



HAL
open science

Sauvegarde et données de la recherche : quel tiers de confiance ?

Mariannig Le Béhec, Philippe Charrier, Gabriel Gallezot

► To cite this version:

Mariannig Le Béhec, Philippe Charrier, Gabriel Gallezot. Sauvegarde et données de la recherche : quel tiers de confiance ?. Annaïg Mahé; Ingrid Mayeur; Elsa Poupardin; Camille Prime-Claverie. Communication scientifique et science ouverte : opportunités, tensions et paradoxes, De Boeck supérieur, pp.173-182, 2023, Information et stratégie, 978-2-8073-5710-5. hal-04481880

HAL Id: hal-04481880

<https://hal.science/hal-04481880v1>

Submitted on 28 Feb 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - ShareAlike 4.0 International License

Version auteur du texte évalué :

Le Béhec, M., Charrier P., Gallezot G. (2023) Sauvegarde et données de la recherche : quel tiers de confiance ? in « Document numérique et société » Liège 2022. Communication scientifique et science ouverte : Opportunités, tensions et paradoxes, Louvain-la-Neuve.: deboeck supérieur, adbs (Information et stratégie), pp. 173–182.

Sauvegarde et données de la recherche : quel tiers de confiance ?

Mariannig le Béhec, Université Claude Bernard Lyon 1, UR ELICO (4147)
Philippe Charrier, Centre Max Weber, UMR 5283 (CNRS/ Université Lyon 2)
Gallezot Gabriel, Université Côte d’Azur, //TransitionS.

Résumé : Les données de la recherche sont un des enjeux des politiques publiques de la science ouverte (PNSO 2, 2021). Dans le cycle de la gestion des données, chaque étape est dépendante l’une de l’autre et stocker des flux de données dans une infrastructure de recherche favorise la mémoire des résultats de la recherche à des fins de réutilisation ou de patrimonialisation. Des infrastructures de recherche (Star, 1999), institutionnelles, existent dans les communautés scientifiques qui se sont mobilisées pour ouvrir leurs données et assurer leur réutilisation depuis les années 1970 par exemple en astrophysique, en sismologie ou en génomique (Denis et Goëta, 2017). Le développement des outils informatiques individuels depuis les années 1980 et le mouvement de la science ouverte tendent à répandre ces pratiques à toutes les communautés épistémiques (Meyer et Molyneux-Hodgson, 2011). Face à une « organisation matérialisée (institutions et politiques mémorielles) » (Merzeau, 2006), quels sont les usages ordinaires (Certeau, 2010) des chercheurs ? Quelle est l’appropriation personnelle (Akrich, 2006, Amiel et al., 2020) de ces outils et de ces enjeux de stockage des données de la recherche ? Les données de recherche et leur gestion se lient-elles à des valeurs (Rebouillat, 2021), à la confiance (Luhmann, 2001) ? La communication des données de la recherche est un pilier de la science ouverte mais une part importante des chercheurs interrogés en France ne développent pas de stratégie à long terme de sauvegarde des données de recherche. Nous soulignons des disparités socio-démographiques, selon le contexte de travail et l’attachement personnel aux données.

1 Introduction

Les données de la recherche sont un des enjeux des politiques publiques en matière de science ouverte et sont mentionnées dans le deuxième Plan national pour la science ouverte, paru en 2021 (PNSO 2, 2021). Dans les modèles vertueux tels que le cycle communément admis de la gestion des données, chaque étape est dépendante l’une de l’autre pour permettre la diffusion des données *in fine*. Pourtant les discours institutionnels peuvent être contradictoires puisque la Commission européenne indique « Aussi ouvert que possible mais aussi fermé que nécessaire » selon la recommandation (UE) 2018/790 de la Commission européenne du 25 avril 2018, relative à l’accès aux informations scientifiques et à leur conservation.

