déployer, à nouveaux frais, des stratégies de recherche et des méthodes adaptées (Miège, 2020).

L'analyse des résultats s'articule autour de trois parties. Nous mettrons d'abord en évidence deux registres de valeur des données : la valeur pour soi et la valeur d'usage. Nous montrerons que si l'une est inhérente au mécanisme intellectuel de conception d'un projet scientifique, la seconde est intimement liée au système actuel d'évaluation de la recherche. Nous avancerons ensuite un troisième registre de valeur : la valeur d'échange, qui tend à être de plus en plus présente dans la communauté scientifique. Nous verrons qu'elle repose tantôt sur l'obligation de mettre à disposition les données au moment de la publication, tantôt sur l'acquisition d'un capital social en échange du partage des données. L'article conclut sur les perspectives d'ajustements entre ces différents régimes de valeur et les nouvelles valeurs autour desquelles l'évaluation de la recherche évolue dans le cadre des politiques de science ouverte.

## AU CŒUR DU LABORATOIRE, À L'OMBRE DES INJONCTIONS

La valeur des données se construit initialement dans le laboratoire, ce lieu clos où le chercheur mature ses idées et ses projets scientifiques. Nous considérerons donc ici la dimension intime du processus de recherche, qui à la fois forge la valeur des données et détermine leur mode de gestion.

## Valeur plurielle, valeur pour soi

À la question « Quelle valeur ont les données que vous avez générées ? », un chercheur a répondu : « *Ce sont nos enfants !* » (chercheur 3). Au cours des entretiens, les chercheurs ont tous témoigné d'un attachement aux données qu'ils avaient collectées. Cette image de filiation des données à leur producteur sous-entend un lien fort, lié vraisemblablement aux efforts investis par le chercheur pour les acquérir, les mettre en forme, les analyser et les conserver.

Un autre aspect qui se dégage de l'analyse des entretiens, toutes disciplines confondues, est que la perception de la valeur des données diffère d'un chercheur à un autre, d'un domaine à un autre. La diversité des propos tenus témoigne de la relativité des perceptions. La question de la valeur des données est de ce fait difficile à saisir car plurielle : elle peut être fonction du coût lié à la production des données, du temps investi ou de l'expertise engagée à les générer.

En Biologie, les chercheurs considèrent certaines données comme « précieuses », du fait de leur coût d'acquisition. L'un des chercheurs expliquait ceci :

« *Pour les données très sensibles comme celles-là (sensibles parce que c'est 1500€ le séquençage), on a un triple stockage* » (chercheur 3).

On retrouve cette même considération en Neurosciences :

« *La valeur qu'on peut donner aux données, c'est toute la masse d'argent qui a été investie pour les obtenir. Par masse d'argent, je pense aux animaux, aux réactifs, aux salaires...* » (chercheur 22).

En Géochimie, c'est la question du volume de travail qui détermine la valeur de certaines données. Le calcul d'une mesure isotopique demande un travail d'acquisition trop important pour que celui qui l'a réalisé se permette de le mettre aussitôt à la disposition de tous. Plus la donnée est complexe à acquérir, moins le chercheur sera enclin à la rendre librement consultable :

« *Typiquement, quand on fait une mesure isotopique, on ne va pas la mettre sur internet le lendemain. Parce qu'obtenir un rapport isotopique, ça représente des mois de travail* » (chercheur 37).

La pluralité des points de vue couverts ci-dessus montre néanmoins un point de convergence : quelles que soient les différences de perception de la valeur des données, celle-ci est une valeur pour soi, pour le chercheur en tant qu'acteur, qui y projette la possibilité de produire et de publier de nouvelles connaissances. Cette étape de définition de la valeur, essentiellement individuelle, est préalable à celle d'une perception de la valeur en regard de la communauté (pairs et collègues). Cette dimension permet d'appréhender l'aspect affectif de la citation ci-dessus, relative aux données comme « enfants », et de comprendre la vision défendue par le géographe ci-dessous :

« *Il y a un aspect un peu précieux de la donnée, qui est qu'on ne la montre pas avant de l'avoir soi-même exploitée. C'est tout l'enjeu de la compétition scientifique* » (chercheur 45).

Notre analyse rend compte d'une définition des données de la recherche qui rejoint complètement les définitions proposées par C. Borgman (2015) et S. Leonelli (2015). En revanche, elle les prolonge avec une facette inédite qui relève de la sphère personnelle et affective du chercheur. Cette facette traduit l'intimité créative du chercheur, sa capacité à porter un regard personnel et singulier sur un objet de recherche.

## Une valeur, malgré les contraintes

Les chercheurs dans le secteur public posent depuis plusieurs années le problème du temps de travail scientifique, qui est « grignoté » par les tâches administratives et gestion-

naires (Collectif P.é.c.r.e.s., 2011). Ce problème du manque de temps s'est retrouvé dans tous les entretiens que nous avons menés. Il a plus particulièrement été abordé sous l'angle de l'arbitrage des priorités et de l'optimisation de la gestion du temps.

