



HAL
open science

L'ouverture des matériaux de recherche ethnographiques en question.

Florence Revelin, Alix Levain, Morgane Mignon, Marianne Noel, Betty Queffelec, Pascal Raux, Hervé Squividant

► To cite this version:

Florence Revelin, Alix Levain, Morgane Mignon, Marianne Noel, Betty Queffelec, et al.. L'ouverture des matériaux de recherche ethnographiques en question. : Rapport d'enquête du projet "Partage et protection des données qualitatives à l'ère du numérique: expériences, enjeux, stratégies". [Rapport de recherche] Centre national de la recherche scientifique. 2021. hal-03238067

HAL Id: hal-03238067

<https://hal.science/hal-03238067>

Submitted on 21 Jun 2021

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 - NoDerivatives 4.0 International License



Programme de recherche PARDOQ | 2020-2021

Partage et protection des données qualitatives à l'ère du numérique : Expériences, enjeux, stratégies

Rapport d'enquête | mai 2021



Auteures : Florence Revelin (CNRS, UMR 7186 LESC) et Alix Levain (CNRS, UMR 6308 AMURE)

Contributeurs.trices : Morgane Mignon (MSHB), Marianne Noël (CNRS, UMR 1326 LISIS), Betty Queffelec (Université de Bretagne Occidentale, UMR 6308 AMURE), Pascal Raux (Université de Bretagne Occidentale, UMR 6308 AMURE), Hervé Squidant (Agrocampus Ouest, UMR 1069 SAS)



Présentation

Le mouvement d'ouverture des données scientifiques constitue, pour les sciences humaines et sociales (SHS), un défi à la fois épistémologique, juridique, éthique et technique. Il se manifeste par des normes et injonctions multiples vis-à-vis des communautés de recherche, qui peinent à s'y conformer et à se saisir des instruments mis à leur disposition. Le projet PARDOQ vise à rendre intelligibles les implications complexes de ce mouvement pour les communautés travaillant à partir de données qualitatives, à travers l'analyse de l'expérience de chercheuses et chercheurs confronté.e.s à la tension entre partage et protection des données ethnographiques. Il s'appuie en particulier, pour ce faire, sur une enquête qualitative menée auprès de chercheurs.euses pratiquant l'ethnographie et de membres de réseaux scientifiques, techniques et juridiques d'appui et à la recherche, dont ce rapport présente les principaux résultats.

Remerciements

Les auteurs.trices et contributeurs.trices de ce rapport tiennent à remercier l'ensemble des personnes consultées et/ou interrogées lors de la réalisation de l'enquête.

Nous souhaitons en particulier insister sur l'importance capitale qu'ont revêtu, pour la construction et la maturation de nos analyses, la confiance que nos interlocuteurs.trices nous ont accordée, la sincérité de leurs propos, la profondeur de leurs réflexions et la volonté, très partagée parmi eux.elles, d'alimenter une réflexion collective, à laquelle nous souhaitons que ce rapport contribue.

Le projet PARDOQ doit beaucoup aux réflexions collectives menées, entre 2016 et 2020, au sein du projet Paroles et chemins de l'agriculture littorale (Parchemins), porté par l'INRAE et soutenu par la Fondation de France et la Région Bretagne. Nous souhaitons donc témoigner notre gratitude à l'ensemble de l'équipe scientifique de Parchemins, en particulier à Anne-Gaëlle Beurier.

Nous remercions également la Région Bretagne et l'équipe du GIS Marsouin, pour le soutien financier qu'elles ont apporté au projet PARDOQ bien sûr, mais également pour leur qualité d'écoute et d'accompagnement dans une période qui complexifie la réalisation des enquêtes de terrain comme les animations scientifiques.

Note aux lectrices et lecteurs

Nous nous sommes efforcées de tendre vers une écriture non sexiste dans la rédaction de ce rapport, en veillant à privilégier les formules épiciènes et neutres chaque fois que cela est possible, afin de fluidifier la lecture, mais en utilisant, lorsque cela nous paraissait nécessaire, l'écriture inclusive.

Pour refléter au mieux les expériences vécues, nous avons fait le choix de ponctuer le texte par des citations assez longues, voire des récits. En contrepartie, nous avons mis en évidence, dans ces citations, les passages particulièrement éclairants en caractères gras.

Les entretiens cités sont pseudonymisés : la majorité de nos interlocuteurs.trices ne souhaitant pas être cité.e.s nommément, nous avons attribué à l'ensemble des personnes interrogées au cours de l'enquête PARDOQ des prénoms factices.

Table des matières

INTRODUCTION	5
LES COMMUNAUTÉS DE LA RECHERCHE QUALITATIVE FACE À L'OUVERTURE DES DONNÉES DANS LE CONTEXTE DE LA SCIENCE OUVERTE	5
ENQUÊTE : PRATIQUES ET EXPÉRIENCES DES COMMUNAUTÉS SCIENTIFIQUES ET TECHNIQUES MOBILISANT DES MATÉRIAUX DE RECHERCHE ETHNOGRAPHIQUES.....	6
MATÉRIEL ET MÉTHODES.....	6
DES ENQUÊTES D'ÉCHELLES MULTIPLES SUR LES PRATIQUES ET LES BESOINS DES CHERCHEURS. EUSES EN MATIÈRE DE GESTION DES DONNÉES	6
L'AUTO-ETHNOGRAPHIE D'UN PROGRAMME DE RECHERCHE AYANT EXPÉRIMENTÉ L'OUVERTURE DE CORPUS ETHNOGRAPHIQUES.....	7
L'ENQUÊTE "PARDOQ" AUPRES DE COMMUNAUTÉS DE RECHERCHE TRAVAILLANT SUR DES MATÉRIAUX ETHNOGRAPHIQUES.....	8
<u>1. APPROCHE CRITIQUE DE L'INTENTION D'OUVRIR LES DONNÉES SCIENTIFIQUES QUALITATIVES : UN BREF ÉTAT DE L'ART.....</u>	10
1.1. L'ATTENTION AUX CONDITIONS DE PRODUCTION DES DONNÉES DE RECHERCHE : UN PRÉREQUIS POUR RENDRE INTELLIGIBLE LA COMPLEXITÉ DU PROCESSUS D'OUVERTURE	10
1.2. L'ANALYSE CRITIQUE DE LA PRODUCTION DES DONNÉES DE RECHERCHE EN SCIENCES HUMAINES ET SOCIALES : ENTRE MODESTIE EMPIRIQUE ET RÉFLEXIONS ÉPISTEMOLOGIQUES.....	11
1.3. PENSER LA TRANSFORMATION DES PRATIQUES ET DU SENS DES PRATIQUES DE PRODUCTION DES DONNÉES DE RECHERCHE À L'ÈRE DU NUMÉRIQUE.....	12
1.4. LA PRATIQUE ETHNOGRAPHIQUE À L'ÉPREUVE DE LA GESTION HÉTÉRONOME DES DONNÉES	14
<u>2. FORMES D'INJONCTION ET INSTITUTIONNALISATION DE L'OUVERTURE</u>	16
2.1. RETOUR SUR LES TRANSFORMATIONS RÉCENTES DU PAYSAGE INSTITUTIONNEL DE LA GESTION DES DONNÉES DE RECHERCHE.....	16
UN CADRE EN CONSTRUCTION ET AUX INTERPRÉTATIONS MULTIPLES.....	16
2.1.1. LE CONTEXTE INTERNATIONAL DU FAIR DATA	18
2.1.2. LA SCIENCE OUVERTE DANS LA LEGISLATION NATIONALE ET L'ARTICLE L533-4 DU CODE DE LA RECHERCHE	21
2.1.3. DES INJONCTIONS RELAYÉES PAR LES PRINCIPAUX BAILLEURS DE LA RECHERCHE ET COLLECTIFS ENGAGÉS ..	21
2.1.4. « AUSSI OUVERT QUE POSSIBLE, AUSSI FERMÉ QUE NÉCESSAIRE » : ARTICULER PROTECTION ET PUBLICISATION DES DONNÉES	23
2.2. LA TRADUCTION DE L'OPEN DATA EN PRATIQUE : FAIRE AVEC UN CADRE EN PLEINE ÉVOLUTION.....	24
2.2.1. LES RÉSEAUX D'APPUI À LA RECHERCHE : UNE ENQUÊTE CENTRÉE SUR L'APPUI SCIENTIFIQUE ET TECHNIQUE ET L'APPUI JURIDIQUE	24
2.2.2. L'APPUI SCIENTIFIQUE ET TECHNIQUE : EXPÉRIENCES DES INGÉNIEUR.E.S EN HUMANITÉS NUMÉRIQUES ET CORRESPONDANT.E.S DE LA TGIR HUMA-NUM	25
2.2.3. L'APPUI JURIDIQUE : EXPERTISE ET EXPÉRIENCES DES DÉLÉGUÉ.E.S À LA PROTECTION DES DONNÉES (DPD)	31