En dépassant les questions de propriété intellectuelle, d’anonymisation des données et réglementaires liées au règlement général sur la protection des données (RGPD) (Desfriches-Doria, 2019) ou à la loi relative aux recherches impliquant la personne humaine (2012), stocker des flux de données dans une infrastructure de recherche favorise la mémoire des résultats de la recherche à des fins de réutilisation et de patrimonialisation. Les enjeux s’associent également à des questions dites de souveraineté et d’innovation. Il est également possible de trouver des discours promoteurs de la transparence et de la confiance des citoyens dans la science grâce à l’ouverture des données de recherche.

Précédemment, les chercheurs ont déjà mis en place des infrastructures de recherche pour répondre à ces enjeux. Les infrastructures de recherche (Star, 1999), institutionnelles, existent dans les

communautés scientifiques qui se sont mobilisées pour ouvrir leurs données et assurer leur réutilisation, leur accessibilité depuis les années 1970, par exemple en astrophysique, en sismologie ou en génomique (Denis et Goëta, 2017). Puis le développement des outils informatiques individuels depuis les années 1980 et l'emprise institutionnelle du mouvement de la science ouverte tendent à répandre ces pratiques, qualifiées parfois de « bonnes pratiques » à toutes les communautés épistémiques (Meyer et Molyneux-Hodgson, 2011).

Face à une « organisation matérialisée (institutions et politiques mémorielles) » (Merzeau, 2006), notre cadre théorique interroge les usages ordinaires (Certeau, 1990) des chercheuses et chercheurs. Quelle est l'appropriation personnelle (Akrich, 2006) des outils principalement numériques et des enjeux de stockage des données de la recherche ? De précédentes recherches développent ces pistes (Amiel et al., 2020). Par outils numériques, nous définissons à la fois des services numériques, des infrastructures de recherche ou des plateformes commerciales. La question se pose de savoir comment les chercheurs peuvent « faire avec » (Certeau, 1990) quand ils disposent de ressources pour sauvegarder des données de recherche mais également quand ils n'en disposent pas. De plus, l'attachement personnel aux données de recherche produites peut diverger selon que la donnée est produite par un instrument de mesure ou obtenue dans le cadre d'entretien avec des personnes lors d'une collecte de terrain. Les données de recherche et leur gestion se lieraient à des valeurs (Rebouillat, 2021). Selon nous, elle se lie également à une forme de confiance « assumée » et « décidée » (Luhmann, 2001) par les chercheuses et chercheurs, selon des formes que nous développerons à partir de la proxémie chez Hall (1990).

Nous constatons que la diffusion des données de la recherche est un pilier de la science ouverte mais une part importante des chercheurs interrogés en France ne développent pas de stratégie de conservation et de valorisation des données de recherche. En 2020, nous avons conduit une enquête à l'échelle nationale en France sur les pratiques numériques et la gestion des données au sein des communautés scientifiques. Nous soulignons qu'au-delà des usages des outils et services informatiques, il existe des disparités sociodémographiques et liées au contexte de travail.

2. Méthodologie de l'enquête Pratiques et Usages des outils numériques et des données de la recherche en France (2022)

Pour documenter les réponses à nos questions, nous nous appuyons sur les résultats de l'enquête SOSP_FR - Pratiques et Usages des outils numériques et des données de la recherche en France (Le Béhec et al., 2022).

La méthode mobilisée est mixte dans le sens où elle mêle une étude quantitative par questionnaire auprès d'un échantillon de chercheurs travaillant sur le territoire français, toutes disciplines confondues, et des entretiens semi-directifs auprès d'une sélection de répondants volontaires ayant manifesté leur intérêt de participer à cette seconde phase de l'enquête.