Les chercheurs disent avoir peu de temps à consacrer à la gestion des données. Dans un mode de travail à flux tendu, celle-ci ne constitue pas une priorité. Les chercheurs se focalisent davantage sur les activités liées à la valorisation des résultats de recherche sous la forme de publications qui entrent en compte directement dans l'évaluation de leur carrière :

*« La seule mesure qu'on a pour passer les concours et devenir directeur de recherche, c'est malheureusement les publications. C'est le seul outil qu'on opère pour pouvoir juger si oui ou non on est un bon chercheur et si on peut avoir une promotion. Donc il faut qu'on publie. Et on est en concurrence avec d'autres labos. Le premier qui a trouvé quelque chose publie bien ; celui qui trouve le même résultat mais deux ans après, il ne publie pas bien ou il ne publie pas. Donc il faut à un moment être le premier à avoir ce résultat » (chercheur 9).*

La problématique du manque de temps pour la gestion des données s'articule avec celle du manque de moyens appropriés. Un chercheur en Biophysique (chercheur 6) évoquait par exemple le problème de l'obsolescence des supports de sauvegarde à propos d'anciennes données stockées sur CD-Rom.

Cette double contrainte de temps et de moyens explique par exemple que la conservation des données, qui est un des aspects essentiels de leur gestion, soit négligée. Il est en effet rare que les fichiers de données soient conservés sur le long terme, notamment en raison des problèmes d'espace de stockage. Le verbatim, ci-dessous, l'illustre particulièrement :

*« On stocke [le matériel histologique] jusqu'à ce que le travail soit complètement valorisé. Après, on a le même problème que tout espace qui n'est pas plastique, à savoir qu'au bout d'un moment il faut qu'on gère de l'encombrement. Et la meilleure façon de gérer l'encombrement, c'est de jeter. Donc, tant que ça n'est pas publié ou que le matériel garde un intérêt, on conserve à la fois les données générées et le matériel brut. Quand c'est publié, en règle générale, on s'en débarrasse au bout de quelques années » (chercheur 18).*

## UNE VALEUR D'USAGE MESURÉE EN UNITÉS COMPTABLES DE L'ÉVALUATION

Un autre registre est la valeur d'usage. Si elle se caractérise en fonction du système d'évaluation de la recherche, nous verrons qu'elle peut aussi varier selon les enjeux de compétition et le stade d'avancement de la carrière du chercheur.

### La valeur d'usage, une valeur en vue de la publication

Il est intéressant d'observer à partir des entretiens que seules les données associées à une publication font l'objet d'une conservation pérenne, avec souvent triple ou quadruple sauvegarde. Ces données sont conservées pour leur valeur de preuve : en cas de revendication des faits énoncés dans la publication, l'auteur pourra justifier sa démonstration, en utilisant les données comme preuve.

De manière corollaire, les données sont conservées aussi longtemps qu'elles présentent un potentiel de publication et donc une valeur potentielle d'usage. Dans cette approche, l'utilité des données pour le chercheur, en vue d'un projet de publication, détermine leur valeur d'usage. Si le chercheur perd ses données, il perd aussi la possibilité de publier et d'obtenir en échange la reconnaissance de ses pairs. C'est donc dans l'exploitation des données pour la publication que le chercheur trouve la rétribution de son travail de collecte.

Cette vision trouve écho dans la littérature scientifique où, selon des travaux d'économie de l'information, un bien est défini comme valeur d'usage dès lors qu'il est considéré dans un rapport d'utilité (Mayère, 1990, Chapitre III). Autrement dit, c'est l'utilisateur, par la consommation du bien, qui lui confère une valeur d'usage. Cette vision permet d'extraire l'information du statut économique de marchandise, en définissant sa valeur d'usage selon un critère d'utilité – l'information s'intégrant toujours dans un processus déterminé par les événements ou les décisions de l'utilisateur.

Selon cette approche, transposée à l'information scientifique, la question de la valeur se discute en fonction de la place occupée par les données dans ce que B. Latour appelle le « cycle de la crédibilité scientifique » (Latour, 2001). Les données sont donc une matière première essentielle à l'article, car c'est sur elles que se fonde le raisonnement scientifique (Rebouillat, 2019b).

Ce cycle de crédibilité s'observe de manière plus évidente en Sciences, Technologies et Médecine, où les recherches sont le plus souvent financées sur projet. Les données, donnant lieu à des publications (de préférence dans des revues prestigieuses), jouent donc un rôle crucial dans le cycle de crédibilité de la science. Cette dynamique est clairement exprimée par les chercheurs :

*« Si [une équipe de recherche] tombe sur un résultat majeur, [elle va] chercher à le publier dans un journal de forte réputation et, par ce biais-là, [elle va] crédibiliser de futures demandes de financement » (chercheur 18).*

Le mode de financement et les temporalités de la recherche en Sciences humaines et sociales, de même que la diversification des vecteurs de l'information scientifique, font que le cycle de la crédibilité se présente dans ce domaine avec moins d'acuité. Les données n'en sont pas moins importantes, mais elles sont moins sollicitées en vue de la recherche de financements.