3. ÉTAT DES CONNAISSANCES SUR LES PRATIQUES DES CHERCHEUR.E.S QUALITATIEN.NE.S EN FRANCE	39
3.1. ENSEIGNEMENTS DES ENQUETES RECENTES A ECHANTILLON LARGE SUR LES PRATIQUES DES COMMUNAUTES DE RECHERCHE EN SHS	39
3.1.1. PANORAMA GENERAL	39
3.1.2. SYNTHESE DES RESULTATS DES ENQUETES A LARGE ECHANTILLON SUR LES PRATIQUES DES CHERCHEUR.E.S	41
3.2. ENSEIGNEMENTS D'ENQUETES QUALITATIVES SUR DES PROJETS DE RECHERCHE EN SHS AYANT INTEGRE UN OBJECTIF DE MISE EN PARTAGE DE CORPUS.....	48
3.2.1. UNE TYPOLOGIE DES PROJETS	48
3.2.2. BEQUALI (UMS 828 CENTRE DE DONNEES SOCIO-POLITIQUES) : LES ENQUETES QUALITATIVES A L'EPREUVE DE LA BANCARISATION.....	49
3.2.3. META-PROJET (UMR CITES, TERRITOIRES, ENVIRONNEMENT ET SOCIETES - CITERES) : EPROUVER LES LIMITES DE L'ANALYSE SECONDAIRE	51
3.2.4. SAWA (LABORATOIRE D'ETHNOLOGIE ET DE SOCIOLOGIE COMPARATIVE) : DE LA RESTITUTION COMME IMPERATIF MORAL A L'ACCOMPAGNEMENT DE L'APPROPRIATION COLLECTIVE DES DONNEES.....	52
4. APPORTS D'UNE APPROCHE ETHNOGRAPHIQUE DES PRATIQUES DES COMMUNAUTES DE RECHERCHE QUALITATIVE EN SCIENCES SOCIALES.....	55
4.1. PISTES OUVERTES PAR L'ENQUETE PARCHEMINS	56
4.1.1. LA DECONSTRUCTION DE L'EVIDENCE DU "CYCLE DE LA DONNEE"	56
4.1.2. L'EXPLICITATION DE L'INDETACHABILITE DU TRIPTYQUE : "CONTEXTE-DONNEE-OBSERVATEUR" ET LA QUESTION DE LA "VIOLENCE" FAITE AUX DONNEES	58
4.1.3. LA DEPENDANCE AUX INSTRUMENTS ET SES EFFETS.....	61
4.1.4. L'INTELLIGENCE DES PRATIQUES: PISTES POUR UNE APPROCHE DE L'OUVERTURE DES DONNEES INTEGRANT LES ENJEUX EPISTEMOLOGIQUES ET ETHIQUES	63
4.2 L'ENQUETE PARDOQ : DISPOSITIF D'ENQUETE ET PROFILS DES CHERCHEUR.E.S INTERROGE.E.S	69
4.2.1. IDENTIFIER DES PROFILS VARIES POUR RECUEILLIR UNE DIVERSITE D'EXPERIENCES	69
4.2.2. LE CORPUS D'INTERLOCUTEURS.TRICES.....	69
4.2.3. THEMES DES ENTRETIENS SEMI-DIRIGES ET POSTURE INCLUSIVE	72
4.3 RESULTATS DE L'ENQUETE PARDOQ MENEES AUPRES DES CHERCHEURS.EUSES	73
4.3.1. DIVERSITE DES RAPPORTS A LA "DONNEE"	73
4.3.2. RAPPORTS A "L'OUVERTURE" DES DONNEES	94
4.3.3. RAPPORTS A LA PROTECTION DES MATERIAUX ETHNOGRAPHIQUES : ENTRE ETHIQUE DE RECHERCHE ET CADRE JURIDIQUE.....	110
4.3.4. RAPPORT A L'INJONCTION D'OUVERTURE.....	116
CONCLUSION GENERALE	120
SYNTHESE DES RESULTATS	123
BIBLIOGRAPHIE.....	124

Introduction

Les communautés de la recherche qualitative face à l'ouverture des données dans le contexte de la Science Ouverte

“La science ouverte est la diffusion sans entrave des publications et des données de la recherche. Elle s’appuie sur l’opportunité que représente la mutation numérique pour développer l’accès ouvert aux publications et – autant que possible – aux données de la recherche”. Plan national pour la Science Ouverte, 2018.

Nous avons à ce jour peu de recul sur la manière dont les communautés de recherche en sciences humaines et sociales (SHS) et en particulier les ethnologues se saisissent du mouvement contemporain de “l’ouverture des données” (*Open Data*) dans le contexte de l’avènement de la “Science Ouverte” (*Open Science*) (Galonnier et al. 2019; Ginouvès et Gras 2018). Ce mouvement international, désormais transposé dans l’agenda politique français, se renforce depuis 5 ans à travers d’abord l’adoption de la *Loi pour une République Numérique* (LRN, 2016) puis du *Plan national pour la Science Ouverte* (PNSO, 2018). Il 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, ainsi présentés par la Commission européenne :

«Open Science est le terme choisi par les parties prenantes lors de la consultation publique de 2014 pour décrire les transitions en cours dans la façon dont la recherche est effectuée, les chercheurs collaborent, les connaissances sont partagées et la science est organisée. Il représente une nouvelle approche du processus scientifique basée sur le travail coopératif et de nouvelles façons de diffuser les connaissances en utilisant les technologies numériques et de nouveaux outils de collaboration. En pratique, les sciences ouvertes peuvent rendre la science plus crédible (amplifier l’intégrité scientifique); plus vérifiable (transparence des données); efficace (éviter la duplication des ressources); et répondant aux défis sociétaux (aider à trouver des réponses à certaines des préoccupations majeures de notre temps »¹.

L’engagement politique en faveur d’une science ouverte 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 la recherche (ERC) qui 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” et d’équiper les communautés de recherche dans ce sens. Enfin, certaines revues scientifiques, en particulier dans le contexte des sciences biophysiques, demandent désormais la publication des corpus de données sur lesquels s’appuient les résultats d’analyses publiés.

Il s’agit dans ce rapport, d’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” attribué à la Commission européenne, et repris de toutes parts par les relais institutionnels promouvant l’ouverture des données. Comprendre ce paysage

¹ Source: FAQ: What is openscience?http://ec.europa.eu , citée par Chartron 2018

en pleine évolution nous permettra de 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**. L'objectif de ce rapport est 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 - l'ethnographie -, produisant des matériaux de recherche (ou "données") particulièrement difficiles à caractériser à l'aune du *continuum* entre "ouvert autant que possible et fermé autant que nécessaire".

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 quelles mesures intègrent-ils.elles cet objectif d'ouverture dans leurs pratiques de recherche et quels déplacements cela suscite-t-il ? Comment prennent-ils.elles en compte les spécificités des corpus ethnographiques dans ce contexte ? Enfin, développent-ils.elles et mobilisent-ils.elles de nouveaux outils pour mettre en œuvre cet objectif d'ouverture ?

Enquête : pratiques et expériences des communautés scientifiques et techniques mobilisant des matériaux de recherche ethnographiques

Pour aborder 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 sciences humaines et sociales (SHS), notre étude s'appuie sur l'expérience de personnels scientifiques et techniques dont les travaux mobilisent des matériaux de recherche ethnographiques (chercheurs.euses, enseignant.es-chercheur.euses, doctorant.e.s et réseaux d'appuis scientifique, technique et juridique à la recherche, incluant par exemple des ingénieur.e.s d'étude et de recherche et des délégué.e.s à la protection des données personnelles). Notre hypothèse est que ces matériaux à la fois très qualitatifs (*e.g.* observations, entretiens, archives, photographies, vidéos), de natures et de formats très hétérogènes, et qui embarquent la relation complexe entre chercheurs.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. Qui plus est, la pratique ethnographique est une méthode de collecte de données ancienne, fortement patrimonialisée, et qui constitue une référence partagée pour plusieurs disciplines des sciences sociales.

Matériel et méthodes

Notre étude s'appuie sur plusieurs ressources complémentaires, qui permettent de croiser différents éclairages sur les questions que nous abordons.

Des enquêtes d'échelles multiples sur les pratiques et les besoins des chercheurs.euses en matière de gestion des données

D'abord, 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, les enquêtes se sont multipliées dans les années récentes pour évaluer et caractériser les pratiques des chercheurs.euses (types de données numériques mobilisées, outils de traitement et de stockage, niveau de connaissance des 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. Ces enquêtes sont de plusieurs types et de plusieurs échelles : il s'agit d'enquêtes menées à l'échelle nationale, comme celle du Consortium Couperin (Rousseau-Hans, Ollendorff, et Harnais 2020) ou du projet CommonData (Amiel et al. 2018), à l'échelle d'Universités comme Lille 3 (Prost et Schöpfel 2015), suivie par Rennes 2 (Serres et al. 2017), Bordeaux Montaigne (Duprat 2019) ou Montpellier Paul Valéry (Dillaerts et al. 2020) (présentées en partie 3); ou en interne à l'échelle de laboratoires ou autres collectifs de recherche (comme celle du Cresson : Mosnier, Acquier, et Dom 2019), 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. Quels que soient l'ampleur et le cadrage de ces enquêtes, elles

allient souvent une approche quantitative par questionnaire, suivie d'une qualification plus fine de certains enjeux, pratiques et tendances grâce à des entretiens individuels avec les chercheurs.euses.

Ces enquêtes constituent pour notre étude une ressource importante à deux niveaux : d'abord au titre des résultats qu'elles produisent et des tendances qu'elles mettent en évidence quant aux pratiques, aux attentes des chercheurs.euses, et à leur adéquation avec des objectifs de conformation à des "bonnes pratiques" de gestion des données (à la fois sur un plan juridique et technique) telles qu'énoncées par les services d'appui à l'information scientifique et technique. Ensuite, ces enquêtes nous montrent aussi la montée en puissance de ces préoccupations pour la gestion, la protection et l'ouverture des données, et la manière dont elles se diffusent et prolifèrent aujourd'hui dans les communautés de recherche (justement par le biais-même de ce type d'enquêtes qui constituent en soi un premier levier pour attirer l'attention des chercheurs.euses sur les enjeux qu'elles abordent). 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) - nous y reviendrons.

L'auto-ethnographie d'un programme de recherche ayant expérimenté l'ouverture de corpus ethnographiques

*Parchemins : Paroles et chemins de l'agriculture littorale*², est un programme de recherche interdisciplinaire et transdisciplinaire initié en 2016, et actuellement en cours de finalisation (début 2021). Visant à documenter et analyser les transformations de l'agriculture dans les zones littorales bretonnes, ce programme présente une double spécificité dans son approche : la première est de considérer l'ouverture de la fabrique de la recherche et la publication des données qualitatives, principalement issues d'enquêtes ethnographiques sur cinq sites d'étude, comme un outil de renforcement de la qualité des connaissances produites et d'implication des acteurs sociaux dans un questionnement sur la place de l'activité agricole dans les territoires littoraux (transdisciplinarité). La seconde est d'allier ethnologie, sociologie, sciences agronomiques et environnementales, sciences et technologies de l'information et de produire un effort important d'intégration des connaissances, par un travail autour de la mise en forme et de la mise en perspective de données quantitatives, qualitatives et spatiales. Cette attention à l'ouverture des données a conduit l'équipe de Parchemins à concevoir et expérimenter différentes architectures de publications de données qui facilitent cette interdisciplinarité.

L'expérimentation mise en place par cette équipe de recherche³ s'appuie sur l'expérience acquise dans le domaine de la publication et de la spatialisation des données environnementales par l'UMR Sols, Agro et hydrosystème, Spatialisation (SAS). Cette configuration a donné deux orientations particulières à l'expérience vécue par cette équipe : d'une part, la recherche de transposition et d'extension de cette expertise des infrastructures de données spatiales aux spécificités de données qualitatives, tels que des corpus ethnographiques. D'autre part, une posture volontariste de la part des membres de l'équipe de recherche, liée à ce contexte expérimental, qui a permis de soutenir l'effort de mise en pratique de l'ouverture des données, même lorsque les circonstances étaient difficiles, dans la perspective de faire aboutir cette expérimentation. Par ailleurs, cet effort bénéficiait de l'existence au sein de l'équipe d'une forte "culture du libre" (en particulier chez les ingénieurs en STI), qui renforçait l'intérêt à expérimenter de nouvelles frontières de l'ouverture de données.