La première partie de l'enquête correspond à la réalisation et la passation d'un questionnaire de 38 questions qui a recueilli 1089 réponses complètes. Il s'organise en 9 chapitres, suivant une logique la plus proche possible du déroulement de l'activité de recherche. Nous mobilisons ici qu'une partie de cette enquête, les questions ayant pour objectif de cerner l'environnement de recherche du répondant, ses pratiques en matière de sauvegardes et d'archivage des données de la recherche et le recueil de données sociodémographiques. Le formulaire a été transmis de manière dématérialisée via une série de réseaux professionnels ou institutionnels le plus large possible (réseau d'unité de formation à l'information scientifique et technique, réseaux professionnels de l'information scientifique et technique, réseaux siconumériques (Twitter et LinkedIn), relais des universités ou organismes de recherche et sollicitations d'entreprises privées afin d'atteindre une grande diversité de disciplines scientifiques et de profils de chercheurs. La passation du questionnaire a été ouverte du 17 juin au 30 septembre 2020 via la plateforme SphynxOnline ; les données ont été sauvegardées sur la plateforme du TGIR Human-Num (CNRS). La seconde méthode est qualitative. Nous avons ainsi réalisé 30 entretiens semi-directifs. Nous faisons le choix de ne pas présenter dans ce travail, car il fait l'objet d'une autre analyse.

Sur la thématique de patrimonialisation, voire de la capitalisation des données de recherche, nous

admettons ici que nous n'avons pas respecté les nomenclatures habituellement développées par les professionnels de l'information pour essayer de ne pas nous éloigner des pratiques des chercheurs (Sbeih et al., 2020). Les données de recherche sont des archives dans la définition qu'en donnent les autrices et auteurs selon le code du patrimoine. Leur article amène également une précision sur ce qu'est le stockage, la sauvegarde, la pérennisation et l'archivage.

« Le stockage est l'action d'enregistrer des informations sur un support pour les rendre accessibles. La sauvegarde permet de dupliquer des informations pour prévenir toute disparition accidentelle. La pérennisation (ou préservation) permet de faire face à la perte d'informations d'identification ainsi qu'à l'obsolescence des supports et des logiciels. Elle consiste en effet à identifier et à conserver des documents et des données pour les rendre accessibles sur le moyen (10 ans et plus) et le long terme (50 ans et plus). L'archivage, qu'il concerne des supports électroniques ou papier, est une démarche qui vise à définir et mettre en œuvre des méthodes et des outils pour gérer et conserver des documents et des données. Le but est de rendre ceux-ci disponibles et accessibles pour répondre aux besoins métiers et scientifiques ainsi qu'aux obligations légales et sauvegarder les documents et données ayant un intérêt patrimonial. L'archivage détermine également les responsabilités de différents acteurs. » (Sbeih et al., 2020).

Ces définitions et cette distinction terminologique demeurent souvent méconnues des membres des communautés scientifiques. Il nous apparaissait donc plus aisé d'essayer de reprendre un terme générique d'archivage défini comme une « conservation à long terme », comprenant ainsi la question des données froides et incluant les questions d'archivage et de pérennisation. Mais nous verrons dans les résultats que cette distinction n'est pas complètement opérante dans le processus de recherche. Nous comprenons donc que cette terminologie n'est pas juste, mais elle nous permet de comprendre les usages et pratiques des communautés scientifiques, sans risque de perdre les répondants sur un questionnaire qui demandait plus de 20 minutes de temps de réponse. Notre analyse porte donc sur deux questions : 15 - Archivez-vous (conservation à long terme) ces données ? et 15 bis, si oui à quel tiers les confiez-vous ? Ces questions venaient suite à la question d'usage « Pour enregistrer et sauvegarder vos données, quels outils utilisez-vous ? » qui via huit items vise à qualifier des usages [questionnaire disponible via <https://10.5281/zenodo.3935958>].

3. Les pratiques entourant l'archivage des données de recherche

Au-delà de la question souvent restreinte de la diffusion des données de recherche que nous pouvons qualifier quantitativement, les résultats de l'enquête SOSP_FR montrent que les pratiques d'archivage des données subissent de multiples variations selon les caractéristiques sociodémographiques du répondant alors qu'un chercheur sur deux développe cette pratique. Les chercheurs qui déclarent une pratique d'archivage comme conservation à long terme (Q15) ont été catégorisés selon d'une part leur adhésion aux principes de la science ouverte et d'autre part, des variables socio-démographiques (statut, fonction, classe d'âge) et professionnelles (employeur, composition de l'équipe de recherche). Nous avons appréhendé les pratiques d'archivage des données par l'intermédiaire d'une question qui portait sur une conservation à long terme, sachant que celle-ci paraît indispensable pour que des pratiques de diffusion des données se concrétisent à une grande échelle. Dans notre enquête 52,9 % des répondants déclarent pratiquer ce type d'archivage, 46,6 % ne pas le faire (0,6 % ne souhaitant pas répondre à la question). Nous pouvons considérer que le poids de cette pratique concerne donc un peu plus de la moitié des chercheurs.

