

L'ouverture des données qualitatives de recherche en sciences sociales : chimère ou refondation ?

Alix LEVAIN, Florence REVELIN, Morgane MIGNON, Marianne NOËL, Betty QUEFFELEC Pascal RAUX, Hervé SQUIVIDANT

Laboratoire AMURE, LESC, LISIS, SAS

Nous disposons à ce jour de peu de recul sur la manière dont les communautés de recherche en sciences humaines et sociales (SHS) se saisissent du mouvement contemporain de "l'ouverture des données" (Open Data) dans le contexte de l'avènement de la "science Ouverte" (Open Science). Ce mouvement international, désormais transposé dans l'agenda politique français, reflète des évolutions profondes qui touchent l'exercice de la recherche scientifique, dans un contexte de numérisation croissante des pratiques et productions scientifiques, et s'appuie sur des grands principes de transparence (crédibilité, reproductibilité et vérifiabilité), d'efficacité d'une recherche cumulative, de démocratisation de l'accès aux savoirs, et d'innovation. Comment les praticien.ne.s de la recherche qualitative (personnels scientifiques et techniques) sont-ils.elles touchés par le mouvement d'ouverture des données ? Comment appréhendent-ils.elles les injonctions en faveur de l'ouverture des données ? Dans quelle mesure intègrent-ils.elles cet objectif d'ouverture dans leurs pratiques de recherche et quels déplacements cela suscite-t-il ? Les résultats récemment publiés du projet PARDOQ apportent un éclairage nouveau sur ces questions.

L'engagement politique en faveur d'une science ouverte¹ est désormais porté auprès des communautés de recherche à travers différents leviers incitatifs ou contraignants, actionnés en particulier par les structures de financement de la recherche d'échelle nationale ou internationale, comme l'Agence nationale pour la Recherche (ANR) ou le Conseil européen de

¹ Transposé en France par la Loi pour une République Numérique (2016) puis le Plan national pour la Science Ouverte (2018 et 2021)

la recherche (ERC) – bien que celles-ci peinent à les adapter aux SHS. Ceux-ci rendent obligatoire depuis 2019 la publication ouverte des productions scientifiques et des données de recherches (non soumises à un régime de protection) issues des programmes financés. Les établissements publics à caractère scientifique et technique (EPST) et les universités déploient également des dispositifs d'accompagnement et d'incitation à l'ouverture. Le CNRS, par exemple, a publié en 2018 sa "*Feuille de route pour la Science Ouverte*" qui présente son positionnement en faveur de l'ouverture des publications (Open Access) et des données (Open Data), suivi en 2020 du "*Plan données de la recherche du CNRS*" dans lequel il réaffirme ses objectifs de promouvoir "*une culture de la donnée FAIR*"² et d'équiper les communautés de recherche dans ce sens. Enfin, certaines revues scientifiques, en particulier pour les sciences biophysiques, demandent désormais la publication des corpus de données sur lesquels s'appuient les résultats d'analyses publiés.

Dans ce contexte, le projet de recherche *Partage et protection des données qualitatives de recherche à l'ère dunumérique : expériences, enjeux, stratégies (PARDOQ)*³ visait à **aborder la complexité, les leviers et les formes du déploiement de ce mouvement** récent en faveur d'une Science Ouverte, dans lequel s'inscrit le projet d'ouvrir "autant que possible" l'accès aux données de la recherche selon le principe directeur du mouvement de l'*Open Data* ("ouvert autant que possible, fermé autant que nécessaire") défendu par la Commission européenne, et repris de toutes parts par les relais institutionnels promouvant l'ouverture des données.

Comprendre ce paysage en pleine évolution apparaissait nécessaire pour le **mettre en regard avec l'expérience des communautés de recherche, qui se trouvent de plus en plus confrontées et questionnées par l'intégration de ces nouveaux enjeux dans leurs pratiques**. Il s'agissait ainsi de comprendre comment le mouvement d'ouverture des données est **vécu, pensé ou impensé, approprié ou suscite des formes de résistance** de la part des communautés de recherche dont les travaux s'appuient sur une démarche qualitative et empirique, produisant des matériaux de recherche (ou "données") particulièrement difficiles à caractériser à l'aune des préconisations de la Science Ouverte.

² Selon les principes FAIR, chaque donnée doit être mise en circulation libre sur le web dans le respect des principes qui lui garantissent d'être "facilement trouvable", "accessible à tous", "interopérable" (au sein des systèmes d'information numériques) et "réutilisable" (Wilkinson et al. 2016), pour être rendue aisément et universellement appropriable.

³ Le projet PARDOQ (2019-2021) a bénéficié du soutien de la Région Bretagne via le GIS Marsouin, et de l'appui logistique de la Maison des Sciences de l'Homme de Bretagne et de l'UMR AMURE.