Les processus conjoints de co-construction d'une instrumentation numérique et de mise en oeuvre de formes de partages et d'ouverture de corpus ethnographiques ont fait l'objet d'une auto-ethnographie, sur

² Le consortium du projet Parchemins, porté par l'Institut national de la recherche agronomique (INRA, devenu depuis INRAE) impliquait initialement à titre principal le Laboratoire Interdisciplinaire Sciences, Innovations, Sociétés (UMR 1326 LISIS – Marne-la-Vallée) et l'UMR Sol, agro et hydrosystème, spatialisation (UMR 1069 SAS – Rennes), qu'a rejoints en 2018 l'UMR Aménagement des Usages des Ressources et des Espaces marins et littoraux (UMR 6308 AMURE – Plouzané).

³ dont les deux auteures de ce rapport sont membres : Alix Levain étant la coordinatrice de Parchemins, et Florence Revelin, ingénieure de recherche en anthropologie.

la durée du projet, visant à documenter les traces des différentes étapes de ces processus et la manière dont ils ont été investis et vécus par les membres du programme. Le présent rapport s'appuie en partie sur ce riche matériau, exposé plus en détails en partie 4.

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

Nous avons mené entre septembre et décembre 2020 une enquête - dite "enquête PARDOQ"-, sur la base d'entretiens semi-dirigés visant à recueillir les expériences de quatre types d'interlocuteurs : des prescripteurs de la science ouverte, des membres de réseaux d'appui à la recherche, des chercheur.euses et enseignant.e.s-chercheur.euses pratiquant l'ethnographie, et des membres de projets institués d'archivage numérique et de partage de corpus qualitatifs. 26 entretiens ont été réalisés au total avec une représentation asymétrique des différentes catégories d'interlocuteurs (voir [FIGURE 1](#)) et une concentration accrue des recueils d'expériences des praticien.ne.s de la recherche et des réseaux d'appui, à la fois techniques et juridiques:

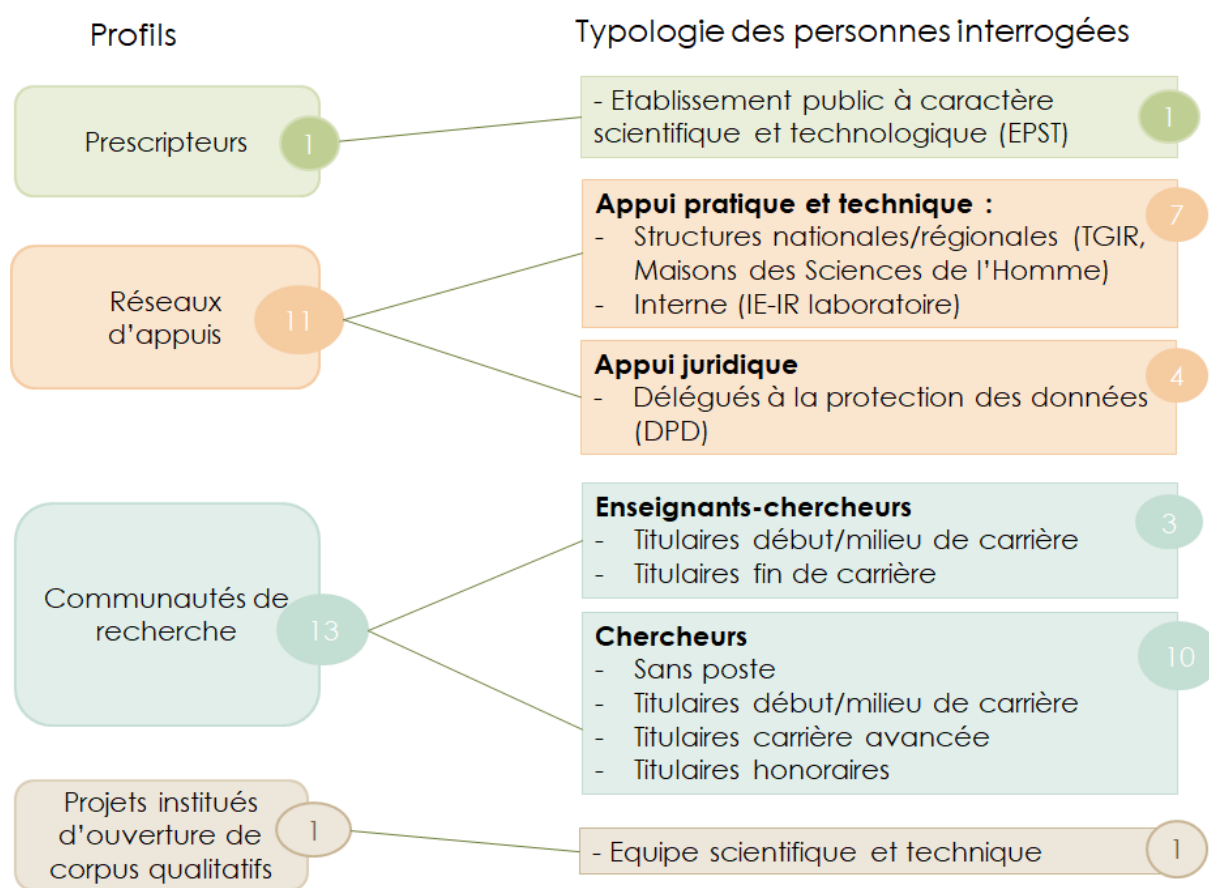


Figure 1. Profils des personnes interrogées dans le cadre de l'enquête PARDOQ.

L'enquête a porté sur les différents rapports de ces interlocuteurs au travail et aux spécificités des matériaux de recherche ethnographiques, sur les formes de partages dans lesquelles ils sont engagés le cas échéant, et sur leurs rapports aux politiques et injonctions 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.trices s'équipent pour traiter ces matériaux et éventuellement les partager ou au contraire les protéger.

1. Approche critique de l'intention d'ouvrir les données scientifiques qualitatives : un bref état de l'art

1.1. L'attention aux conditions de production des données de recherche : un prérequis pour rendre intelligible la complexité du processus d'ouverture

Depuis une trentaine d'années, l'étude sociale des sciences s'est considérablement enrichie d'approches critiques tendant à décrire la science non pas à partir des énoncés qu'elle propose *ex post*, mais davantage à partir de l'analyse de la façon dont ces énoncés sont construits. Cette analyse se fonde sur l'examen des activités concrètes par lesquelles la recherche, en tant qu'assemblage d'institutions, de professionnels, d'objets et de flux, construit, établit et consolide des *faits*. Le travail scientifique apparaît alors comme un ensemble d'activités visant à "sécuriser des ressources, conduire des recherches, construire des arguments, ouvrir (et protéger) des espaces pour l'enquête, et produire des preuves suffisamment convaincantes pour passer le cap de la revue par les pairs et aboutir à une publication." (Hackett et al. 2017).

Toute réflexion sur les processus d'ouverture des données oblige ainsi à se pencher sur la façon dont les données se fabriquent et dont elles constituent, en elles-mêmes, le résultat de ces activités très quotidiennes, de ce travail qui enrôle et engage des instrumentations techniques, mais également des dispositifs et des conventions professionnelles et, plus largement, des conventions sociales. La production des données apparaît alors comme une construction, et les données davantage comme des construits, des "obtenus" (Latour 1993), des "capta" (Drucker 2011), que comme le matériau brut à partir duquel le savoir scientifique s'élabore : comme le rappellent Denis et Goëta (2016), les données ne sont pas "passivement récoltées, mais façonnées" [p.25]. Ainsi conçues, les "données" incorporent le monde social dans lequel celui ou celle qui les produit évolue, et les représentations que celui ou celle-ci se fait de ce monde, des principes et valeurs qui l'organisent.

Les travaux fondamentaux dans le domaine Sciences, techniques, Sociétés (STS), ont principalement porté sur le domaine des sciences dites exactes. Ces travaux ont mis en évidence à quel point les processus de validation des connaissances scientifiques étaient historicisés, situés. Ils ont par exemple mis en évidence la succession de régimes d'administration de la preuve reconnus comme valides (Shapin et al. 1993). Le travail probatoire comporte ainsi une dimension sociale, et ne peut être considéré comme indépendant du processus de production des "données" scientifiques elles-mêmes, en ce sens que celui-ci est orienté par des fins, des règles et des standards (de présentation, de format, d'interprétation etc.) que le producteur de ces données a été socialisé pour maîtriser. Dans leur livre consacré à la construction de l'objectivité scientifique, Daston et Galison (2010) montrent ainsi que les scientifiques portent des vertus épistémiques qui sont "internalisées et respectées en faisant référence de manière constante à des valeurs éthiques et une efficacité pragmatique. Ces normes ont pour objectif de rendre assuré le savoir scientifique" [p.40]. Ces auteurs questionnent ainsi la pertinence d'aborder l'histoire des sciences comme une histoire cumulative, marquée par l'apprentissage de capacités supérieures d'objectivation, et insistent sur l'importance de s'attacher à rendre compte de la "construction du moi scientifique", à l'échelle individuelle mais également collective, pour mettre en évidence les formes d'objectivation qui s'imposent comme légitimes dans un contexte donné. Ils insistent également sur **l'hétéronomie constitutive de la production des données, c'est-à-dire sur le caractère politique de la mise en base de données**, faisant écho à la définition des données proposée par L. Gitelman et V. Jackson, dans l'introduction qu'elles donnent à l'ouvrage collectif *Raw Data is an Oxymoron* (Gitelman et Jackson 2013) :

"Data are units or morsels of information that in aggregate form the bedrock of modern policy decisions by government and nongovernmental authorities. [...] At first glance data are apparently before the fact: they are the starting point for what we know, who we are, and how we communicate. This shared sense of starting with data often leads to an unnoticed assumption that data are transparent, that information is self-evident, the fundamental stuff of truth itself."[p.2]

1.2. L'analyse critique de la production des données de recherche en sciences humaines et sociales : entre modestie empirique et réflexions épistémologiques

S'appuyant le plus souvent d'études de cas issues des sciences dites dures, le travail de déconstruction de la notion de "donnée" réalisé au sein des études sociales des sciences permet de mettre en évidence **à la fois les similitudes et les singularités de la production de données en sciences humaines et sociales, ainsi que les rapports complexes qu'entretiennent leurs producteurs avec celles-ci.**