La pratique n'est pas plus développée chez les adhérents aux principes de la science ouverte, sachant qu'il est possible de supposer que l'ensemble des répondants avaient une sensibilité vis-à-vis de la science ouverte. Nous avons répertorié et sélectionné ces répondants en fonction de leur positionnement sur certains items du questionnaire (n=172) que sont la diffusion en accès libre des publications par choix personnel, la diffusion des publications sur des réseaux sociaux numériques et un positionnement favorable à la diffusion large des données de la recherche. Ainsi, cette sous-population pratique la conservation à long terme à 51,2 %, soit une proportion proche et même légèrement en dessous des pratiques déclarées dans l'échantillon. D'autres variables viennent caractériser les pratiques, notamment les communautés épistémiques, au sens de pratiques partagées qui dépassent les

catégorisations par discipline, le statut et la classe d'âge.

Toutefois, nos résultats montrent que la pratique d'archivage ne se répartit pas uniformément parmi les chercheuses et chercheurs. L'hypothèse la plus évidente est de considérer que les personnes les plus sensibles aux thèmes de la science ouverte sont nettement plus enclins à la pratiquer, si nous la définissons ici comme un aboutissement de l'application des principes FAIR, soit des données découvrables, accessibles, interopérables, réutilisables.

À la lecture des résultats, la relation de dépendance n'est pas évidente. Par exemple, parmi les chercheurs déclarant être favorable à une large diffusion des données de recherche, au-delà même du cercle universitaire et académique, seuls 51,4 % d'entre eux pratiquent un archivage, soit une proportion très proche de la moyenne des répondants.

Il faut donc croiser les réponses liées à la pratique d'archivage avec d'autres variables pour déceler l'influence de variables explicatives. Ainsi, nous avons constaté que les résultats subissent de multiples variations selon les caractéristiques sociodémographiques, au point qu'il nous paraît indispensable de les prendre en considération, afin de saisir comment se construisent les pratiques d'archivage.

Par variables sociodémographiques, nous entendons dans notre enquête le statut, la fonction dans la recherche, la classe d'âge, de même que le contexte de travail, identifié par le nombre de personnes avec lequel l'activité est réalisée permettant de distinguer collectif et exercice solitaire (Le Béhec et al. 2022).

L'élément le plus saillant tient au contexte de travail. Ainsi, réaliser son activité au sein d'un collectif de plus de 20 personnes influe favorablement sur la pratique d'archivage. Le développement d'infrastructures de recherche ou leur accessibilité dans ce contexte de travail semblent avoir un effet d'entraînement sur cette pratique, au point que le recours à l'archivage se produit dans 8 cas sur 10 lorsque le répondant travaille dans un contexte regroupant plus de 20 personnes, alors qu'il n'est le fait que de 48,3 % pour les personnes travaillant seules.

Les résultats à propos du croisement avec la fonction exercée par le répondant sont également discriminants. Les chercheurs ayant des responsabilités, comme les professeurs d'université et les directeurs de recherches, sont plus sensibles à la question de l'archivage. Ils sont respectivement 61 % et 69 % d'entre eux à déclarer cette pratique, ce que font bien moins les doctorants ou les ingénieurs de recherche (41 et 49 %).