L'ethnographie est une pratique de plus en plus légitime et partagée au sein des SHS. Qu'elle intervienne en complément d'autres méthodes de recherche ou de façon exclusive, elle constitue une référence commune pour de nombreux.euses chercheur.euses dans des champs d'étude aussi divers que la santé, l'environnement, la politique, les pratiques culturelles, les organisations. Son caractère intégrateur, le poids qu'occupent les données à caractère personnel, l'hétérogénéité et le foisonnement des matériaux auxquels elle donne naissance, la façon dont elle engage la subjectivité des chercheurs.euses, font de **l'ethnographie une "pratique limite" lorsqu'elle est appréhendée sous l'angle de la standardisation, de la normalisation et de l'institutionnalisation de la gestion des données de la recherche.**

L'enquête PARDOQ est la première enquête qualitative menée en France sur les pratiques et représentations du partage et de la protection des données au sein des communautés de recherche en sciences sociales pratiquant ou accompagnant le recueil de matériaux ethnographiques⁴. L'approche qualitative retenue pour cette enquête rend visible plusieurs dimensions de ce rapport aux données, qu'une enquête par questionnaire s'appuyant sur l'usage de notions largement issues de terminologies juridiques ou techniques et faisant l'objet d'un traitement statistique ne permettrait pas d'identifier avec la même finesse, et peut-être pas du tout. Elle vient donc enrichir les études récentes, de plus en plus nombreuses, qui abordent ces questions dans une double perspective de quantification et d'appui aux politiques émergentes de gestion des données de recherche à l'échelle des laboratoires et des établissements d'enseignements supérieur et de recherche.

La pratique ethnographique à l'épreuve de la gestion hétéronome des données

Ce "moment "données"", qui s'étend bien sûr bien au-delà du champ de la recherche publique, peut ainsi être abordé comme un moment de densification des controverses et des dilemmes suscités par la gouvernance contemporaine des sciences à l'ère du numérique⁵. Cette politique de mise à disposition des données spécifiquement dirigée vers la recherche scientifique se croise avec le mouvement plus général de la protection des données à

⁴ Pour une présentation détaillée des résultats de l'enquête, voir le rapport d'enquête complet, accessible sur HAL: <https://hal.archives-ouvertes.fr/hal-03238067>

⁵ Bigot, Jean-Édouard. 2018. « Instruments, pratiques et enjeux d'une recherche numériquement équipée en sciences humaines et sociales ». PhD Thesis, Compiègne ; Galonnier, Juliette, Stefan Le Courant, Anthony Pecqueux, et Camille Noûs. 2019. « Les sciences humaines et sociales au travail : Que faire des données de la recherche ? » <https://doi.org/10.4000/traces.10518>

caractère personnel rendu nécessaire par le big data, avec lequel il entre régulièrement en tension, en particulier dans la recherche publique. En effet, l'évolution du cadre législatif relatif aux données à caractère personnel (en particulier le Règlement Général sur la Protection des Données - RGPD et sa transposition dans la Loi n° 2018-493 du 20 juin 2018 relative à la protection des données personnelles) pose à la recherche qualitative en SHS des questions particulièrement aiguës. En premier lieu, elle oblige à **considérer à nouveaux frais la relation avec les personnes enquêtées et le statut des matériaux de recherche recueillis auprès d'elles, et grâce à elles**. En second lieu, elle conduit à **l'irruption de tiers extérieurs à l'enquête dans cette relation, positionnés comme garants du respect de normes juridiques et éthiques par les chercheurs et chercheuses**. En troisième lieu, elle **bouleverse le rapport à l'enquête de terrain, en remettant en cause les vertus et fonctions heuristiques attribuées traditionnellement à celle-ci** (la possibilité de faire émerger de nouvelles hypothèses et de nouvelles problématiques, l'indétermination et la disponibilité au réel, l'articulation d'une temporalité longue avec des temporalités plus resserrées, le partage d'expériences et l'importance du matériau collecté autour d'interactions informelles), au profit d'une vision planifiée, d'une anticipation des interactions et d'une trajectoire maîtrisée. Les sociologues et anthropologues ayant formalisé leurs questionnements à ce sujet à partir de leur propre expérience sont encore peu nombreux ; mais certains vont jusqu'à considérer que le respect de ces nouveaux cadres rend certaines recherches impossibles en raison du caractère sensible et subversif de certaines enquêtes et de l'incompatibilité entre respect de l'éthique professionnelle et partage⁶.

Les SHS sont directement et particulièrement concernées par les données "fermées". Dans certains cas, l'encadrement juridique et les prescriptions éthiques préexistantes sont très explicites et lisibles : c'est le cas des données médicales. C'est également le cas *a priori* des données dites "sensibles"⁷. Pour ces dernières, cependant, la question posée n'est pas seulement celle de leur fermeture, mais se formule également en termes de possibilité-même

⁶ Voir par exemple, à propos d'une enquête au Front National: Bizeul, Daniel. 2020. « Faut-il tout dévoiler d'une enquête au Front national ? Réflexions sur le partage des données et le devoir éthique en sociologie ». Bulletin of Sociological Methodology/Bulletin de Méthodologie Sociologique, octobre, 0759106320960887. <https://doi.org/10.1177/0759106320960887>