### La compétition, facteur de variation

Une approche des données de la recherche par le concept de valeur d'usage permet de comprendre que plus elles détiennent un potentiel de publication, plus elles sont convoitées et plus leur valeur d'usage augmente. Elles entrent donc directement dans la dynamique de compétition scientifique qui caractérise la science moderne (Bucchi, 2015). Cette dynamique s'observe de manière particulièrement exacerbée dans les domaines des Sciences et Techniques et des Sciences de la vie. Les chercheurs, les équipes et les universités peuvent être en concurrence. Le secret devient alors une des règles du jeu de la compétition et de la course à la publication. Il y a donc un réflexe spontané de protection des données : celles-ci doivent rester confidentielles, tant que l'article n'a pas été publié.

*« En colloque, la politique est de ne montrer que des choses déjà publiées ou déjà sous forme d'article qu'on va soumettre. On a pris le parti pris de ne pas présenter des choses non publiées. Parce que, dans les colloques, il y a forcément nos compétiteurs. Il suffit qu'ils notent la bonne idée et qu'ils refassent la manip en deux temps, trois mouvements. Ils vont publier avant nous une histoire moins complète et, nous, on se sera fait squeezer » (chercheur 1).*

Le même mécanisme est à l'œuvre quand il s'agit de breveter une connaissance ou une technologie (Rebouillat, 2019b) :

*« Si on ne protège pas les données, on ne sera pas en mesure de les transférer à un industriel. Parce que l'industriel, ça ne l'intéresse pas si c'est public. Je pense à des entreprises qui veulent exploiter, faire de la licence sur ce qu'on fait ; eux, ils veulent des droits de licence pour être sûrs d'être en mesure d'exclure toute personne qui voudrait faire la même chose » (chercheur 35).*

## Autre facteur de variation, le stade de la carrière

La valeur d'usage des données peut varier également en fonction du stade d'avancement de la carrière du chercheur. Plus il sera jeune et en début de carrière, pratiquant sa recherche via des emplois précaires (CDD, post-doctorat, ATER...) et aspirant à décrocher un poste stable, plus le chercheur accordera une valeur importante à ses données. Pour les doctorants et jeunes chercheurs, les données constituent un matériau à partir duquel ils peuvent envisager une stratégie de publication à même de leur fournir du capital symbolique. Ce dernier sera mobilisé auprès des pairs ou lors de concours de recrutement :

*« Je connais beaucoup de collègues allemands qui n'ont pas de position fixe. En Allemagne, les positions fixes sont très rares. [...] Je connais des gens qui ont 45 ans et qui n'ont pas encore de position fixe. Donc il y a un enjeu économique derrière, mais pour eux, pour leur propre vie. C'est-à-dire qu'il faut qu'ils publient un maximum. Du coup, ils sont un peu moins partageurs sur certaines données. On sent qu'il y a un enjeu personnel pour eux. Alors que quand on a une position permanente, on n'a plus forcément cet enjeu personnel » (chercheur 16).*

Cet aspect a été mis en évidence dans les études de Van den Eynden *et al.* (2016) et de Tenopir *et al.* (2015). Elles montrent que les jeunes chercheurs sont moins disposés à partager leurs données, car ils craignent de perdre des opportunités de publications futures en rendant leurs données disponibles. Des travaux relatifs aux stratégies de carrière des jeunes chercheurs ont également montré que des directeurs de thèse ou de post-doctorat demandaient explicitement (voire exigeaient) que les données ne soient pas partagées (Nicholas *et al.*, 2017).

## VALEUR D'ÉCHANGE, OU LES PERSPECTIVES DE RECONFIGURATION DE LA VALEUR DES DONNÉES

La question de la valeur des données, abordée ci-dessus comme préalable à la publication, peut également être pensée en termes d'échange. L'étude montre que certaines pratiques naissantes autour des modalités de partage des données apportent une forme de rétribution, nous invitant à relativiser les conclusions de Fecher *et al.* (2017).

### Valeur des données avec la publication

Notre étude montre que la question de la valeur des données peut s'envisager en complément de l'article publié. Car, depuis les années 2010, de plus en plus de revues scientifiques demandent aux chercheurs de fournir les données associées au manuscrit soumis (Colavizza *et al.*, 2020). Les arguments qui président à cette démarche relèvent à la fois d'un souci d'intégrité scientifique et d'ouverture des données. L'accessibilité de ces dernières par les évaluateurs et les lecteurs joue un rôle d'administration de la preuve. Elle est perçue comme un gage de qualité, qui permet de limiter la fraude scientifique. La démarche est portée à la fois par des éditeurs *Open Access*, comme *Public Library of Science* (Plos) (Byrne, 2017), et par des éditeurs historiques, comme *SpringerNature*. Ce dernier a mené en 2017 une des plus vastes enquêtes internationales (7719 répondants) sur les pratiques de partage des données de la recherche (Stuart *et al.*, 2018). L'étude a montré que 63% des répondants partageaient les données associées à leur publication dans des entrepôts de données et/ou comme matériel supplémentaire à l'article.