La modicité même de ces études empiriques peut faire l'objet de plusieurs types d'interprétations. Nous souhaiterions insister ici sur l'une d'entre elles : le fait que, d'une certaine façon, la réflexion autour des SHS et, en particulier, de la nature des matériaux qu'elles mobilisent, a plutôt fait historiquement l'objet de **réflexions épistémologiques au sein même des disciplines concernées, que d'enquêtes empiriques dans ce domaine.** C'est particulièrement notable pour le matériau de recherche qui ne relève pas de la construction ni de la remobilisation de séries statistiques, c'est-à-dire du **matériau faiblement standardisé qui constitue le fonds à partir duquel s'élabore la recherche qualitative en sciences sociales et ses méthodes** (entretien de type qualitatif, observations en situation, etc.), qui de façon générale assume sur le plan épistémologique la dimension interprétative consubstantielle de la production des savoirs scientifiques, sans toujours toutefois en expliciter toutes les implications (Freund 1978). Il résulte de cela que les enquêtes en SHS comportent généralement une forte dimension réflexive (enquête sur soi et sur *d'où on parle*) et ne peuvent être dégagées d'une réflexion sur les différents paradigmes qui sous-tendent la diversité des approches en SHS. De ce point de vue, **une première distinction peut-être simplement opérée entre perspective objectivante et perspective subjectiviste.** Dans une perspective objectivante, fondée sur la possibilité et l'objectif d'une possibilité de disjonction, de découplage entre le chercheur et l'information qu'il produit, comme condition de la validité de la science, perspective qui est par exemple très présente dans certaines traditions de l'économie, de la sociologie ou de la géographie. Il est alors possible, de façon très similaire à la démarche qui prévaut au sein des sciences biophysiques, d'envisager un détachement de l'objet de recherche et sa symbolisation par une quantification. **Ce détachement ne s'opère pas de la même façon dans une perspective subjectiviste, dans laquelle l'attachement s'opère et se pense à deux niveaux. Le premier niveau est celui du caractère contextuel et situé de la connaissance produite,** et donc du fait que sa valeur dépend de la qualité de l'énoncé de ses conditions de productions et de validité, forcément précaires et limitées dans le temps. Au-delà de cette indissociabilité donnée-contexte, **le 2ème niveau est celui de l'indissociabilité entre ce couple et la lecture qu'en propose l'observateur ou l'observatrice ;** dit autrement, le travail interprétatif est considéré comme consubstantiel, incorporé dans le matériau de recherche, ce qui rend fortement sujette à caution la pertinence, voire la possibilité, d'une remobilisation par d'autres que ceux ou celles qui l'ont produit. C'est ainsi cette perspective qui, selon Max Weber, fonde la spécificité de l'épistémologie des sciences sociales, comme "sciences du contexte", historiques et situées.

Le primat de l'une ou l'autre de ces perspectives correspond à des moments historiques importants et souvent conflictuels dans la structuration de certaines des disciplines et sous-disciplines des SHS, même si elles coexistent souvent dans la pratique des chercheurs et des chercheuses. Malgré cette hybridité et cette diversité des pratiques, il faut considérer dès lors que le processus d'ouverture des données de recherche pose des questions d'ordre fondamentalement épistémologique et méthodologique aux sciences humaines et sociales. Celles-ci portent, en particulier, sur la reproductibilité et la vérifiabilité de la science (Galonnier et al. 2019).

Ainsi, si peu d'études de cas ont été consacrées en STS à la recherche en sciences humaines et sociales, en revanche, de nombreuses d'études ont établi, dans une diversité de contextes, le caractère subalterne de certains de leurs paradigmes structurants dans la façon dont les politiques des sciences étaient conçues et mises en oeuvre. C'est le cas de la politique d'*Open Science* (Stérin et Nous 2019), et en particulier de l'*Open Access* et des politiques d'évaluation qui se développent en prenant appui sur celui-ci (voir par exemple : Dassa 2019).

1.3. Penser la transformation des pratiques et du sens des pratiques de production des données de recherche à l'ère du numérique

Le déploiement à large échelle des technologies numériques stimule de nouvelles méthodes de recherche, en même temps qu'il produit un renouvellement des cadres d'exercice de l'activité scientifique, qui touche l'ensemble des disciplines et des objets de recherche (Millerand 2011; Borgman 2012; Bourdeloie 2014).

Comme le soulignent Denis et Goëta (2016), l'ouverture des données publiques implique des séries d'opérations complexes comme des enquêtes permettant l'identification, au sein de chaque institution, de ce qui peut être considéré comme des "jeux de données", des processus d'extraction (souvent longs et contrariés) consistant à détacher les données des systèmes d'information dans lesquels elles sont inscrites, et enfin des opérations attentionnées de "nettoyage" - de "brutification", écrivent Denis et Goëta - visant à leur "assurer une double intelligibilité, humaine et technique". Les investissements de forme (Thévenot 1986) que requièrent l'effort des institutions publiques pour "rendre les données présentes", c'est-à-dire visibles, stables, circulables sont donc considérables (Leonelli, Rappert, et Davies 2017). Leur observation révèle les normativités épistémiques incorporées dans les instruments.

La multiplication des infrastructures techniques (qu'elles se nomment archives, entrepôts ou portails) venant appuyer ce mouvement a été maintes fois soulignée ces dernières années, créant des effets de structuration et des irréversibilités de nature à générer un durcissement des contraintes et une orientation par la donnée des activités de recherche dans les années à venir, allant jusqu'à une réallocation structurelle des moyens dévolus à ces activités (Galonnier et al. 2019; Jarrige, Le Courant, et Paloque-Bergès 2018). Conjugée à un renforcement de la recherche sur financement et des injonctions à l'ouverture, elle peut également conduire à (devoir) penser le réusage de la donnée avant même sa collecte et à qualifier et informer ce réusage en ignorant les propriétés d'émergence des données. L'enrôlement de communautés d'appui à la recherche dans ces déploiements alimente une forme d'"impératif technologique" qui tend à invisibiliser la dimension politique du mouvement d'ouverture. Ce mouvement ancien, qui s'est appuyé sur des arguments mettant en avant le caractère de bien commun du savoir scientifique est d'ailleurs traversé par des tensions, puisqu'une partie de ses promoteurs revendiquent une conscience aiguë des risques associés à la marchandisation des données (Wyatt 2017).

Car, si comme le rappelle F. Millerand (2011), la vision des données comme bien commun fait écho à la thèse mertonienne d'une science moderne fondée sur le partage des données et s'inscrit ainsi dans une forme de continuité avec les représentations dominantes du travail scientifique et de la « bonne » science, les comportements de fermeture et de protection des données sont constitutifs du processus même de découverte scientifique à l'échelle individuelle et collective (Biagioli 2012). En effet, la confidentialité agit comme un « sas » protecteur permettant de gérer une phase d'instabilité inhérente à la validation par les pairs lors d'un processus de porter à connaissance qui passe par l'élaboration de stratégies sophistiquées de publication. Le numérique bouleverse la façon dont le partage des données est envisagé : la structure informationnelle des matériaux de recherche dont la bancarisation et l'ouverture sont encouragées et organisées au sein de grandes infrastructures numériques « conduit les chercheurs à se poser de nouvelles questions sur la nature de leurs données et sur les possibilités de leur réutilisation par d'autres » (Millerand 2011 : 233).

Dans ce contexte, l'*Open Data* et ses appuis institutionnels apparaissent davantage aux observateurs critiques comme un mouvement externe à la science, tendant à un renforcement au nom de la "transparence" du contrôle du travail des chercheuses et chercheurs, davantage qu'il ne témoigne d'un "retour aux sources" qui serait issu d'une dynamique interne au champ scientifique, bien que les vertus épistémiques de cet éventuel retour soient mises en avant par certains d'entre eux (Galonnier et al. 2019). **Or, ces enjeux de transparence et de contrôle ne sont pas du tout posés de la même manière au sein des champs de recherche qui assument une dimension interprétative comme consubstantielle de la production du matériau de recherche, qu'au sein des disciplines** (comme la psychologie par exemple) **où se pose un enjeu de répliquabilité et de lutte contre la fraude.** C'est ainsi la dimension normative de ce mouvement qui est mise en lumière, dans la mesure où il est susceptible d'orienter des choix à l'aune, non

d'un renforcement du contrôle interne par les pairs, mais plutôt de l'introduction d'un contrôle bureaucratique standardisé de l'activité scientifique. **Le risque d'une bureaucratisation de la recherche est ainsi régulièrement mis en exergue, qui conduirait à "inscrire durablement de bonnes raisons de produire de mauvaises données"** (Weller 2019 à propos de Denis 2018).

Quel sens, dans ce contexte, donner à la notion d'"ouverture"? Car c'est davantage à un double mouvement d'ouverture-fermeture, de mise en visibilité et d'invisibilisation de ce qui construit leur savoir et leur légitimité (Leonelli, Rappert, et Davies 2017), que se confrontent les praticiens de la recherche. Cette mise en ordre et cette standardisation des pratiques, dont la continuité avec les normes disciplinaires et la socialisation professionnelle des praticiens mérite d'être interrogée, en ce qu'elle les conduit à se confronter, dans un même mouvement et selon une même temporalité, **deux dimensions du rapport aux données pourtant issues de dynamiques et traduites dans des dispositifs juridiques distincts : celle de la protection et celle du partage.**

Dans ce contexte, les travaux les plus récents consacrés aux relations complexes qu'entretiennent le mouvement d'ouverture des données et la construction des savoirs scientifiques insistent sur le caractère contingent de leur articulation et appellent à une attention soutenue aux pratiques micro-institutionnelles, à l'échelle des collectifs de recherche et des laboratoires, pour comprendre la façon dont les acteurs et actrices de la recherche s'engagent - ou pas - dans le processus d'ouverture (Levin et Leonelli 2017). **Ce resserrement du regard au plus près des acteurs sociaux de la recherche, des cadres dans lesquels ils exercent leur activité et des interactions dans lesquels ils se trouvent engagés, implique notamment d'observer avec la même acuité tous les collectifs engagés dans ce processus, qu'ils conçoivent les politiques d'ouverture, les relaient, les promeuvent, et/ou qu'ils en soient les destinataires, qu'ils s'en arrangent ou s'en maintiennent à l'écart, les contestent éventuellement.**

De ce point de vue, la complexité croissante du cadre juridique applicable à la diffusion des données de la recherche rend son appropriation directe par les praticiens de la recherche particulièrement malaisée (Stérin et Noûs 2019), et s'appuie très largement sur l'investissement et la transformation des priorités et des métiers d'appui à la recherche, dans un contexte où la gestion des données en SHS n'a jusqu'à une période récente ni constitué une question largement discutée au sein des communautés de recherche, ni été prise en compte dans les moyens alloués aux laboratoires *via* les dotations des tutelles ou les financements de projets (pour une critique de cette situation, voir : Weil et Cases 2012). Les politiques de gestion des données s'inscrivent en effet dans un bouleversement plus large suscité par la généralisation de l'instrumentation numérique dans les milieux académiques et dont les implications politiques restent encore à mesurer, pour "affiner l'hypothèse des enjeux de pouvoir et de la « normativité » à l'oeuvre dans les logiques d'instrumentation numérique en abordant les dispositifs au croisement de leurs « lignes de force » et de leurs « lignes de fuite » (Bigot 2018 : 311).