⁷ L'article 9 du RGPD définit les données sensibles comme "des données à caractère personnel qui révèle(nt) l'origine raciale ou ethnique, les opinions politiques, les convictions religieuses ou philosophiques ou l'appartenance syndicale, ainsi que le traitement des données génétiques, des données biométriques aux fins d'identifier une personne physique de manière unique, des données concernant la santé ou des données concernant la vie sexuelle ou l'orientation sexuelle d'une personne physique sont interdits."

de collecte : les formes de justification de la collecte de données sensibles à des fins de recherche et leur gestion pérenne⁸, une fois collectées, posent de nombreuses difficultés. Qui plus est, le respect du cadre légal, en particulier l'inscription préalable au registre des activités de traitement prévue à l'article 30 du RGPD implique une gestion temporelle et d'échelle très fine dans les cas, nombreux, où les données ne contiennent pas a priori de données sensibles, mais où la collecte de matériaux qualitatifs - par exemple dans le cadre d'observations participantes ou d'entretiens ouverts - en incorporent in fine, de façon non anticipable.

Le cas des données sensibles montre à quel point, d'une part, tous les champs de recherche ne sont pas également exposés à la complexité de mise en œuvre du droit qui s'applique aux données de recherche, et d'autre part, que les SHS, en particulier celles qui déploient des méthodes d'enquête qualitative, le sont particulièrement. **Entre l'ouverture de principe et l'exception de fermeture, subsiste donc une large zone grise** qui nécessite d'une part une qualification juridique requérant un haut niveau d'expertise, et d'autre part des arbitrages minutieux, au cas par cas : **c'est à cet endroit-même que se trouvent une grande partie des données ethnographiques.**

Bien qu'il se présente comme la traduction dans une formule accessible et synthétique d'un ensemble de dispositions juridiques contraignantes et complexes, le principe « Aussi ouvert que possible, aussi fermé que nécessaire » a également, de fait, une portée normative sur un plan plus large : il affecte une valeur positive à l'ouverture, vers laquelle un ensemble d'incitations et de ressources doivent être orientées, et positionne la fermeture des données dans le champ des pratiques illégitimes et relictuelles, si elles ne sont pas justifiées par une caractérisation juridique ad hoc. L'écart entre le « fermé observé » et le « fermé légitime » apparaît ainsi comme un nouveau critère d'évaluation de la qualité des pratiques de gestion des données et comme un point de focalisation des tutelles et des réseaux d'appui.

Les enquêtes récentes

Les travaux qui rendent compte de ce contexte "post RGPD" et "post LRN (Loi pour une République numérique)" sont encore peu nombreux, en particulier pour ce qui concerne la recherche qualitative en SHS, pour laquelle l'articulation entre ouverture des données

⁸ L'article 9 énonce plusieurs exceptions au principe de non traitement des données personnelles sensibles, dont le cas où le "traitement est nécessaire à des fins archivistiques dans l'intérêt public, à des fins de recherche scientifique ou historique ou à des fins statistiques".

(principe d'ouverture par défaut) et protection des données personnelles se pose de façon aiguë, le principe d'ouverture par défaut apparaissant en l'état difficilement applicable⁹. C'est *a fortiori* le cas pour l'ethnographie, pratique qui dépasse aujourd'hui largement le cercle des seuls ethnologues (ou anthropologues, selon la terminologie qui s'impose progressivement), mais constitue également le coeur de la matière empirique à partir de laquelle de nombreux sociologues, des politistes et, dans une moindre mesure, des géographes, construisent leurs analyses.

L'histoire des pratiques de partage des matériaux et des résultats de recherche est en effet marquée par un impératif déontologique de restitution et la qualité de la relation enquêteur-enquêté constitue une instance d'évaluation centrale de la recherche dans ces champs disciplinaires¹⁰. L'attention à la matérialité des cultures et l'inscription de la pratique ethnographique dans le champ de la patrimonialité renforce le poids de cette exigence. Toutefois, ces pratiques restent artisanales, dispersées, et font l'objet d'un contrôle épistémologique et déontologique distribué parmi les pairs plutôt que d'une gestion par les établissements de rattachement, les autorités de tutelle, les bailleurs de fond et, bien sûr, les juridictions. La façon inégale dont le contrôle éthique et déontologique s'est développé dans ce champ montre à quel point les communautés d'ethnologues s'inscrivent dans des traditions nationales, voire locales, dans lesquelles les sujets, terrains et collaborations déterminent davantage que les appartenances institutionnelles ou les cadres juridiques la codification des pratiques.