Notre enquête de terrain a permis d'observer cette pratique. Sur les 46 chercheurs auxquels nous avons posé la question, 28 d'entre eux ont déclaré partager les données associées à leur publication en réponse à la politique de la revue (le tableau en [annexe 2](#) détaille la

répartition par laboratoire). Nous précisons que cette pratique était moins présente en Sciences humaines et sociales (5 chercheurs sur 14), car les revues dans ces domaines sont pour le moment moins prescriptives.

Lorsque les chercheurs répondent à la demande des revues, les données sont majoritairement partagées sous forme de fichiers liés à l'article, dans la section appelée « matériel supplémentaire » ou « *Supplementary Materials* ». Dans le panel, seuls les biologistes et les astronomes ont recours à des banques de données (ou entrepôts de données). Pour les astronomes, ce sont les observatoires (où sont situés les télescopes), qui se chargent de rendre les données disponibles (chaque observatoire dispose de sa propre archive de données). En Biologie, l'article n'est publié que si les données de génomique ont été déposées au préalable dans la banque de données *GenBank*.

Malgré leur réticence au partage, les chercheurs respectent cette clause d'ouverture. Un des chercheurs rencontrés explique qu'il n'est pas dans son intérêt de refuser la requête de l'éditeur. Pour lui, refuser c'est peut-être perdre l'opportunité d'être publié :

*« Pour moi, l'idée c'est que ma publication soit acceptée. Donc, si l'éditeur me demande de faire ça, je le fais. Quand on veut publier, on rend les choses fluides » (chercheur 41), [il s'agissait en l'occurrence de données génétiques, dont l'éditeur requiert le dépôt dans GenBank].*

Ce que nous montre notre analyse est que les chercheurs ne sont pas directement à l'initiative du partage. Si 28 des 57 participants à l'enquête disent mettre à disposition les données associées aux publications, il ne s'agit pas de pratiques systématiques. A l'exception des biologistes, cela est même relativement rare. Ce constat permet de nuancer à la fois les résultats de la vaste enquête de l'éditeur *SpringerNature*, mais aussi ceux de notre terrain.

Les pratiques des chercheurs sont donc conditionnées par la politique de la revue, à laquelle ils ont soumis leur manuscrit. On comprend que les éditeurs ont une forte influence sur les pratiques des communautés scientifiques, dès lors qu'ils détiennent les revues, lieux de légitimation dans le système d'évaluation de la recherche. Cela fait écho aux résultats de l'enquête réalisée par Nicholas *et al.* : plusieurs des jeunes chercheurs interrogés déclarent en effet ne mettre leurs données à disposition que si la revue dans laquelle ils prévoient de publier le leur demande (Nicholas *et al.*, 2017, p.214).

## La valeur sociale des données

Le temps et le travail consentis autour des données peut être visible et reconnu de manière inattendue chez les chercheurs qui en font l'expérience. Nous avons eu l'occasion de l'observer sur notre terrain auprès d'un des chercheurs, celui-ci ayant acquis de la visibilité et de la reconnaissance pour son expertise autour des données qu'il a partagées.

Comme il l'explique dans le verbatim ci-dessous, il a vécu au cours de sa formation doctorale une expérience de partage de données avec un autre doctorant, qui travaillait sur le même sujet que lui et avec lequel il a préféré entrer dans un rapport de collaboration (plutôt que de compétition). En rendant ses données accessibles, il a apporté de la visibilité à son travail, jouant ainsi la carte de la reconnaissance sociale. Celle-ci s'est révélée gagnante, car elle lui a permis d'augmenter ses chances d'être connu et sollicité par ses pairs :

*« Ça n'est pas juste partager les données pour partager les données. Il y a toute une mentalité derrière, qui est concomitante. La modélisation, que j'ai faite durant ma thèse sur les primates, a ensuite été réutilisée par d'autres personnes, qui ont utilisé mes formules chez les humains et qui m'ont posé des questions auxquelles j'ai répondu. Du coup, c'est intéressant, parce que vous n'êtes pas forcément co-auteur de ces publications mais vous êtes cités. Et ensuite les personnes vous connaissent. Ça veut dire que, quand elles ont besoin d'organiser un symposium, elles vont faire appel à vous. Donc vous n'êtes pas récompensés tout de suite, mais un peu plus tard » (chercheur 16).*

Cet exemple unique dans les entretiens menés mérite d'être traité et développé. Dans ce cas précis, le partage des données a joué un rôle de levier vers une forme de reconnaissance sociale, à travers l'identification d'opportunités de collaboration. Cette démarche, qui prend le contrepied de la compétition, rejoint de la même manière les principes de la science ouverte. Dans cette analyse, il est intéressant de souligner que le chercheur en question n'est pas un militant de la science ouverte et qu'il n'en mobilise pas les principes dans ses propos. Ayant une double formation disciplinaire, dont celle de l'Informatique, il a acquis une sensibilité aux principes de l'*Open Source*, qu'il estime transposables à son domaine de recherche – l'Éthologie. Ce faisant, il adopte aujourd'hui les pratiques de partage de données qu'il a découvertes pendant sa thèse, dans le but de développer de nouvelles collaborations scientifiques.