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

Ce "moment "données"" 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 numérique (Bigot 2018; Galonnier et al. 2019). 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 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 sciences sociales 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 rendent certaines recherches impossibles en raison du caractère sensible et subversif de certaines enquêtes et de l'incompatibilité entre respect de l'éthique professionnelle et partage (voir par exemple, à propos d'une enquête au Front National: Bizeul, 2020).

Les travaux qui rendent compte de ce contexte "post RGPD" et "post LRN" sont encore peu nombreux, en particulier pour ce qui concerne la recherche qualitative en sciences sociales, pour laquelle l'articulation entre ouverture des données (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 applicable (Maurel 2018). 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 cœur de la matière empirique à partir de laquelle de nombreux sociologues, des politistes et, dans une moindre mesure, des géographes, construisent leurs analyses. Ils et elles revendiquent, par la pratique ethnographique et l'affirmation de sa pertinence, un certain type de rapport à l'enquête et à la collecte de données, et une forme d'engagement (Cefaï 2010). Daniel Cefaï propose dans l'introduction et la postface qu'il donne au recueil de textes *L'engagement ethnographique* une lecture de cet engagement par et dans la pratique en distinguant 3 temps :

- l'indexation du texte ethnographique sur des comptes-rendus d'événements et des récits d'acteurs sociaux réalisés sur le "terrain", au cours d'opérations d'observation et d'enregistrement menées par l'ethnographe ;
- le choix du ou des sites d'étude et les formes de présence que l'ethnographe y privilégie, qui se construit en fonction d'un questionnement toujours ouvert et prêt à être formulé en cours d'enquête;
- la restitution et la transmission des résultats auprès de divers publics, associée pour Cefaï au travail de publication, entendu au sens large de publicisation des résultats et non au sens étroit de production de textes académiques.

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 disciplinaires (Katz 2001; 2002).

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'ethnographes 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.

Ainsi, si les cadres éthiques de la discipline sont questionnés de longue date notamment à partir de ce qu'on pourrait nommer des "situations limites" ou à risque (Deschamp 2018), ayant historiquement donné lieu à des abus (par exemple, l'exploitation du savoir ethnographique dans un contexte de conflit ou à des fins de renseignement stratégique, l'observation participante clandestine, la collecte d'artefacts sans consentement éclairé des communautés auxquelles ils appartiennent...). Ce sont ainsi les pratiques de terrain qui sont au cœur de la codification éthique propre aux disciplines qui pratiquent l'ethnographie qui a émergé tardivement en France pendant la décennie 2000 (Cefai et Costey 2009), mais bien plus précocement, à partir des années 1960, dans d'autres pays où la critique postcoloniale et féministe était plus développée, en particulier aux Etats-Unis et dans les pays du Commonwealth). Cette longue histoire met en évidence l'intensité des tensions qui accompagnent l'institutionnalisation des cadres éthiques; dans bien des cas en effet, la façon dont ces cadres s'énoncent rend de fait la pratique impossible:

"Le problème de base des IRB (Institutional Review Boards) est la mise hors la loi d'une bonne partie de la recherche ethnographique, laquelle, par définition, n'a pas de conception (design) préétablie. La plupart des interactions sur le terrain ne sont pas perçues, au moment où elles sont accomplies, comme des moments de recueil de données : faut-il dès lors exclure toute expérience et toute mémoire autobiographique de l'enquête ? Comment faire quand la plupart des rencontres sur le terrain sont contingentes, non prévisibles, qu'elles relèvent à la fois de la vie personnelle et de la méthode d'investigation, et que c'est d'elles que dépend la constitution même de l'objet d'enquête ? Et au-delà, faut-il renoncer à enquêter sur les activités immorales, illégales et criminelles, qui échappent aux dispositifs d'enquête formelle et quantitative – abandonner toute forme d'observation, directe ou participante, sur les radicalismes religieux ou politiques, sur la criminalité en col blanc, la délinquance, le narcotrafic ou la prostitution ? Faut-il s'abstenir de toute enquête dans des régimes autoritaires, par crainte des usages politiques, policiers ou militaires qui pourraient en être faits ? Et sous le prétexte de ne pas nuire aux enquêtés, individus ou institutions, et de réguler des pratiques d'enquête, ce qui est en soi louable, faut-il bafouer le droit à l'enquête, qui est constitutif de l'expérience démocratique ?" (Cefai et Costey 2009: 20).

Ce que soulignent les expériences nord-américaines auxquelles D. Cefai fait référence dans cette contribution à l'élaboration d'une Charte éthique au sein de l'Association française de sociologie, c'est ainsi que les formes déclaratives a priori, les processus de standardisation et la centralisation heurtent frontalement ce type de pratique et la fragilisent. **On comprend bien, ainsi, comment il est possible que des ethnologues aient été parmi les plus rapides en France pour investir des outils de publication numérique d'archives à destination des communautés de recherche en SHS** (Mouton 2001; Abélès et Mouton 2008; Molinié et Mouton 2008) **et à penser la question de la restitution en lien avec la numérisation** (Mouton 2018; Ginouvès et Gras 2018), tout en étant singulièrement peu diserts sur les apories auxquelles l'évolution du droit et l'organisation institutionnelle de la gestion des données les expose. L'affirmation d'une nécessité de contrôle externe sur les conditions dans lesquelles le partage et la protection des données de recherche s'effectue change les termes du débat et les acteurs qui y prennent part. Or, le temps long de la délibération collective, de la mise en perspective de la portée éthique et politique des choix, heurte aujourd'hui celui de l'interprétation et de la mise en œuvre dans l'urgence de la règle de droit, et celui, plus incertain, du déploiement des usages et réusages des matériaux numérisés.

2. Formes d'injonction et institutionnalisation de l'ouverture

Dans cette seconde partie, nous proposons de mettre en perspective les évolutions récentes de l'encadrement des pratiques de gestion des données de la recherche en France. A partir d'une analyse documentaire, nous les replacerons dans un premier temps dans le mouvement plus large d'ouverture des données et en explorerons les formes les plus concrètes. Nous nous attacherons ensuite, en nous appuyant sur les entretiens semi-directifs et les observations ethnographiques réalisées entre septembre 2020 et janvier 2021, à rendre compte de l'expérience des communautés d'appui à la recherche qualitative qui se trouvent engagées dans ce mouvement, en particulier à travers l'exemple des ingénieurs en humanités numériques des Maisons des Sciences de l'Homme et les délégués à la protection des données.

2.1. Retour sur les transformations récentes du paysage institutionnel de la gestion des données de recherche

2.1.1. Un cadre en construction et aux interprétations multiples

Le projet transnational d'une "Science Ouverte" agrège principalement deux mouvements majeurs que sont l'ouverture des publications scientifiques (*Open Access* - OA), et l'ouverture des données publiques (*Open Data* - OD) : la conjonction de l'OA et de l'OD peut être considérée comme le fait générateur d'un mouvement d'ouverture plus spécifique à la recherche publique (*Open Research Data* - ORD). Les temporalités et formes de déploiement de ces deux volets sont différentes : ils suivent des trajectoires complexes qui impliquent une diversité d'acteurs, de perspectives, de motivations et d'interprétations.

Le premier (l'OA) a émergé dans les années 1990 au sein de communautés de chercheurs en physique théorique, et sa trajectoire d'émergence n'est pas homogène, comme le souligne Moore (2017) : pour schématiser, il se déploie d'une part en réaction au monopole exercé par les éditeurs privés de revues scientifiques, et d'autre part dans le contexte de la "culture du logiciel libre" ou "culture du libre". Chartron (2014) explique ainsi que "l'OA a connu des phases différentes, les enjeux n'ayant cessé d'être reformulés par différents groupes d'acteurs dont la vision tend aujourd'hui à ne plus converger véritablement", voire à se polariser entre d'une part une vision libérale portée par les politiques européennes et d'autre part, une idéologie des savoirs-communs de la connaissance défendue par des groupes de chercheurs et chercheuses et des bibliothèques militantes (*ibid.*). Ce mouvement est généralement considéré comme trouvant sa source dans les dynamiques de consolidation éthique engagées à partir de certains champs de recherche dans les années 1990, en particulier le domaine de la recherche biomédicale concernée par le développement et le partage des données génomiques. Les principes directeurs dits "des Bermudes" (1996), l'Initiative de Budapest (Initiative de Budapest pour l'Open Access 2002⁴), la Déclaration de Bethesda (2003)⁵ et la Déclaration de Berlin⁶ constituent aujourd'hui le socle de la définition dite "BBB" du libre accès, à savoir **l'accès gratuit, en ligne, immédiat et permanent à la version intégrale d'un article scientifique ou académique.**

Ce mouvement a connu un essor majeur depuis les années 2010⁷. Les pratiques de publication ouverte des résultats formalisés de recherche sont désormais avancées (Masclat De Barbarin 2018), même si l'objectif

⁴ <https://www.budapestopenaccessinitiative.org/>

⁵ https://www.ouvrirlascience.fr/wp-content/uploads/2018/11/D%C3%A9claration-Bethesda_fr.pdf

⁶ <https://openaccess.mpg.de/2365/en>

https://oa2020.org/wp-content/uploads/pdfs/14thBerlinOpenAccessConference_Report.pdf

⁷ Les politiques de dépôt en accès ouvert mises en œuvre par les établissements de recherche, de documentation et d'enseignement supérieur ou leurs subdivisions sont recensées au sein d'un registre numérique coopératif intitulé ROARMAP, qui en recense à ce jour plus de 1000. L'expansion la plus importante sur la décennie 2010 concerne les politiques construites à l'échelle même des établissements. Source : <http://roarmap.eprints.org/>, consulté le 27 avril 2021.

de 100% des publications en accès libre des recherches financées publiquement pour 2020 (fixé par exemple par le Président directeur général du CNRS) n'est pas atteint à ce jour (Chartron et Schöpffel 2017). Le mouvement vers l'OA est ainsi fortement marqué par **une vision de la science dont le caractère de bien commun est menacé par l'intégration des publications scientifiques dans une logique de marché.**

C'est aussi une vision de la science comme bien commun qui est à l'œuvre dans le mouvement d'ouverture des données de la recherche (ORD). Schématiquement, **on peut considérer que l'ORD s'inscrit dans une double généalogie, externe au champ scientifique, et interne à celui-ci.** La dynamique exogène est celle de l'ouverture des données administratives et du renforcement des droits des citoyens vis-à-vis de l'administration (accès aux documents administratifs, ouverture des archives...), observable en France à partir des années 1970. La dynamique plus endogène, d'autre part, est celle de l'OA, dans la mesure où la logique de l'OA poussée à son terme consiste en une accessibilité directe au public des données de recherche fondant les résultats rendus publics par les pratiques d'auto-archivage ("*green route*" ou "voie verte") et de publication dans les revues dites "alternatives", nativement en accès libre ("*gold route*" ou voie "dorée").