Deux types d'enquêtes récentes permettent de documenter empiriquement les écarts entre d'une part les cadres incitatifs du mouvement d'*ouverture des données* et ses traductions dans des injonctions institutionnelles, et d'autre part, les pratiques des membres des communautés de recherche en SHS. Tout d'abord, les enquêtes menées dans le contexte d'un effort d'équipement technique et numérique adapté aux enjeux de gestion, d'archivage pérenne et de sécurisation des données de la recherche, et qui se sont multipliées dans les années récentes pour évaluer et caractériser les pratiques des chercheur.euses (types de données numériques mobilisées, outils de traitement et de stockage, niveau de connaissance des

⁹ Maurel, Lionel. 2018. « La réutilisation des données de la recherche après la loi pour une République numérique ». In La diffusion numérique des données en SHS - Guide de bonnes pratiques éthiques et juridiques. Presses Universitaires de Provence. <https://hal.archives-ouvertes.fr/hal-01908766>

¹⁰ Katz, Jack. 2001. « From how to why: On luminous description and causal inference in ethnography (Part I) ». *Ethnography* 2 (4): 443-73; et: 2002. « From How to Why: On Luminous Description and Causal Inference in Ethnography (Part 2) ». *Ethnography* 3 (1): 63-90. <https://doi.org/10.1177/1466138102003001003>

enjeux juridiques liés au traitement et à la sécurisation des données, pratiques de partage et d'ouverture, etc.) et leurs besoins en accompagnement et instrumentation techniques se sont multipliés ces dernières années. Il s'agit d'enquêtes menées à l'échelle nationale, comme celle du Consortium Couperin¹¹ ou du projet CommonData¹², à l'échelle d'Universités comme Lille 3¹³, suivie par Rennes 2¹⁴, Bordeaux Montaigne¹⁵ ou Montpellier Paul Valéry¹⁶; ou encore, menées à l'échelle de laboratoires ou autres collectifs de recherche (comme celle du Cresson¹⁷), et dont les résultats n'ont, pour la plupart, pas vocation à être publiés mais davantage à informer une politique interne de soutien à la gestion des données. Ces enquêtes allient souvent une approche quantitative par questionnaire, suivie d'une qualification plus fine de certains enjeux, pratiques et tendances grâce à des entretiens individuels.

Elles se concentrent principalement sur certains enjeux juridiques et techniques liés au souci de mise en conformité avec les cadres juridiques et institutionnels, au détriment d'autres (épistémologiques, méthodologiques et/ou de qualité scientifique). Elles témoignent toutes à des degrés divers d'une distance importante des répondant.es vis-à-vis des attentes institutionnelles, bien souvent aussi d'une forme d'incompréhension face à la formulation des enjeux ou d'une difficulté à faire le lien entre leurs pratiques et les questions posées. Le second corpus est constitué de retours d'expériences individuels ou collectifs de professionnel.les de la recherche qui expérimentent ces processus de partage et d'ouverture, pris en charge par les équipes scientifiques elles-mêmes, et qui éclairent justement ces zones

¹¹ Rousseau-Hans, Françoise, Christine Ollendorff, et Vincent Harnais. 2020. « Les pratiques de publications et d'accès ouvert des chercheurs français en 2019 », 94.

¹² Amiel, Philippe, Francesca Frontini, Pierre-Yves Lacour, et Agnès Robin. 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):147-68. doi: 10.4000/cdst.2061

¹³ Prost, Hélène, et Joachim Schöpfel. 2015. « Les données de la recherche en SHS. Une enquête à l'Université de Lille 3. » Report. <https://hal.univ-lille3.fr/hal-01198379>

¹⁴ Serres, Alexandre, Marie-Laure Malingre, Morgane Mignon, Cécile Pierre, et Didier Collet. 2017. « Données de la recherche en SHS. Pratiques, représentations et attentes des chercheurs : une enquête à l'Université Rennes 2 ». Research Report. Université Rennes 2. <https://hal.archives-ouvertes.fr/hal-01635186>

¹⁵ Duprat, Julie. 2019. « Les données de la recherche à l'Université Bordeaux Montaigne ». Research Report. Université Bordeaux Montaigne. <https://hal.archives-ouvertes.fr/hal-02020141>

¹⁶ Dillaerts, Hans, Céline Paganelli, Lise Verlaet, et Catherine Hugo. 2020. « Usages et pratiques en lien avec les données de recherche. Une enquête menée auprès des chercheurs de l'université Paul-Valéry Montpellier 3 ». PhD Thesis, Université Paul-Valéry Montpellier 3.

¹⁷ Mosnier, Léa, Françoise Acquier, et Véronique Dom. 2019. « Les données de la recherche AAU-CRESSON: résultats de l'enquête sur les usages des chercheurs, doctorants et ingénieurs en matière de gestion de données ». PhD Thesis, Centre de recherche sur l'espace sonore et l'environnement urbain.

d'ombre (en particulier les publications réflexives et méthodologiques associées aux projets BeQuali¹⁸, Meta-Projet¹⁹, SAWA²⁰ et Parchemins²¹)²².

L'enquête "PARDOQ" auprès de communautés de recherche travaillant sur des matériaux ethnographiques

L'enquête PARDOQ s'est principalement appuyée sur le recueil de l'expérience de 26 personnels scientifiques et techniques dont les travaux mobilisent des matériaux de recherche ethnographiques (chercheur.euses, enseignant.es-chercheur.euses, doctorant.es et réseaux d'appuis scientifique, technique et juridique à la recherche, incluant par exemple des ingénieur.es d'étude et de recherche et des délégué.es à la protection des données personnelles). Notre hypothèse est que ces matériaux à la fois très qualitatifs (observations, entretiens, archives, photographies, vidéos...), de natures et de formats très hétérogènes, et qui embarquent la relation complexe entre chercheur.euses, terrains et sujets de recherche, sont susceptibles de mettre en lumière une diversité de configurations et de problématiques lorsque leur mise en partage est envisagée.