Ce cas particulier apporte un éclairage sur les politiques de science ouverte, qui portent attention à la question de la reconnaissance des efforts des chercheurs dans le partage de leurs données, considérant ces efforts (et le temps corollaire investi) comme un travail scientifique (MESRI, 2018). Cette tendance politique tend à s'accompagner d'indicateurs alternatifs aux indicateurs traditionnels de citation (comme le facteur d'impact) et positionnés sur le partage des données (*Déclaration de San Francisco sur l'évaluation de la recherche*, 2012). Le groupe de travail « Évaluation » du Comité pour la Science Ouverte (CoSO) a justement fourni des recommandations pour une légitimation et une reconnaissance de la gamme complète des activités de recherche (COSO, 2019). En Europe, on assiste aujourd'hui aux prémices de cette légitimation, avec des annonces récentes d'universités engagées dans la reconnaissance de différents profils de chercheurs, avec des activités de recherche variées, dans l'évaluation de leur carrière (VSNU *et al.*, 2019).

## CONCLUSION

Le présent article a investi la question du partage des données de la recherche à partir de l'angle inédit des régimes de valeur associés aux données. Cet angle permet de se positionner en amont de l'activité de partage, afin de traiter d'aspects souvent écartés des débats et des discussions scientifiques. L'approche choisie permet alors de mieux appréhender les freins et les résistances des chercheurs aux prescriptions politiques de partage des données.

L'article a montré que les trois régimes de valeur (valeur pour soi, valeur d'usage et valeur d'échange) sont certes distincts, mais néanmoins articulés les uns aux autres. Le régime qui agit le plus fortement sur les pratiques des chercheurs est celui de la valeur d'usage, qui est étroitement liée à la publication scientifique et donc à la reconnaissance symbolique. C'est la valeur la plus rétributrice pour le chercheur, celui-ci étant soumis à une évaluation de son travail et à des contraintes de gestion de temps. Dans cette perspective, la valeur des données est conditionnée par une forme d'assujettissement à la valeur de la publication.

Les politiques favorables au partage des données sont donc confrontées à l'impensé de l'évaluation de la recherche, encore centrée sur l'article. Elles ne prennent en compte ni la valeur pour soi, ni la valeur d'usage des données, qui sont des valeurs fondées sur l'investissement du chercheur et le rôle de preuve scientifique des données, justifiant le fait même qu'elles aient été collectées et conservées et les légitimant comme faisant partie intégrante du raisonnement scientifique. On ne peut donc pas s'attendre à ce que les communautés de recherche se mobilisent autour du partage des données, tant que les politiques de science ouverte ne prendront pas en considération les logiques qui structurent la valeur des données au sein des réseaux scientifiques. L'enjeu actuel autour du partage des données se joue donc sur les ajustements qui pourront se faire entre les différents régimes de

valeur et les nouvelles formes d'évaluation de la recherche. L'exemple du chercheur en Éthologie renforce ce constat et permet de dégager quelques perspectives d'évolution des formes de reconnaissance symbolique, fondées sur les données.

C'est pourquoi, en prolongement de ce travail, il nous paraît pertinent de redéployer les questions de recherche mobilisées ici au regard de l'émergence d'un nouvel environnement plus favorable, grâce à l'évolution des politiques de partage de données et des politiques d'évaluation de la recherche, qui tendent à prendre en compte une plus large diversité des formes de productions scientifiques. Nous étudierons ainsi dans quelle mesure ce nouvel environnement prend en considération les différents régimes de valeur de la donnée (valeur pour soi, valeur d'usage, valeur d'échange) et analyserons s'il agit sur les pratiques scientifiques.

## RÉFÉRENCES BIBLIOGRAPHIQUES

Andreoli-Versbach, P. ; Mueller-Langer, F. (2014), Open access to data : An ideal professed but not practised, *Research Policy*, vol. 43, n°9, p. 1621-1633, <https://doi.org/10.1016/j.respol.2014.04.008>

Bates, M. J. (2005), Information and knowledge : An evolutionary framework for information science. *Information Research*, 10(4). <http://informationr.net/ir/10-4/paper239.html>

Borgman, C. L. (2015), *Big data, little data, no data : Scholarship in the networked world*, The MIT Press.

Bourdieu, P. (1975), La spécificité du champ scientifique et les conditions sociales du progrès de la raison, *Sociologie et sociétés*, vol. 7, n° 1, p. 91-118, <https://doi.org/10.7202/001089ar>

Broudoux, E. ; Chartron, G. (Éds.) (2015), *Big Data—Open Data : Quelles valeurs ? Quels enjeux ? Actes du colloque « Document numérique et société »*, Rabat, 2015, De Boeck Supérieur. <https://doi.org/10.3917/dbu.chron.2015.01>

Bucchi, M. (2015), Norms, competition and visibility in contemporary science : The legacy of Robert K. Merton, *Journal of Classical Sociology*, vol. 15, n° 3, p. 233-252, <https://doi.org/10.1177/1468795X14558766>

Bullich, V. ; Clavier, V. (2018), Production des données, « Production de la société ». Les Big Data et algorithmes au regard des Sciences de l'information et de la communication, *Les Enjeux de l'information et de la communication*, vol. 19, n° 2, p. 5-14, <https://doi.org/10.3917/enic.025.0005>