Le déploiement de l'ORD dans le domaine spécifique de la recherche publique est cependant plus récent et se heurte plus intensément à la grande hétérogénéité des contextes de fabrication de la recherche scientifique - qui implique une hétérogénéité tout aussi grande de types de données et de rapports à ces données - et soulève des problématiques complexes au moment de sa mise en œuvre. Par contraste avec l'OA, le mouvement de l'ORD n'en est qu'à ses débuts et le cadre de son déploiement est en construction (Masclat De Barbarin 2018). En témoignent la multiplication en France, sur une très courte période (2018-2021), des documents stratégiques conçus à différentes échelles⁸, mais qui ont tous en commun de contribuer à l'émergence et à l'affirmation d'une politique polycentrique d'administration des données de la recherche publique⁹.

Selon Paschetto, Sands, et Borgman (2015), deux propriétés majeures de l'ORD émergent des définitions que ces auteurs ont pu recenser grâce à une analyse bibliographique exhaustive : la disponibilité "légale" et "technique" des données. Cela souligne la subtilité de la mention "autant que possible", si problématique dans l'interprétation des directives du PNSO pour les communautés de recherche qui entreprennent de s'y conformer et de s'engager dans un projet d'ouverture par défaut : quelles sont les données de leurs corpus qui doivent et peuvent être rendues publiques et comment ? Dans ce contexte, on voit fleurir, en particulier depuis l'adoption du PNSO en 2018, une multitude de guides, d'enquêtes, d'ateliers pratiques, de tutoriels, de formations ou d'outils abordant la question de la gestion des données et de la mise en œuvre de leur ouverture sur un plan principalement juridique, technique et pratique - nous y reviendrons dans le rapport.

En continuité avec l'impulsion donnée par l'OA, le mouvement de l'ORD a pris forme dans la convergence entre d'une part des valeurs et principes de démocratisation de l'accès à la construction des savoirs scientifiques, d'autre part des objectifs de transparence prenant appui sur des enjeux d'accumulation liée aux réutilisations des données plutôt qu'à leur reproduction, mais aussi des enjeux plus controversés, tels que la vérifiabilité et la reproductibilité des opérations scientifiques (Desquilbet et al. 2019).

Mais tout comme son *alter ego* (l'OA) auquel l'ORD appliqué à la recherche publique est presque systématiquement associé, les définitions et les enjeux auxquels l'associent divers groupes d'acteurs sont hétérogènes et parfois contradictoires (Paschetto et al. 2016). Cette convergence de cadrages et d'agenda s'inscrit dans un contexte commun : celui de l'intensification de la production de données numériques, reliée

⁸ Parmi les plus connus : le Plan national pour une Science Ouverte (2018), le Plan données de la recherche du CNRS (2020). Le Coso (Comité pour la Science Ouverte, recense) ainsi 196 références liées à la mise en œuvre de la science ouverte, dont 18 documents de politiques publiques et 22 guides. Parmi ces 196 ressources inventoriées, 137 sont publiées entre 2018 et 2021 (https://www.ouvrir.lascience.fr/category/ressources/?type=politiques_publicques, consultée le 29/04/2021)

⁹ Voir par exemple, la *Feuille de route du CNRS pour la science ouverte* (2019) : https://www.cnrs.fr/sites/default/files/press_info/2019-11/Plaque ScienceOuverte.pdf; et, pour l'INRAE, les *Principes pour la gouvernance des données* (2020) : <http://geowww.agrocampus-ouest.fr/web/wp-content/uploads/2021/02/20201210-INRAEGouvernancedesdonnees%CC%81es.pdf>

à la numérisation croissante des activités et productions scientifiques (Le Marec et Mairesse 2017). Comme le relève S. Pouyllau, directeur technique de la TGIR Huma-Num, dans l'éditorial qu'il donne en 2018 à l'ouvrage *La diffusion numérique des données en SHS - Guide de bonnes pratiques éthiques et juridiques* (Ginouvés et Gras 2018), les sciences humaines et sociales, dans leur diversité, sont toutes concernées par ce mouvement:

« Ce qui est commun aujourd'hui, c'est l'ampleur de la dimension numérique des dispositifs de collecte (de l'enquête sur les terrains, au dépouillement de fonds d'archives, en passant par des relevés à l'aire de drones), de la façon d'organiser les corpus, de l'instrumentation d'analyses et enfin de la façon d'envisager le partage des données, les chemins pour travailler les données sont multiples » (Pouyllau 2018: 8)

L'ouverture des données telle que promue par le Plan national pour la Science Ouverte (2018) implique de **publier sur des entrepôts de données en accès ouvert**, c'est-à-dire sous licence ouverte telle que celle proposée par *Etalab* par exemple (Licence Ouverte V 2.0, Etalab 2017) ou par *Creative Commons* (CC), tous les matériaux de recherche produits sur la base de financements publics et non soumis à des régimes de protection juridique (Stérim et Noûs 2019). Ce principe d'ouverture par défaut est que, sauf exception légitime, aucune donnée n'est vouée à rester cloisonnée et protégée dans un outil de stockage privé ou fermé. Chaque donnée doit être mise en circulation libre sur le web dans le respect des principes FAIR 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.

2.1.2. Le contexte international du FAIR Data

Les principes FAIR ont été formalisés à partir de 2014 par un collectif hybride œuvrant pour la diffusion numérique des savoirs scientifiques, à l'occasion d'un atelier à Leiden (Pays-Bas):

"From these discussions, the notion emerged that, through the definition and widespread support of a minimal set of community-agreed guiding principles and practices, data providers and data consumers - both machine and human - could more easily discover, access, interoperate, and sensibly re-use, with proper citation, the vast quantities of information being generated by contemporary data-intensive science. These simple principles and practices should enable a broad range of integrative and exploratory behaviors, and support a wide range of technology choices and implementations, just as the Internet Protocol (IP) provided a minimal layer - the "waist" of an hourglass - that enabled the creation of a vast array of data provision, consumption, and visualization tools on the Internet."¹⁰

La démarche est ainsi fondée sur le constat du développement d'une science "intensive en données" nécessitant l'élaboration de nouvelles conventions partagées de partage et de réutilisation, sans préjuger ni contrôler les conditions de ces réutilisations. Ces principes ont connu une diffusion rapide dans le cadre de la publication académique (Wilkinson et al. 2016). Dans le même temps, ils ont fait l'objet d'une appropriation institutionnelle à l'échelle de l'Union européenne : dès 2017, la commission européenne incite les projets de recherche bénéficiaires du programme H2020 à rendre leurs données de recherche trouvables, accessibles, interopérables et réutilisables (FAIR) (European Commission, 2017). Pour les ingénieur.e.s en sciences de l'information qui assurent la gestion de données de recherche dont la publicité était déjà obligatoire au titre de la directive INSPIRE (données spatiales et environnementales), ce changement de contexte a des effets immédiats, comme en témoigne par exemple le concepteur et

¹⁰ Source: <https://www.force11.org/group/fairgroup>, consulté le 28 avril 2021.

administrateur du Portail de l'information géographique d'un grand laboratoire de sciences de l'environnement breton :

*“Le fait que les programmes de recherche financés par la Commission européenne soient fortement incités à produire des données compatibles avec le FAIR data a complètement changé la donne. On est passé d’initiatives isolées répondant « vaguement » aux préconisations de la directive **INSPIRE** avec une forme de désintérêt des tutelles à de nouvelles et fortes injonctions émanant des instituts de recherche pour que leurs équipes produisent des **FAIR data**. Le **financement** des projets de recherche en est l’enjeu, pas moins. »¹¹*

Les principes FAIR acquièrent dès lors une dimension prescriptive pour toute recherche financée totalement ou partiellement sur fonds publics (Stérim 2018; Stérim et Noûs 2019), et les instruments de financement de la recherche sur projet constituent l'espace privilégié dans lequel cette prescription est associée à un contrôle effectif, avant même la reprise de ces principes dans des textes législatifs et réglementaires.

La Directive européenne du 20 juin 2019 sur les données ouvertes et la réutilisation des informations du secteur public intègre ce principe dans la législation de l'Union Européenne (ENCADRE 1).

¹¹ <http://geowww.agrocampus-ouest.fr/web/?p=2977>, consulté le 27 octobre 2020

Encadré 1. Les données de recherche dans la Directive sur les données ouvertes

Le texte de la directive **précise l'encadrement des données de la recherche (Art. 10) et définit celles-ci (Art. 2-9).**

« Article 10 Données de la recherche

1. Les États membres encouragent la mise à disposition des données de la recherche en adoptant les politiques et en prenant les mesures nécessaires à l'échelon national afin de rendre librement accessibles les données résultant de la recherche financée au moyen de fonds publics (« politiques de libre accès») qui respectent le principe d'ouverture par défaut et sont compatibles avec les principes FAIR. Dans ce contexte, il y a lieu de tenir compte des préoccupations liées aux droits de propriété intellectuelle, à la protection des données à caractère personnel et à la confidentialité, à la sécurité et aux intérêts commerciaux légitimes dans le respect du principe «aussi ouvert que possible, mais aussi fermé que nécessaire». Ces politiques de libre accès visent les organismes exerçant une activité de recherche et les organisations finançant une activité de recherche.

2. Sans préjudice de l'article 1er, paragraphe 2, point c)¹², les données de la recherche sont réutilisables à des fins commerciales ou non commerciales, conformément aux chapitres III et IV, dans la mesure où elles sont financées au moyen de fonds publics et où des chercheurs, des organismes exerçant une activité de recherche ou des organisations finançant une activité de recherche les ont déjà rendues publiques par l'intermédiaire d'une archive ouverte institutionnelle ou thématique. À cette fin, il est tenu compte des intérêts commerciaux légitimes, des activités de transmission des connaissances et des droits de propriété intellectuelle préexistants. »

L'article 2.9 précise la définition des données de la recherche dans le cadre de cette Directive : « «données de la recherche», des documents se présentant sous forme numérique, autres que des publications scientifiques, qui sont recueillis ou produits au cours d'activités de recherche scientifique et utilisés comme éléments probants dans le processus de recherche, ou dont la communauté scientifique admet communément qu'ils sont nécessaires pour valider des conclusions et résultats de la recherche; »

Comme le rappelle l'exposé des motifs de ce texte présenté dans la proposition de la Commission a son sujet, « Il s'agit notamment de fournir un accès en temps réel aux données dynamiques par des moyens techniques appropriés, de fournir davantage de données publiques de grande valeur aux fins de réutilisation, de prévenir l'apparition de nouvelles formes d'accords d'exclusivité, de limiter le recours aux exceptions au principe de tarification au coût marginal et de clarifier les relations entre la directive ISP et certains instruments juridiques connexes. »¹³. Cette volonté de développer la masse de données librement disponible pour favoriser l'économie numérique s'inscrit dans la stratégie pour le marché unique numérique¹⁴, développée quelques années plus tôt, en 2015, par la Commission Européenne après avoir été déclarée prioritaire dans les orientations politiques de la Commission Juncker¹⁵.