L'enquête a porté sur les différents rapports de ces interlocuteur.rices au travail et aux spécificités des matériaux de recherche ethnographiques, sur les formes de partage dans lesquelles ils sont engagés le cas échéant, et sur leurs rapports aux politiques et injonctions

¹⁸ Voir : Cadorel, Sarah, Guillaume Garcia, Émilie Groshens, Émeline Juillard, Jérémie Vandebunder, Stefan Le Courant, Anthony Pecqueux, et Camille Noûs. 2019. « Archiver, documenter, enquêter sur l'enquête qualitative. Le travail de l'ombre de beQuali ». Tracés. Revue de Sciences humaines, no #19 (décembre): 183-98. <https://doi.org/10.4000/traces.10968>

¹⁹ Voir notamment: Huyghe, Marie, Laurent Cailly, et Nicolas Oppenheim. 2018a. « Partage des données d'enquêtes en recherche qualitative: approche juridique, éthique, déontologique et épistémologique ». Et : 2018b. « Le méta-projet, retour d'expérience d'une analyse secondaire ». <https://halshs.archives-ouvertes.fr/halshs-01922702>

²⁰ Voir notamment : Vapnarsky, Valentina. 2020. « Retour aux sources ? Circulation et virtualités des savoirs amérindiens à l'ère du numérique ». Journal de la Société des américanistes 106 (106-2): 79-103. <https://doi.org/10.4000/jsa.19003>; Camargo, Éliane, Veronica Holguin Lew, Sara Tandar, et Équipe Wayana et Apalaï. 2020. « L'Amazonie amérindienne dans l'ère du numérique : le portail multilingue WATAU ». Patrimoines du Sud, no 12 (septembre). <https://doi.org/10.4000/pds.5127>

²¹ Voir notamment : Squidant, Hervé, et Florence Revelin. 2021. « S'hybrider sous le 4e paradigme de la "science dirigée par les données massives" : l'ouverture des données favorise-t-elle l'interdisciplinarité et la transdisciplinarité ? » Communication, technologies et développement, no 9 (mars). <http://journals.openedition.org/ctd/3791>

²² Pour une présentation de certaines de ces expériences, voir également la captation vidéo de l'atelier PARDOQ L'ouverture des matériaux de recherche ethnographiques en question (Rennes, 2 juin 2021), accessible sur Canal-U : https://www.canalu.tv/video/mshb/atelier_pardog_l_ouverture_des_materiaux_de_recherche_ethnographiques_en_question.62487

liées au mouvement d'ouverture de données d'un côté, et à la protection des données de l'autre. Une attention particulière a été accordée aux catégories émic de ces différents profils d'interlocuteurs et en particulier à leurs manières de penser et nommer les matériaux de recherche auxquels ils sont confrontés dans leur travail soit de recherche, soit d'appui à la recherche ; d'autre part, nous nous sommes intéressés à la manière dont ces interlocuteurs s'équipent pour traiter ces matériaux et éventuellement les partager ou au contraire les protéger.

Principaux enseignements : les formes de la résistance aux « données »

Le premier résultat de cette enquête est la mise en évidence de la façon dont les chercheurs.euses conçoivent les matériaux empiriques qu'ils collectent: **la notion de "donnée" n'est pour eux.elles une évidence ni sur le plan épistémologique, ni a fortiori sur le plan juridique.** Cela se traduit par exemple par le fait que tou.te.s les praticien.ne.s interrogé.e.s ne l'utilisent pas pour désigner les informations et matériaux empiriques sur lesquels ils.elles fondent leurs analyses. L'une des conséquences très concrètes de ce rapport spécifique aux données est que **les chercheurs.euses pratiquant l'ethnographie ne se projettent qu'exceptionnellement dans la réutilisation de données ethnographiques produites par d'autres**, et ne pensent symétriquement pas leurs matériaux comme appropriables ex post par d'autres, c'est-à-dire détachables de leur contexte de production et de leur *auteur.e*. Cet *ethos* de l'auteur, dont l'enquête PARDOQ a montré qu'il était dominant chez certain.e.s ethnographes, mais présent chez tou.te.s, positionne les enjeux de propriété intellectuelle plutôt de façon sélective sur des objets circonscrits, des "oeuvres", et pas sur des séries de données.