Le statut joue un rôle tout aussi important, avec une séparation assez nette entre les chercheurs bénéficiant d'un emploi stable et permanent où le positionnement en termes d'archivage à long terme est plus favorable que la moyenne des répondants (58,7 % de réponse positive), alors que les chercheurs non-permanents sont moins nombreux à le faire (49,3 %). Par non-permanents nous avons qualifié les répondants qui ont des contrats doctoraux, post-doctoraux et à durée déterminée.

La variable de la classe d'âge n'est pas en reste, même si elle semble minorer le poids de la variable statutaire. Car plus le répondant fait partie des jeunes générations de chercheurs, plus sa pratique en matière d'archivage s'affaiblit, au point qu'elle finit par ne concerner que 37,1 % de ceux nés après 1995 dont les trois quarts sont des doctorants. À l'inverse, cette pratique devient majoritaire à partir des générations nées avant 1975.

Parallèlement à ces résultats, pour qu'une pratique soit envisagée par un acteur, elle doit également disposer d'un espace jugé propice à sa réalisation. Ainsi, la sous-question permet de comprendre à quel tiers les chercheuses et chercheurs confient la conservation à long terme des données de recherche. Les réponses étaient libres ; nous travaillons ici à partir d'un recodage des réponses. Les résultats viennent renforcer le constat que les caractéristiques socio-démographiques des chercheuses et chercheurs influent sur le choix du tiers mobilisé pour l'archivage. Nous qualifions ce tiers de tiers de confiance selon la distinction que Luhmann (2001) opère entre confiance assumée et confiance décidée, la première résidant dans un environnement permettant la confiance, la seconde qualifiant le choix, en fonction notamment de facteurs de risques.

À partir des réponses des 479 personnes qui ont répondu « oui » à la question sur la pratique de l'archivage des données et renseigné la question du tiers, quatre types de tiers de confiance se distinguent. Le premier est un tiers d'archivage qui n'en est pas véritablement un, dans le sens où c'est le chercheur, plus précisément son matériel, qui est le lieu de sauvegarde : clé USB, ordinateur, disque

dur externe, voir *network access server* (NAS), etc. ; 30,5 % des répondants à la question sont dans ce registre. Ils sont probablement plus nombreux si l'on ajoute les 62 répondants qui expliquent ne pas recourir à un tiers et nous pouvons supposer que ces personnes utilisent du matériel personnel. L'ensemble représente 43,1 % des répondants.

Vient ensuite le deuxième tiers d'archivage qui correspond aux solutions institutionnelles proposées par les universités ou les grands organismes de recherche : 34 %. Il faudrait peut-être y ajouter les répondants qui déclarent se tourner vers les solutions proposées plus localement, par un réseau interne au laboratoire, mais également par un système de sauvegarde collaboratif : 17,7 %. Ainsi, l'utilisation de solutions institutionnelles concerne 51,7 % des répondants. La part des recours à d'autres solutions, externes aux institutions ou à son environnement de travail (cloud non institutionnel...), s'élèvent à 11 % pour le troisième tiers d'archivage.

Enfin, une pratique assez marginale mais tout de même notable consiste à recourir aux services de son entourage familial ou amical (2,4 %). Même marginaux, ces résultats demeurent inattendus comme quatrième tiers de confiance.

Les types de tiers de confiance ne se répartissent pas harmonieusement en fonction des caractéristiques des répondants. À grand trait, nous retiendrons que l'usage d'archivage dans des ressources informatiques institutionnelles est nettement plus pratiqué par les chercheuses et les chercheurs non-permanents et stables, mais bien moins par les travailleurs indépendants et surtout les doctorants et les post-doctorants. Dans le même temps, les universitaires font moins appel aux ressources informatiques institutionnelles et optent plus souvent pour une conservation personnelle ou sans tiers : c'est le cas de 36,2 % des réponses d'universitaires. Les doctorants et post-doctorants ajoutent une appétence particulière supplémentaire pour les solutions proposées par des sociétés privées et commerciales (20,7 %) ou les conservations personnelles (39,1 %). D'une manière générale, les résultats montrent que les ressources institutionnelles sont sous-utilisées par les répondants.