Byrne, M. (2017), Making Progress Toward Open Data : Reflections on Data Sharing at PLOS ONE, *PLOS Blogs*, <https://everyone.plos.org/2017/05/08/making-progress-toward-open-data/>

Cabrera, F. (2014), *Les données de la recherche en sciences humaines et sociales : Enjeux et pratiques. Enquête exploratoire*, mémoire d'études de l'INTD, Cnam, [http://memsic.ccsd.cnrs.fr/mem\\_01128394/document](http://memsic.ccsd.cnrs.fr/mem_01128394/document)

Chartron, G. ; Schöpfel, J. (2017), Open access et Open science en débat. *Revue française des sciences de l'information et de la communication*, n° 11, <https://doi.org/10.4000/rfsic.3331>

Colavizza, G. ; Hrynaszkiewicz, I. ; Staden, I. ; Whitaker, K. ; McGillivray, B. (2020), The citation advantage of linking publications to research data, *PLOS ONE*, vol. 15, n° 4, e0230416, <https://doi.org/10.1371/journal.pone.0230416>

Collectif P.é.c.r.e.s. (2011), *Recherche précarisée, recherche atomisée. Production et transmission des savoirs à l'heure de la précarisation*, Éditions Raisons d'Agir.

COSO - Comité pour la Science Ouverte (2019), *Types de documents, productions et activités valorisées par la science ouverte et éligibles à une évaluation*, <https://www.ouvrirlascience.fr/types-de-documents-productions-et-activites-valorisees-par-la-science-ouverte-et-eligibles-a-une-evaluation>

CE - Commission européenne. (2012). Communication de la Commission au Parlement européen, au Conseil, au Comité économique et social européen et au comité des régions. Pour un meilleur accès aux informations scientifiques : Dynamiser les avantages des investissements publics dans le domaine de la recherche. <https://eur-lex.europa.eu/legal-content/FR/TXT/?uri=CELEX:52012DC0401>

Da Costa, M. P. ; Lima Leite, F. C. (2019), Factors influencing research data communication on Zika virus : A grounded theory, *Journal of Documentation*, vol. 75, n°5, p.910-926, <https://doi.org/10.1108/JD-05-2018-0071>

*Déclaration de San Francisco sur l'évaluation de la recherche* (2012), <https://sfdora.org/read/>

Fecher, B. ; Friesike, S. ; Hebing, M. ; Linek, S. (2017), A reputation economy : How individual reward considerations trump systemic arguments for open access to data, *Palgrave Communications*, vol. 3, n° 17051, <https://doi.org/10.1057/palcomms.2017.51>

Fecher, B. ; Friesike, S. ; Hebing, M. ; Linek, S. ; Sauer mann, A. (2015), A Reputation Economy : Results from an Empirical Survey on Academic Data Sharing, *DIW Berlin Discussion Paper*, n° 1454, <http://dx.doi.org/10.2139/ssrn.2568693>

Foray, D. ; Lundvall, B.-A. (1997), Une introduction à l'économie fondée sur la connaissance, in B. Guilhon, P. Huard, M. Orillard, & J.-B. Zimmermann (Éds.), *Économie de la connaissance et des organisations. Entreprises, territoires, réseaux*, p. 16-37, L'Harmattan.

Garvey, W. D. (1979), *Communication : The Essence of Science. Facilitating information exchange among librarians, scientists, engineers and students*, Pergamon Press, <https://doi.org/10.1016/C2013-0-05843-4>

Gingras, Y. (2018), Les transformations de la production du savoir : De l'unité de connaissance à l'unité comptable, *Zilsel*, vol. 4, n° 2, p. 139-152, <https://doi.org/10.3917/zil.004.0139>

Homolak, J. ; Kodvanj, I. ; Virag, D. (2020), Preliminary analysis of COVID-19 academic information patterns : A call for open science in the times of closed borders, *Scientometrics*, vol. 124, n° 3, p. 2687-2701, <https://doi.org/10.1007/s11192-020-03587-2>

Latour, B. (2001), *Le métier de chercheur : Regard d'un anthropologue. Une conférence-débat à l'INRA*, INRA Editions.

Latour, B. ; Woolgar, S. (1979), *La vie de laboratoire : La production des faits scientifiques*. La Découverte.

Leonelli, S. (2013), Why the Current Insistence on Open Access to Scientific Data? Big Data, Knowledge Production, and the Political Economy of Contemporary Biology, *Bulletin of Science, Technology and Society*, vol. 33, n° 1-2, p. 6-11, <https://doi.org/10.1177/0270467613496768>

Leonelli, S. (2015), What Counts as Scientific Data ? A Relational Framework, *Philosophy of Science*, vol. 82, n° 5, p. 810-821, <https://doi.org/10.1086/684083>

Levin, N. ; Leonelli, S. (2017), How Does One "Open" Science ? Questions of Value in Biological Research, *Science, Technology, & Human Values*, vol. 42, n° 2, p. 280-305. <https://doi.org/10.1177/0162243916672071>