La **Stratégie pour le marché unique numérique** précise ainsi dans son paragraphe 4.1 "Construire une économie des données" :

« Les mégadonnées, l'informatique en nuage et l'internet des objets sont indispensables à la compétitivité de l'UE. Les données sont souvent considérées comme un catalyseur de croissance économique, d'innovation et de conversion au numérique dans tous les secteurs économiques, en particulier pour les PME (et les *start ups*) et pour la société dans son ensemble. Les mégadonnées et le calcul à haute performance font également évoluer les pratiques dans le domaine de la recherche et du partage des connaissances, participant d'une transition vers une «science ouverte» plus performante et réactive.

Le secteur des mégadonnées connaît une croissance de 40 % par an, sept fois supérieure à celle du marché de l'informatique.

Un marché fragmenté n'offre pas suffisamment d'espace pour exploiter pleinement le potentiel de l'informatique en nuage, des mégadonnées, de la science fondée sur les données et de l'internet des objets en Europe. Pour valoriser au mieux les technologies numériques et les technologies des données, nous devons lever une série d'obstacles techniques et législatifs. » COM(2015)192 p. 16

Par ailleurs, le préambule qui précède le dispositif de la Directive européenne du 20 juin 2019 sur les données ouvertes et la réutilisation des informations du secteur public témoigne de la façon dont sont associés à l'Open data des enjeux de contrôle et de rationalisation des pratiques scientifiques, dans une vision finaliste et utilitariste du progrès de la connaissance:

« Le libre accès contribue à améliorer la qualité, à réduire la duplication inutile des recherches, à accélérer le progrès scientifique, à lutter contre la fraude scientifique, et peut globalement favoriser la croissance économique et l'innovation » (considérant 27 de la Directive européenne du 20 juin 2019 sur les données ouvertes et la réutilisation des informations du secteur public).

¹² Le champ d'application de la directive est défini à l'art. 1er de la Directive. Il intègre les données de la recherche mais ne s'applique pas "aux documents dont des tiers détiennent les droits de propriété intellectuelle"

¹³ Commission Européenne « Proposition de directive du Parlement Européen et du Conseil concernant la réutilisation des informations du secteur public (refonte) » COM(2018)234 final, 50p., p. 1 et 2.

¹⁴ Ibid. p. 3.

¹⁵ Commission Européenne, « Communication de la Commission au Parlement Européen, au Conseil, au Comité Économique et Social Européen et au Comité des Régions – Stratégie pour un marché unique numérique en Europe » COM(2015)192.

2.1.3. La science ouverte dans la législation nationale et l'article L533-4 du code de la recherche

La Loi n°2016-1321 du 7 octobre 2016 sur la république numérique (art. 30) intègre une disposition concernant la science ouverte dans le code de la recherche. Son premier paragraphe concerne l'ouverture des publications, son deuxième et troisième paragraphe l'ouverture des données. Il est notable que le quatrième paragraphe spécifie que ces dispositions sont d'ordre public. En effet, cela signifie qu'il n'est pas possible d'y déroger par contrat, et ce « même si l'éditeur l'interdit dans le contrat d'édition, fût-ce un contrat de droit étranger » souligne Anne-Laure Stérin (Stérin 2018, p. 25¹⁶).

Encadré 2. L'article L533-4 du code de la recherche

« I. Lorsqu'un écrit scientifique issu d'une activité de recherche financée au moins pour moitié par des dotations de l'Etat, des collectivités territoriales ou des établissements publics, par des subventions d'agences de financement nationales ou par des fonds de l'Union européenne est publié dans un périodique paraissant au moins une fois par an, son auteur dispose, même après avoir accordé des droits exclusifs à un éditeur, du droit de mettre à disposition gratuitement dans un format ouvert, par voie numérique, sous réserve de l'accord des éventuels coauteurs, la version finale de son manuscrit acceptée pour publication, dès lors que l'éditeur met lui-même celle-ci gratuitement à disposition par voie numérique ou, à défaut, à l'expiration d'un délai courant à compter de la date de la première publication. Ce délai est au maximum de six mois pour une publication dans le domaine des sciences, de la technique et de la médecine et de douze mois dans celui des sciences humaines et sociales.

La version mise à disposition en application du premier alinéa ne peut faire l'objet d'une exploitation dans le cadre d'une activité d'édition à caractère commercial.

II.-Dès lors que les données issues d'une activité de recherche financée au moins pour moitié par des dotations de l'Etat, des collectivités territoriales, des établissements publics, des subventions d'agences de financement nationales ou par des fonds de l'Union européenne ne sont pas protégées par un droit spécifique ou une réglementation particulière et qu'elles ont été rendues publiques par le chercheur, l'établissement ou l'organisme de recherche, leur réutilisation est libre.

III.-L'éditeur d'un écrit scientifique mentionné au I ne peut limiter la réutilisation des données de la recherche rendues publiques dans le cadre de sa publication ».

IV.-Les dispositions du présent article sont d'ordre public et toute clause contraire à celles-ci est réputée non écrite. »

2.1.4. Des injonctions relayées par les principaux bailleurs de la recherche et collectifs engagés

Le déploiement de l'*Open Data* au sein de la recherche publique est ainsi porté par deux catégories principales d'acteurs. Les premiers sont des collectifs transnationaux hybrides, impliquant fortement les professionnels de la recherche et en particulier les communautés d'appui, qui sont par exemple engagés et structurés en France au sein du Comité pour la Science Ouverte (COSO). Les seconds sont les bailleurs institutionnels de la recherche publique à l'échelle européenne, qui formalisent ces prescriptions dans leurs appels à projets (bourses de recherche de l'European Research Council, programme H2020 de l'Union européenne) et sont bientôt rejoints par les agences nationales de financement de la recherche. Ceux-ci sont fédérés au sein de l'initiative cOAlition S (qui porte le Plan S).

Dans la pratique, ces deux catégories d'acteurs interagissent et négocient de façon intense la construction, la consolidation de standards visant la transformation des pratiques de recherche et de diffusion de la recherche.

L'exemple de l'Agence nationale de la recherche (ANR) en France est à cet égard parlant: le site de l'ANR détaille ainsi la façon dont sa propre politique de Science Ouverte s'articule avec les dynamiques en cours à l'échelle de l'Union européenne, en mettant à la fois l'accent sur son implication dans « plusieurs initiatives transnationales au sein desquelles elle porte la position française en faveur de la Science Ouverte et de la bibliodiversité. Elle est membre de la cOAlition S qui regroupe plusieurs organismes de financement pour

¹⁶ Voir également : <https://ethiquedroit.hypotheses.org/1748>; ainsi que L533-4 Code de la recherche – et le rapport de l'assemblée nationale : <https://www.assemblee-nationale.fr/14/rapports/r3399.asp>

accélérer la transition vers un accès complet et immédiat aux publications scientifiques et soutient le Plan S. L'Agence est également membre du bureau GO FAIR français.”¹⁷. Les éléments de discours produits par l'ANR dans le cadre de ses documents de politique générale, comme de ses actions de diffusion auprès des communautés de recherche, mettent l'accent sur les valeurs qui sous-tendent l'engagement dans la science ouverte et sur la concertation avec les réseaux de professionnel.le.s de la recherche. La politique de science ouverte est ainsi, comme dans la plupart des institutions de la recherche et de l'enseignement supérieur, décrite et mise en récit comme un engagement au service de la connaissance et de la citoyenneté (FIGURE 2).



Figure 2. Présentation schématique de l'évolution de la politique de la Science Ouverte par l'ANR¹⁸

Le principe fondamental « aussi ouvert que possible, aussi fermé que nécessaire »¹⁹ est repris dans le discours officiel de l'ANR, mais aussi reformulé, accompagné d'explications nuancées qui viennent justifier la "demande" formulée auprès des porteurs de projet d'élaboration d'un plan de gestion des données pour les projets qu'elle finance à partir de 2019, "document qui synthétise la description et l'évolution des jeux de données, et prépare le partage, la réutilisation et la pérennisation des données"²⁰. Les discours institutionnels sur la gestion des données de la recherche oscillent systématiquement entre deux registres de justification faisant appel au civisme scientifique : **le rappel des règles de droit, et l'invocation d'une nécessité et d'un fondement endogènes aux communautés de recherche**, qui le rendraient plus légitimes et renforceraient la mobilisation, attendue, des professionnel.le.s de la recherche.

¹⁷ Page "Science ouverte" du site de l'ANR: <https://anr.fr/fr/lanr-et-la-recherche/engagements-et-valeurs/la-science-ouverte/>, consultée le 27 avril 2021.

¹⁸ Source : <https://anr.fr/fr/lanr-et-la-recherche/engagements-et-valeurs/la-science-ouverte/>, consulté le 26 octobre 2020

¹⁹ Directive européenne du 20 juin 2019 sur les données ouvertes et la réutilisation des informations du secteur public, article 10.

²⁰ Site de l'ANR, consulté le 21/09/2020 : <https://anr.fr/fr/lanr-et-la-recherche/engagements-et-valeurs/la-science-ouverte/>

Cette ligne de crête entre incitation et obligation est ainsi difficile à tenir pour les organismes prescripteurs: les règles de droit sont formulées de façon très générale et le travail d'interprétation que leur bonne application requiert, en l'absence de jurisprudence, est considérable. Les interactions entre prescripteurs institutionnels, communautés d'appui à la recherche, chercheurs.euses et enseignant.e.s-chercheurs.euses au sujet de la gestion des données sont ainsi marquées par des incertitudes fondamentales quant à la compréhension de la portée juridique, organisationnelle et pratique de ces changements - voire sur leur dimension politique.

2.1.5. « Aussi ouvert que possible, aussi fermé que nécessaire » : articuler protection et publicisation des données

Sur un plan juridique et pratique, le respect des obligations définies par les textes en vigueur implique de circonscrire, dans un premier temps, le champ de ce qui doit rester "fermé", en prenant par ailleurs en compte les dispositions spécifiques prévoyant la publicité de principe pour certains types de données. C'est le cas, selon les termes de la directive INSPIRE et des textes qui en assurent la transposition en droit français, des données spatiales et des données environnementales, dont on peut considérer qu'elles constituent, pour les sciences humaines et sociales, plutôt des données contextuelles posant moins souvent des problèmes d'articulation entre publicité et protection (notamment liés au respect du secret statistique).