À ces distinctions, il faut en ajouter l'opposition entre travail en solitaire et travail en collectif. Plus le chercheur est en situation de travail solitaire, plus sa pratique d'archivage (lorsqu'elle existe) repose sur une conservation personnelle avec 47,5 % des personnes qui travaillent seules qui archivent avec ou sur du matériel informatique personnel. À l'inverse, les chercheuses et chercheurs qui travaillent dans les collectifs ont nettement plus tendance à se tourner vers les solutions institutionnelles ou proposées par leur structure de proximité, comme un laboratoire. Ainsi, parmi les répondants qui archivent les données de recherche et qui travaillent dans des groupes de 6 à 10 personnes, 55,8 % d'entre choisissent l'un ou l'autre de ces tiers. Il devient alors compréhensible d'explicitier cette expression d'attachement personnelle « mes données de recherche » alors même que les données de recherche, comme précédemment indiqué, dans le cadre de la recherche publique, sont des archives selon le code du patrimoine (Sbeih et al., 2020).

4. Les pratiques d'archivage : équipement informatique personnel et tiers de confiance

Les résultats présentés laissent entrevoir plusieurs hypothèses qui viendraient expliquer des liens de causalité observés.

La première tient au poids de la proxémie (Hall, 1990) dans l'exécution des pratiques d'archivage. La proxémie peut être définie comme un modèle d'anthropologie de l'espace, de sphères de proximité avec notamment la distinction des distances intimes, personnelles, sociales et publiques. Elle permet de mettre en évidence des dimensions culturelles. Il est plus aisé de faire rentrer un proche collègue qu'un pair partenaire dans une bulle d'usage ou de pratique.

Joachim Schöpfel (2018 : 18), dans une étude qualitative, tente de cerner les dimensions qui interviennent dans le choix des modes d'archivage des chercheurs. Le premier choix est la mise en forme des données archivées qui peut conduire à des solutions hétérogènes, le deuxième est la finalité des données produites et enfin la troisième est le niveau d'adéquation entre les solutions concrètement utilisées et les solutions espérées. Toutefois l'analyse par les usages omet l'hétérogénéité des personnes répondantes. Dans les analyses d'usages, les supports de sauvegarde peuvent différer mais le chercheur répondant demeure unique et indivisible. En effet des groupes de recherche (ou des

« clusters ») peuvent être aisément identifiés via l'étude de leurs usages, mais cette perspective occulte l'hétérogénéité intrinsèque de statut et de fonction des membres dans la recherche publique française et leurs détournements possibles au sens de « braconnage » chez de ? Certeau (2010). Comme nous avons pu le découvrir via nos résultats, les caractéristiques sociodémographiques, les conditions dans lesquelles la recherche est exercée, le contexte de travail (même si sur ce point une analyse plus détaillée serait encore plus éclairante pour notre problème) font varier considérablement les pratiques. Il convient donc de passer des usages d'outils pré-sélectionnés aux pratiques entendues comme « faire avec » (Certeau, 2010 ; Jeanneret 2019) et il convient également de réinscrire dans les différences sociodémographiques et le contexte de travail des membres des communautés scientifiques.

Ainsi, les chercheuses et chercheurs paraissent développer des pratiques d'archivage qui prêtent le flanc à la proxémie même si nous ne pouvons pas discerner par une analyse quantitative des cultures façonnées par une multitude d'éléments. Nous empruntons donc la théorie pour aller vers des concepts opératoires pour qualifier des manières de faire, peut-être pourrions-nous également les qualifier de conventions ?

La première est celle de la proxémie spatiale et symbolique. Les membres de la recherche agissent dans des institutions et des organisations avec lesquelles ils ont des liens sociaux plus ou moins forts. N'oublions pas que l'activité de recherche demeure une activité professionnelle qui a connu (en France et ailleurs) des mutations profondes de son marché du travail. Ce monde professionnel est sans doute plus hétérogène socialement et statutairement qu'il ne l'a jamais été. Ainsi, quelle proximité entretient un doctorant avec son institution de rattachement par rapport à un chercheur plus aguerris et travaillant dans un organisme de recherche depuis 10 ans, alors que tous les deux concourent à produire des données de recherche ? Au centre nous ne plaçons donc pas l'intime mais l'institution dans laquelle la chercheuse ou le chercheur réalise ses missions de recherche c'est-à-dire l'« organisation matérialisée (institutions et politiques mémorielles) » de la recherche (Merzeau, 2006).