Dans quelles circonstances et à quelles conditions le processus de "détachement" se produit-il alors ? La qualité de "donnée", au sens d'une unité élémentaire d'information stabilisée et partageable, s'acquiert dans différentes configurations, dont l'enquête PARDOQ livre un aperçu:

- par **hybridation de paradigme**, c'est-à-dire par imprégnation initiale et/ou intégration dans des communautés de recherche interdisciplinaires dans lesquelles préexiste une "culture de la donnée" ou dans lesquelles, comme dans le cas du programme Parchemins étudié de façon détaillée à l'occasion de l'enquête PARDOQ, celles-ci servent d'objet intermédiaire pour consolider des approches interdisciplinaires. Dans ce contexte s'opère une différenciation et une sélection des matériaux ethnographiques bancarisables;

- par **l'attribution d'une valeur patrimoniale** à des matériaux en raison de leur importance culturelle ou scientifique. Ce cas est illustré dans le cadre de cette étude, à l'échelle individuelle, par la situation de transmission des matériaux de recherche en fin de carrière (Figure 2), et à l'échelle collective, sur deux registres différents, par l'exemple de la patrimonialisation des archives anciennes des ethnologues et celui du renforcement de l'impératif moral et épistémologique de restitution des matériaux à leurs communautés d'origine ;

Dans ces deux configurations, les enjeux scientifiques que les chercheur.euses portent, c'est-à-dire les dynamiques endogènes au champ scientifique, sont susceptibles de converger avec les cadrages externes qui accompagnent l'institutionnalisation de la gestion des données. Car si elles préexistent aux évolutions récentes du droit, celles-ci viennent cependant renforcer leur légitimité scientifique, avec des conséquences matérielles de plus en plus visibles : l'allocation de moyens d'ingénierie et d'appui à la recherche plus conséquents, ou l'obtention de financements sur projets adossés aux infrastructures de publication de données.

Si ces exemples témoignent de dynamiques endogènes fortes au sein des communautés ancrées dans la pratique ethnographique, ils restent néanmoins très minoritaires : **l'ethnographie reste une pratique le plus souvent solitaire, très artisanale, de plus en plus réalisée dans des conditions statutaires et matérielles précaires**. Tous les ethnographes n'ont ainsi pas également accès aux ressources, au premier rang desquelles le temps de travail et les compétences à acquérir ou sur lesquelles s'appuyer au sein d'une équipe.

De ce point de vue, les tensions posées d'une part par la mise en œuvre du RGPD et les dispositions de la Loi dite Lemaire en ce qui concerne la protection des données à caractère personnel et d'autre part celles qui sont soulevées par l'obligation « aussi souvent que possible » d'ouvrir les données de recherche peuvent *a priori* paraître de nature différente, bien qu'elles soient mises simultanément à l'agenda et généralement présentées ensemble par des acteurs institutionnels sous l'angle de la transformation des pratiques que toutes deux requièrent.

Les premières concernent principalement la protection des enquêté.es. L'histoire de la pratique ethnographique est marquée par la lente construction de conventions collectives le plus souvent informelles qui visent à réguler le risque pour les enquêté.es à la fois à l'aune d'une protection contre l'intrusion d'acteurs tiers et contre d'éventuelles négligences des ethnographes eux-mêmes. Ces régulations imprègnent fortement les normes professionnelles

visant la protection des enquêté.es et de leurs droits, puisqu'elles reposent autant sur la protection de l'indépendance des chercheur.euses, que sur les pratiques et les choix de techniques littéraires *ad hoc* permettant, suivant les cas, une anonymisation très rigoureuse, une protection sourcilleuse des informateurs ou une exposition volontaire de certains acteurs sociaux dominants. **C'est donc une approche différenciée et casuistique qui prévaut, fondée sur une déontologie qui s'accorde mal avec l'imposition de cadrages uniformes *ex situ*.** A cet égard, les résultats de cette enquête mériteraient d'être complétés et éclairés par une enquête sur la socialisation des chercheur.euses pratiquant les méthodes qualitatives et l'ethnographie, et sur la place qu'y occupe l'apprentissage de la déontologie professionnelle.

Par contraste, les tensions qui naissent de l'injonction à l'ouverture ne sont pas déontologiques, mais plutôt d'ordre épistémologique. L'enquête PARDOQ montre à quel point **les pratiques de documentation de métadonnées et d'ouverture des données ethnographiques sont nombreuses et parfois très anciennes**, par exemple lorsqu'elles s'inscrivent dans une démarche systématique classificatoire et d'inventorisation de la diversité biologique et culturelle, intégrant dès la collecte une perspective de conservation et de constitution de collections muséales par exemple. A cet égard, le développement des outils et infrastructures numériques constitue, dans ce domaine comme ailleurs, une transformation dans les pratiques dans lesquelles les chercheur.euses enquêté.es et les projets étudiés dans le cadre de cette enquête s'inscrivent pleinement et dont ils et elles constituent parfois de fervents promoteurs. Cependant, ces pratiques ne sont pas spontanément conçues ni reliées par ceux et celles qu'elles concernent au cadre légal et aux offres institutionnelles de services dans lesquelles elles s'inscrivent ou pourraient s'inscrire. **Ce qui s'observe n'est pas un mouvement global vers l'ouverture**, dans lequel les praticien.nes repéreraient une rupture ou une accélération à partir de 2016. C'est plutôt un exercice permanent de composition et d'arbitrage dans lequel les formes de l'ouverture et de partage sont très peu standardisées et interviennent à différents moments du processus de recherche, en fonction de considérations ayant trait à la nature des données et aux objectifs de leur collecte et de leur exposition. Les pratiques de partage des matériaux ethnographiques sont donc, le plus souvent, éloignées des schèmes descriptifs du "cycle de vie" de la donnée, inspirés des sciences de l'information et dans lesquels le processus d'ouverture est représenté comme une série d'opérations successives et distinctes. L'expérience pratique montre combien la notion de "donnée", et *a fortiori* de "donnée brute" est questionnée dans les initiatives de bancarisation et mise en partage des matériaux des chercheur.euses pratiquant l'enquête ethnographique, si toutefois