- Mayère, A. (1990), *Pour une économie de l'information*, CNRS Editions, <https://doi.org/10.3917/cnrs.mayer.1990.01>
- Merton, R. K. (1973), *The Sociology of Science. Theoretical and Empirical Investigations*, University of Chicago Press.
- Miège, B. (2007), *La société conquise par la communication. Tome III : Les TIC entre innovation technique et ancrage social*, Presses Universitaires de Grenoble.
- Miège, B. (2020), *La numérisation en cours de la société. Points de repères et enjeux*, Presses Universitaires de Grenoble.
- Ministère de l'Enseignement Supérieur, de la Recherche et de l'Innovation (2018), *Plan national pour la science ouverte*, [http://cache.media.enseignementsup-recherche.gouv.fr/file/Actus/67/2/PLAN\\_NATIONAL\\_SCIENCE\\_OUVERTE\\_978672.pdf](http://cache.media.enseignementsup-recherche.gouv.fr/file/Actus/67/2/PLAN_NATIONAL_SCIENCE_OUVERTE_978672.pdf)
- National Science Board (2005), *Long-Lived Digital Data Collections : Enabling Research and Education in the 21st Century* (NSB-05-40), National Science Foundation, <https://www.nsf.gov/pubs/2005/nsb0540/>
- Nicholas, D. ; Rodríguez-Bravo, B. ; Watkinson, A. ; Boukacem-Zeghmouri, C. ; Herman, E. ; Xu, J. ; Abrizah, A. ; Świgoń, M. (2017), Early career researchers and their publishing and authorship practices, *Learned Publishing*, vol.30, n° 3, p.205-217, <https://doi.org/10.1002/leap.1102>
- Nicholas, D. ; Watkinson, A. ; Boukacem-Zeghmouri, C. ; Rodríguez-Bravo, B. ; Xu, J. ; Abrizah, A. ; Świgoń, M. ; Herman, E. (2017), Early career researchers : Scholarly behaviour and the prospect of change, *Learned Publishing*, vol.30, n° 2, p. 157-166, <https://doi.org/10.1002/leap.1098>
- OCDE - Organisation de Coopération et de Développement Economiques (2004), *Déclaration sur l'accès aux données de la recherche financée par des fonds publics*, <https://legalinstruments.oecd.org/fr/instruments/157>
- OCDE - Organisation de Coopération et de Développement Economiques (2007), *Principes et lignes directrices de l'OCDE pour l'accès aux données de la recherche financée sur fonds publics*, Editions OCDE, <http://www.oecd.org/fr/science/sci-tech/38500823.pdf>
- Prost, H. ; Schöpfel, J. (2015), *Les données de la recherche en SHS. Une enquête à l'Université de Lille 3*, Rapport de recherche, Université Lille 3, <https://hal.univ-lille.fr/hal-01198379>
- Rebouillat, V. (2019a), *Ouverture des données de la recherche : De la vision politique aux pratiques des chercheurs*, thèse en Sciences de l'information et de la communication, Cnam, <https://tel.archives-ouvertes.fr/tel-02447653>
- Rebouillat, V. (2019b), Les données scientifiques face aux enjeux de la recherche en Sciences, Technologie et Médecine : Enquête exploratoire à l'Université de Strasbourg, *Etudes de communication*, vol. 52, n° 1, p. 99-116, <https://doi.org/10.4000/edc.8681>
- Serres, A. ; Malingre, M.-L. ; Mignon, M. ; Pierre, C. ; Collet, D. (2017), *Données de la recherche en SHS. Pratiques, représentations et attentes des chercheurs : Une enquête à l'Université Rennes 2*, Rapport de recherche, Université Rennes 2, <https://hal.archives-ouvertes.fr/hal-01635186>
- Stuart, D. ; Baynes, G. ; Hrynaszkiewicz, I. ; Allin, K. ; Penny, D. ; Lucraft, M. ; Astell, M. (2018), *Whitepaper : Practical challenges for researchers in data sharing*, figshare, <https://doi.org/10.6084/m9.figshare.5975011.v1>
- Tenopir, C. ; Allard, S. ; Douglass, K. ; Aydinoglu, A.U. ; Wu, L. ; Read, E. ; Manoff, M. ; Frame, M. (2011), Data Sharing by Scientists : Practices and Perceptions, *PLOS ONE*, vol. 6, n° 6, e21101. <https://doi.org/10.1371/journal.pone.0021101>

Tenopir, C. ; Dalton, E. D. ; Allard, S. ; Frame, M. ; Pjesivac, I. ; Birch, B. ; Pollock, D. ; Dorsett, K. (2015), Changes in Data Sharing and Data Reuse Practices and Perceptions among Scientists Worldwide, *PLOS ONE*, vol. 10, n° 8, <https://doi.org/10.1371/journal.pone.0134826>

Tenopir, C. ; Kaufman, J. ; Sandusky, R. ; Pollock, D. (2019), The Time Has Come... To Talk About Why Research Data Management Isn't Easy, *Proceedings of the Charleston Library Conference*, p. 298-301, <https://docs.lib.purdue.edu/charleston/2019/libraryservices/7/>