L'hypothèse que nous défendons ici est celle d'une proxémie spatiale et symbolique qui prend la forme d'une série de cercles de confiance plus ou moins rapprochés, qui aurait en son cœur le cercle contenant les institutions fournissant des ressources informatiques de sauvegarde les plus en phase avec les principes FAIR et à sa plus grande périphérie le cercle de confiance le plus restreint autour du chercheur lui-même, à savoir sa propre personne, ses proches et ses outils d'archivage, soient l'informatique personnelle dont il a la propriété. Le problème est que la défaillance possible de l'informatique personnelle amène à recourir à des plateformes commerciales disposant de ressources accessibles sans compte informatique institutionnel.

La seconde manière d'entendre la proxémie est la définition psycho-affective reposant sur une proximité avec autrui, une pratique, une histoire, en raison de l'écho symbolique et affectif qu'elle produit en nous. D'une certaine manière, cette perspective rejoint celle développée par Rebouillat même si nous nous portons moins sur la définition des valeurs que celle de confiance. Rebouillat pointe l'importance de la valeur accordée par le chercheur aux données de recherche, qui entraîne des représentations variables de la valeur (d'échange, d'usage et symbolique) associée à celles-ci et conduit à mettre en œuvre des pratiques de sauvegarde et d'archivage plus ou moins investies. Elle montre que les techniques d'archivage sont corrélées aux différentes formes de valeurs associées aux données que le chercheur produit. D'ailleurs la notion même de propriété des données de recherche se vit très différemment. Mais la question de la valeur et des risques induit également la notion de confiance. Or la confiance dépend de la familiarité (Luhmann, 2001). La notion de familiarité peut donc expliquer le recours à un tiers provenant de l'entourage familial et/ou amical.

5. Conclusion

Les conséquences de cette mise en perspective de nos résultats remettent en question le déterminisme technologique diffusé généralement au sein des recherches tentant de saisir la compréhension des pratiques de sauvegarde ou d'enregistrement à long terme des données de la recherche. En effet, la multiplication des outils à disposition, leur accessibilité accrue n'entraîne pas forcément une

adéquation et une augmentation des pratiques répondant aux principes FAIR. Il apparaît que le contexte dans lequel se déroule le travail de recherche pèse considérablement dans les pratiques d'archivage et de sauvegarde. Et l'émergence de savoir-faire et de bricolage ordinaire (Certeau, 2010) demeurent opérants dans le sens où ils répondent à une situation et un contexte de travail, à « faire-avec ». Ainsi, un chercheur qui utilise l'ordinateur d'un membre de sa famille ou d'un ami proche pour archiver ses données bricole un système original mais qui répond avant tout à la recherche du meilleur tiers de confiance.

En conséquence, les offres de sauvegarde et d'archivage, malgré les incitations institutionnelles qui peuvent les soutenir, se heurtent à la question de la construction de la confiance du chercheur envers ses supports et leur pérennité, sachant que sa confiance s'élabore à partir d'une situation de travail concrète, situations qui sont en France marquées par une hétérogénéité croissante. Il convient aussi d'expliquer les finalités et la potentialité d'une mise en commun pour les chercheurs des données de recherche. Le terme ouverture et partage masquent la possibilité première de mutualiser les recherches dans les communautés épistémiques ou des collectifs de travail.