ils s'adonnent à l'exercice sous cette forme. Les différentes formes de traitement du matériau d'enquête impliquent un travail fin de sélection de ce qui peut ou non faire l'objet d'une forme de publicisation. Si le choix se porte sur une publication dans un catalogue, un entrepôt ou autre forme de banque de données, alors ce matériau est retravaillé pour être nettoyé, mis en forme ou au format, éventuellement anonymisé, numérisé, et documenté par des métadonnées qui visent à le décrire et à le rattacher à son contexte de production de sorte qu'il puisse être approprié et réutilisé dans des conditions pratiques et éthiques acceptables pour son producteur primaire. Cette superposition d'étapes de traitement du matériau - aux fins de permettre son accessibilité et sa réutilisation secondaire, ou de rendre transparent le processus d'élaboration de résultats publiables, et non seulement à des fins de renseigner une question de recherche posée par le.la chercheur.euse

- consiste à en détacher des formes d'unités élémentaires qui puissent être bancarisées, elles-mêmes rattachées et enrichies d'une somme d'informations qui permette de replacer cette unité élémentaire à des éléments du contexte dont elle a été extraite. Ce processus d'assèchement/enrichissement façonne et module indéniablement le matériau et, loin de ne constituer qu'une étape dans un processus par ailleurs inchangé, engage et remet en cause l'ensemble du processus de recherche. Cet écart nous paraît, à l'issue de cette enquête, essentiellement se traduire par **une distance forte et par des incompréhensions régulières entre chercheur.euses pratiquant l'ethnographie et réseaux d'appui à la recherche** qui se révèlent, en matière de gestion des données, constituer un maillon essentiel, très actif dans la diffusion des informations juridiques et techniques, et parfois militant de l'*Open Science*.

Pour ces raisons, une lecture critique située du processus d'ouverture des données de la recherche est indispensable. De ce point de vue, les résultats inédits issus des enquêtes Parchemins et PARDOQ convergent avec les principaux enseignements de la littérature scientifique en études sociales des sciences : si les valeurs portées par le mouvement de l'Open Science, comme l'impératif catégorique de protection des enquêté.es trouvent un écho plus que favorable au sein des communautés étudiées, en revanche les formes contemporaines de leur institutionnalisation révèlent des inégalités structurelles et favorisent une hiérarchisation des savoirs au détriment de la recherche qualitative en sciences sociales, plaçant les professionnel.les au coeur d'une série de tensions, voire de conflits de valeurs, difficiles à surmonter.

Une lecture possible - mais sans doute restrictive - des tensions que l'enquête Parchemins et l'enquête PARDOQ mettent à jour est **que la situation observée au tournant des années 2020**

est une situation d'apprentissage social: ces tensions seraient avant tout liées au fait que les communautés concernées, qu'il s'agisse des chercheur.euses ou des métiers d'appui à la recherche, ne maîtrisent pas encore les compétences nécessaires à l'ouverture des données, à la fois individuellement, et collectivement - dans le sens où la portée de cette transformation oblige à un réajustement des interactions entre métiers. Il est probable, en effet, que ces tensions soient particulièrement vives dans une période de transition entre deux régimes de rapport à la donnée : le cadre légal a connu des bouleversements récents dont la portée est encore difficilement évaluable, non seulement parce qu'il n'est encore effectivement mis en oeuvre que dans un nombre très limité de cas, mais aussi parce que la jurisprudence associée est encore quasiment inexistante. D'autre part, les infrastructures qui supportent ce changement ne sont pas stabilisées, ce dont témoigne la multiplicité des outils techniques, la restructuration des métiers, la difficulté à organiser et à rendre lisible la chaîne de traitement par des procédures connues et comprises à l'échelle des établissements.

Cependant, à l'issue de cette enquête, il nous semble fondamental de resituer ce régime émergent dans une problématisation plus large des conceptions dominantes du savoir, des façons de faire preuve, et de l'autonomie du champ scientifique. Le mouvement d'institutionnalisation de la gestion des données et la façon dont il rencontre, ou pas, les fondements épistémologiques sur lesquels les savoirs scientifiques sont construits, sont aussi - et peut-être doivent être au regard des valeurs dont se réclame ce mouvement - **des processus sociaux à rendre intelligibles, en observant minutieusement ce qui se cache derrière et se niche dans la norme et l'instrument.** Réalisée par et pour les communautés de recherche en SHS, cette enquête peut ainsi contribuer à renforcer l'appropriation de ces processus, au premier chef pour ceux et celles qui en sont les chevilles ouvrières.