Van den Eynden, V. ; Knight, G. ; Vlad, A. ; Radler, B. ; Tenopir, C. ; Leon, D. ; Manista, F. ; Whitworth, J. ; Corti, L. (2016), *Towards Open Research : Practices, experiences, barriers and opportunities*, Wellcome Trust, <https://doi.org/10.6084/m9.figshare.4055448.v1>

VSNU, NFU, KNAW, NWO, & ZonMw (2019), *Room for everyone's talent towards a new balance in the recognition and rewards of academics*, <http://www.vsnul.nl/files/documenten/Domeinen/Onderzoek/Position%20paper%20Room%20for%20everyone%E2%80%99s%20talent.pdf>

Wallis, J. C. ; Rolando, E. ; Borgman, C. L. (2013), If We Share Data, Will Anyone Use Them ? Data Sharing and Reuse in the Long Tail of Science and Technology, *PLOS ONE*, vol. 8, n° 7, <https://doi.org/10.1371/journal.pone.0067332>

## REMERCIEMENTS

Je tiens à remercier Chérifa Boukacem-Zeghmouri (Professeur à l'Université Lyon 1) et Hans Dillaerts (Maître de conférences à l'Université Montpellier 3) pour leurs commentaires sur la première version de cet article. Je souhaite également remercier les 57 chercheurs et enseignants-chercheurs qui ont accepté de participer à l'enquête.

## ANNEXES

## Annexe 1 : Disciplines des 57 participants à l'enquête

Grand domaine scientifique	Laboratoire	Chercheur	Discipline	
Vie & Santé	Laboratoire 1	Chercheur 1	Génétique	
	Laboratoire 2	Chercheur 2	Chimie thérapeutique	
	Laboratoire 3	Chercheur 3	Biologie	
		Chercheur 4	Biologie	
		Chercheur 5	Biologie	
		Chercheur 6	Biophysique	
	Laboratoire 4	Chercheur 7	Biologie	
		Chercheur 8	Biologie	
		Chercheur 9	Biologie	
		Chercheur 10	Biologie	
		Chercheur 11	Chimie thérapeutique	
	Laboratoire 5	Chercheur 12	Médecine	
		Chercheur 13	Écologie	
		Chercheur 14	Écologie	
		Chercheur 15	Écologie	
	Laboratoire 6	Chercheur 16	Éthologie	
		Chercheur 17	Bioinformatique	
		Chercheur 18	Neurosciences	
		Chercheur 19	Neurosciences	
		Chercheur 20	Neurosciences	
	Sciences & Technologies	Laboratoire 7	Chercheur 21	Primatologie
			Chercheur 22	Neurosciences
Laboratoire 8			Chercheur 23	Astronomie
			Chercheur 24	Astronomie
			Chercheur 25	Astronomie
			Chercheur 26	Astronomie
			Chercheur 27	Astronomie
		Chercheur 28	Astronomie	
Chercheur 29		Astrophysique		
Laboratoire 9		Chercheur 30	Bioinformatique	
	Chercheur 31	Géographie		
	Chercheur 32	Informatique		
	Chercheur 33	Informatique		
	Chercheur 34	Informatique		
	Chercheur 35	Sciences de l'ingénieur		
Laboratoire 10	Chercheur 36	Sciences de l'ingénieur		
	Chercheur 37	Géologie		
	Chercheur 38	Géologie		
	Chercheur 39	Géologie		
		Chercheur 40	Géologie	

Sciences humaines et sociales	Laboratoire 11	Chercheur 41	Ecologie
		Chercheur 42	Géo-archéologie
		Chercheur 43	Géographie
		Chercheur 44	Géographie
		Chercheur 45	Géographie
	Laboratoire 12	Chercheur 46	Anthropologie
		Chercheur 47	Démographie
		Chercheur 48	Démographie
		Chercheur 49	Droit
		Chercheur 50	Histoire
		Chercheur 51	Sciences politiques
		Chercheur 52	Sociologie
		Chercheur 53	Sociologie
	Laboratoire 13	Chercheur 54	Sociologie
		Chercheur 55	Théologie
Chercheur 56		Théologie	
Chercheur 57		Théologie	

## Annexe 2 : Nombre de chercheurs du panel à avoir déjà partagé des données liées à une publication

Grand domaine scientifique	Laboratoire	Discipline	Nombre de chercheurs...	
			Rencontrés	Ayant déjà partagé des données liées à une publication
Vie & Santé	Laboratoire 1	Biologie	1	1
	Laboratoire 3	Chimie et Biologie	3	3
	Laboratoire 4	Chimie et Biologie	5	5
	Laboratoire 5	Ecologie	3	3
	Laboratoire 6	Neurosciences	4	4
	Laboratoire 7	Neurosciences	1	1
Sciences & Technologies	Laboratoire 8	Astronomie	7	2
	Laboratoire 9	Sciences de l'ingénieur	5	2
	Laboratoire 10	Sciences de la Terre	3	2
Sciences humaines et sociales	Laboratoire 11	Géographie	4	3
	Laboratoire 12	Sciences sociales	8	2
	Laboratoire 13	Théologie	2	0
			<b>46</b>	<b>28</b>