Les chercheurs développent donc des savoir-faire et l'informatique n'exonère pas une réflexion sur l'archivage et une possible perte des données de recherche. Le développement de pratiques en matière de données de recherche interroge vivement les pratiques d'évaluation et de valorisation des activités de recherche. Par exemple, nos résultats montrent que les responsabilités dans la recherche amènent à une meilleure prise en considération de la question du stockage, de la sauvegarde et de l'archivage des données de recherche. Selon les modèles de gestion en management des connaissances et les définitions des communautés scientifiques qu'elles soient épistémiques ou de pratiques, il semblerait opportun de former les chercheurs non en fonction de disciplines mais selon leurs pratiques de recherche et leur collectif de recherche.

6. Bibliographie

Akrich, M. (2006). Les utilisateurs, acteurs de l'innovation. In Akrich, M., Callon, M., & Latour, B. (Eds.), *Sociologie de la traduction : Textes fondateurs*. Presses des Mines. <https://doi.org/10.4000/books.pressesmines.1181>

Amiel, P., Frontini, F., Lacour, P.-Y., & Robin, A. (2020). « Pratiques de gestion des données de la recherche : Une nécessaire acculturation des chercheurs aux enjeux de la science ouverte ? Résultats d'une enquête exploratoire dans le bassin montpelliérain (juin 2018) ». *Cahiers Droit, Sciences & Technologies*, 10, 147-168. <https://doi.org/10.4000/cdst.2061>

Certeau, M. de, (2010). *Arts de faire* (Nouvelle éd). Gallimard.

Desfriches-Doria, O. (2019). Glossaire : Les notions clés du RGPD. I2D - Information, données & documents, n° 1(1), 87. <https://doi.org/10.3917/i2d.191.0087>

Denis, J., & Goëta, S. (2017). « Les facettes de l'Open Data : Émergence, fondements et travail en coulisses ». In P.-M. Menger & S. Paye (Éds.), *Big data et traçabilité numérique* (p. 121-138). Collège de France. <https://doi.org/10.4000/books.cdf.5005>

Hall, E. T. (1990). *The hidden dimension*. New York: Anchor Books.

Jeanneret, Y. (2019). *La fabrique de la trace*. ISTE. <https://www.istegroup.com/fr/produit/la-fabrique-de-la-trace/>

Le Béhec, M., Bouchard, A., Charrier, P., Denecker, C., Gallezot, G., & Rennes, S. (2022). *Pratiques et usages des outils numériques dans les communautés scientifiques en France*. [Research Report]. Comité pour la science ouverte. <https://doi.org/10.52949/5>

LOI n° 2012-300 du 5 mars 2012 relative aux recherches impliquant la personne humaine (1), 2012-300 (2012).

Luhmann, N. (2001). « Familiarity, Confidence, Trust : Problems and alternatives ». *Réseaux*, 108(4), 15-35.

Merzeau L. (2006), « Mémoire », *Médium*, 9(1), 153-163.

Meyer, M., & Molyneux-Hodgson, S. (2011). « Communautés épistémiques : Une notion utile pour théoriser les collectifs en sciences ? » *Terrains travaux*, 18(1), 141-154.

Ministère de l'Enseignement supérieur et de la recherche. (2021). Deuxième Plan national pour la science ouverte. <https://www.ouvirlascience.fr/deuxieme-plan-national-pour-la-science-ouverte>

Rebouillat V. (2021) « Le partage des données vu par les chercheurs : une approche par la valeur », *Les Enjeux de l'Information et de la Communication*, 22(1), 35-53, <https://lesenjeux.univ-grenoble-alpes.fr/2021/varia/03-le-partage-des-donnees-vu-par-les-chercheurs-une-approche-par-la-valeur/>

Sbeih L. et al., (2020) « L'archivage des données de la recherche à l'Inra. Éléments de réflexion, démarche et perspectives », *Archive Ouverte d'INRAE*, ID : 10670/1.ficf4q

Schöpfel, J. (2018). *Vers une culture de la donnée en SHS* [Research Report]. Université de Lille. <https://hal.archives-ouvertes.fr/hal-01846849>

Star, S. L. (1999). « The Ethnography of Infrastructure ». *American Behavioral Scientist*, 43(3), 377-391. <https://doi.org/10.1177/00027649921955326>