Encadré : Synthèse des résultats de l'enquête PARDOQ sur l'ouverture des données de la recherche qualitative en sciences sociales

- Dans sa forme contemporaine dominante, l'ouverture des données ne s'inscrit pas dans une dynamique interne à la recherche qualitative en sciences sociales mais se produit davantage par des contacts volontaires ou contraints avec d'autres disciplines ou d'autres modèles de construction du savoir.
- La conformation aux cadres juridiques et politiques de l'*Open Research Data* nécessite un travail d'ajustement considérable et multiforme, dans un contexte où les ressources disponibles pour ce type de recherche sont mises à mal. La question de l'ouverture des

données ne se pose pas indépendamment des conditions dans lesquelles se produit la recherche par ailleurs (peu de moyens, contractualisation croissante, question des métiers et des statuts).

- Encore peu nombreuses, les enquêtes sur les pratiques des équipes de recherche travaillant sur des matériaux qualitatifs témoignent toutefois d'un écart considérable et persistant entre les pratiques et les attentes institutionnelles, tant en ce qui concerne les pratiques d'ouverture que le respect du droit applicable en matière de protection des données (sécurisation, protection des données à caractère personnel).
- Cela ne signifie pas pour autant que les communautés SHS ne mettent pas en place un ensemble de dispositifs visant à la fois à protéger leurs sources et à rendre accessibles leurs matériaux de recherche ; elles déploient pour ce faire un large éventail de pratiques à la fois "ancrées" (relations de confiance et éthique de recherche) et "bricolées", témoignant d'une préoccupation réelle et de capacités d'adaptation et d'innovation, dans des cadres non standardisés et peu formalisés.
- L'enquête PARDOQ montre en particulier l'engagement et le rôle moteur à titre individuel de certain.e.s chercheur.euses et ingénieur.es dans les activités de promotion de *la Science Ouverte*, non pas centrées sur l'ouverture des données mais dans le cadre d'approches plus globales de démocratisation de l'activité scientifique, qui peuvent adopter des postures très contrastées, voire polarisées vis-à-vis de la question spécifique de l'ouverture des données qualitatives. Le niveau de connaissances et de compétence au sujet de l'ouverture n'est donc pas prédictif de l'attitude vis-à-vis de celle-ci.
- Les communautés SHS pratiquant la recherche qualitative fondée sur l'ethnographie ne sont pas toutes exposées avec la même intensité à la nécessité d'intégrer les principes FAIR : les chercheur.euses travaillant sur des données sensibles, au sein de communautés interdisciplinaires et/ou ayant recours à des financements externes sont les plus familiers des cadres de l'*Open Research Data* et de la protection des données identifiantes au sens juridique du RGPD.
- Pour les chercheur.euses engagé.es dans ce type de démarche, il existe une confrontation à un ensemble d'épreuves : décalage entre la temporalité d'instruction des enjeux d'ouverture des données et celle de la recherche, identification des interlocuteurs et des cadres applicables, des outils, protocoles et des ressources (ce dans le contexte d'une nébuleuse d'offre d'outils qui évolue constamment). Dans le cas des recherches ethnographiques, seule

une petite partie du matériau collecté peut être aisément caractérisable *a priori* à l'aune d'un régime juridique stabilisé.

- Quelle que soit la qualité des ressources externes dont il.elle dispose, le.la chercheur.euse/collecteur.rice reste dans la pratique l'instance principale d'arbitrage et d'appréciation du statut ouvert ou fermé des matériaux qu'il.elle recueille. La recherche d'externalisation (auprès de personnels d'appui scientifique et technique) associée à ces pratiques demeure en grande majorité impossible.
- Ces questions sont plus largement discutées et débattues dans les communautés d'ingénieur.es (appui scientifique et technique à la recherche), qui plaident pour une meilleure intégration de ces enjeux dès la conception des projets. Ce constat soulève la question des métiers et d'une meilleure intégration des personnels de la BAP D et F dans les dynamiques actives de recherche. Or, ils.elles ne sont pas toujours présent.es aux côtés des chercheur.euses, pour des raisons historiques, budgétaires et/ou organisationnelles.
- Le retrait des communautés SHS vis-à-vis de l'ouverture des données de la recherche ne peut donc être interprété uniquement à l'aune d'un déficit de compétence ou d'une réticence dans une période de transition. Il a à voir avec les paradigmes auxquels ces communautés épistémiques se réfèrent (nature des données, intimité du chercheur embarquée dans son matériau de recherche, éthique de recherche), ainsi qu'à la connaissance et à la reconnaissance institutionnelle de leur pertinence et de leur validité.



Photographie 1. L'ethnobotaniste Claudine Friedberg devant son herbier, Octobre 2009 (© A. Epelboin)