Les SHS sont en revanche directement et particulièrement concernées par les données "fermées". Dans certains cas, l'encadrement juridique est très explicite et lisible : c'est le cas des données médicales. C'est également le cas *a priori* des données dites "sensibles"²¹. Pour ces dernières, cependant, la question posée n'est pas seulement celle de leur fermeture, mais se formule également en termes de possibilité-même de collecte : les formes de justification de la collecte de données sensibles à des fins de recherche et leur gestion pérenne²², 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 et la fermeture de principe, 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

²¹ L'article 9 du RGPD définit les données sensibles comme "des données à caractère personnel qui révèlent 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."

²² 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".

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.

2.2. La traduction de l'*Open Data* en pratique : faire avec un cadre en pleine évolution

2.2.1. Les réseaux d'appui à la recherche : une enquête centrée sur l'appui scientifique et technique et l'appui juridique

Les réseaux d'appui à la recherche sont particulièrement mobilisés dans le contexte du déploiement du mouvement de l'ouverture des données et de la mutation des pratiques qu'il implique. Dans le cadre du programme PARDOQ, nous avons donc orienté notre enquête autour de l'expérience de deux types d'acteurs particulièrement impliqués dans l'accompagnement de ce mouvement récent : d'une part des communautés d'appui technique (personnels ingénieurs de l'information scientifique et technique et des humanités numériques²³) et d'autre part, des communautés d'appui juridique, à travers en particulier les délégués à la protection des données (DPD, ou DPO en anglais).

Notre enquête auprès de ces deux types d'interlocuteurs, menée principalement sur la base d'entretiens semi-dirigés et d'observations, est relativement ciblée : nous avons réalisé des entretiens assez courts (entre 30 minutes et 1 heure) autour de questions portant sur :

- le périmètre de leurs missions,
- le type d'interactions avec les communautés de recherche en SHS : types de sollicitations, formes d'accompagnements,
- les spécificités rencontrées en lien avec les matériaux ethnographiques dans les expériences d'appui vécues, et les problèmes rencontrés,
- l'évolution des types de demandes et d'accompagnement dans les années récentes.

En tout, **7 entretiens** ont été réalisés avec des membres des **réseaux d'appui technique et scientifique**, auprès d'ingénieur.e.s d'études ou de recherche dont 2 avec des chargés de mission de la Très grande infrastructure de recherche (TGIR) Huma-Num, qui constitue une instance stratégique d'envergure nationale pour l'équipement numérique des communautés de la recherche qualitative, et la promotion du mouvement de l'ORD auprès de ces communautés. Pour les entretiens avec les 5 autres membres, nous avons fait le choix de cibler un réseau en particulier, important pour les communautés de recherche en SHS (mais qui n'est pas représentatif de l'ensemble de la communauté d'appui à la recherche) : le réseau des MSH (Maisons des sciences de l'Homme). Il constitue un réseau structuré à l'échelle du territoire national avec 22 MSH, réparties par régions. Chaque MSH propose une offre d'appui numérique aux chercheurs.euses et ingénieur.e.s des universités, unités de recherche et écoles de leurs régions. Elles représentent ainsi, pour reprendre les termes d'un interlocuteur de l'enquête, "un guichet d'accompagnement à la recherche", proposant une expertise avancée sur les enjeux émergents liés à la Science Ouverte et la numérisation des données. Les MSH sont par ailleurs dotées d'ingénieurs en humanités numériques qui endossent le rôle de correspondants régionaux pour la TGIR Huma-Num. Nous avons fait le choix de centrer notre enquête sur ce type de profil d'interlocuteurs.trices, se situant à l'interface entre la TGIR et les instances régionales que représentent les MSH d'une part, et d'autre part, les communautés de recherche. De ce fait, notre enquête cible une communauté particulièrement active, structurée à l'échelle nationale et régionale, et formée au regard des questionnements qui nous intéressent : ces réseaux d'appui assument en effet un engagement institutionnel en faveur de la Science Ouverte et de

²³ Notons que ces professionnel.le.s sont en général formé.e.s à ces enjeux, mais ne sont pas également réparti.e.s dans tous les segments de l'enseignement supérieur et de la recherche en SHS : ils.elles sont très représenté.e.s au sein des infrastructures inter-établissements (TGIR, MSH...), beaucoup moins dans les unités de recherche ou les services des universités. Dans ce dernier segment, ils.elles sont en général affecté.e.s ou dédié.e.s à des observatoires, à la gestion d'outils spécifiques ou à du traitement de données thématiques davantage qu'à la gestion de celles-ci. Sans que notre enquête permette de le quantifier, il semble donc que le mouvement de l'Open Data se traduise par des transformations des métiers, mais aussi par des translations de moyens entre segments de la recherche.

l'ouverture des données. Elle permet ainsi de dégager des tendances éclairantes au regard des questions soulevées dans ce rapport, en documentant l'expérience d'une communauté particulièrement engagée. Néanmoins, elle ne reflète pas la grande diversité des situations que l'on peut rencontrer au sein des réseaux d'appui scientifique et technique en SHS, notamment en nous intéressant aux personnels internes aux unités de recherche qui sont peu représentés ici (un seul de nos interlocuteurs correspond à ce profil), aux personnels bibliothécaires ou impliqués dans la gestion de services communs de documentation (SCD), ou encore aux métiers émergents tels que les "chargé.e.s de données de recherche", les chargé.e.s Science Ouverte", ou les "Data Librarians" par exemple, qui constituent des branches variées de ces réseaux d'appui.

D'autre part, **4 entretiens ont été réalisés avec des membres de réseaux d'appui juridique**, auprès de DPD affiliés à 4 EPST ou universités différents.

Du fait du caractère restreint et très ciblé de ce volet de l'enquête PARDOQ, les résultats présentés dans cette partie doivent être appréhendés comme exploratoires²⁴, et méritent des approfondissements. Une exploitation plus fine des matériaux d'enquête collectés dans le cadre de PARDOQ serait utilement complétée par un prolongement de cette étude, élargie à d'autres sphères d'acteurs.trices de ces différents réseaux d'appui; en particulier les réseaux d'appui internes aux structures de recherche, et également les différents types de profils cités ci-dessus et qui ne sont pas représentés dans notre étude.

2.2.2. L'appui scientifique et technique : expériences des ingénieurs en humanités numériques et correspondants de la TGIR Huma-Num

Les missions des personnels d'appui techniques et scientifiques que nous avons interrogés, et qui travaillent au contact des équipes de chercheurs.euses, comportent une double dimension :

- **l'accompagnement des chercheurs.euses sur les volets numériques** de leurs projets (répondre à leurs besoins en équipements numériques, les aiguiller vers les outils existants ou émergents – en particulier ceux de la TGIR Huma-Num s'ils correspondent à leurs besoins, accompagner la conception du volet numérique de la gestion des données d'un projet, les conseiller en matière de prestation). Dans ce cadre, les ingénieur.e.s des MSH expliquent jouer **un rôle « d'aiguillage »** vers les bons interlocuteurs si la demande qui leur est formulée n'entre pas dans leurs champs de compétences (autres personnels des MSH, personnels internes aux structures de recherche par exemple). Ils.elles représentent en cela à la fois un « guichet » et un « filtre » pour réceptionner et rediriger les demandes et besoin des chercheurs.euses, selon les termes de Benjamin, ingénieur au sein de la TGIR (entretien en novembre 2020).

- **la coordination de services numériques** à l'échelle de leur institution (suivi et déploiement des services et infrastructures numériques, communication et sensibilisation des communautés de recherche autour de cette offre).

Outre l'accompagnement des praticien.ne.s directement lié à la première dimension de leur mission, les ingénieur.e.s interrogés collaborent largement avec d'autres métiers impliqués dans le développement de l'offre numérique et l'appui à la recherche, en particulier des métiers techniques tels que des informaticien.ne.s, des gestionnaires de bases de données, ou des chargés web-design par exemple.

Promouvoir les « bonnes pratiques » de gestion des données numériques

Dans leurs missions d'appui à la recherche, une attention particulière est portée à la promotion de **« bonnes pratiques » de gestion des données dans un environnement numérique** - pour reprendre l'expression communément utilisée dans les entretiens et au sein de cette communauté. En particulier, les enjeux de documentation, d'archivage, de stockage sécurisé et de sauvegarde des données numériques (nativement numériques ou numérisées) font l'objet d'une vigilance dont ils témoignent largement au cours des

²⁴ Limitées par le format de cette enquête, nous avons fait le choix d'accentuer les efforts d'analyse sur les matériaux d'enquête collectés auprès des communautés de chercheurs.euses et présentés en partie 4.

entretiens, et d'un travail de sensibilisation dans les processus d'accompagnement des chercheurs.euses. Il s'agit d'inciter les praticien.n.e de la recherche à évoluer vers des pratiques plus standardisées de nommage et de classement numériques de leurs fichiers de données, à adopter un protocole rigoureux et anticipé de sauvegarde de ces données sur plusieurs supports, et à être attentifs aux propriétés des outils choisis pour stocker, sauvegarder et éventuellement diffuser ces données.

Le choix des outils (de traitement ou de stockage des données) est orienté dans la mesure du possible vers des logiciels s'appuyant sur des licences ouvertes (*Open Source*) : une « culture du libre » est communément partagée et encouragée au sein de ces réseaux comme l'explique Marie, ingénieure en humanités numériques dans une MSH, lors d'un entretien en novembre 2020. Les solutions de stockage développées « localement » sont privilégiées, c'est-à-dire de préférence les outils proposés par les MSH via la TGIR Huma-Num dans le cas de nos enquêt.e.s (comme *Huma-Num Box* ou *SHAREDOCS* par exemple) et des outils institutionnels déployés par les tutelles ou institutions auxquelles les chercheurs.euses sont affiliés, dont les serveurs sont hébergés en France ou en Europe dans le respect du cadre réglementaire du RGPD. Marie (ingénieure de recherche dans une MSH) souligne que promouvoir ces « bonnes pratiques » en matière de gestion des données numériques, et sensibiliser à l'utilisation d'outils *Open Source* est une ligne politique poursuivie par toutes les instances de la recherche publique en France. Elle adhère à ces valeurs et considère en être une porte-parole dans le cadre de ses missions, tout en nuancant la réalité des faits : ces « bonnes pratiques » sont un horizon vers lequel elle tente de pousser les chercheurs.euses, mais il existe des écarts entre ce modèle et les pratiques ancrées des chercheurs.euses. Elle explique ces écarts par une diversité de raisons, notamment des incohérences structurelles et la prégnance d'usages partagés de certains outils propriétaires :

« Promouvoir l'Open Source, c'est pas tant une politique de la MSH, que du CNRS, du monde de la recherche. Avec toutes ses incohérences. Puisque par exemple à l'Université [voisine], tout ce qui est outils numériques c'est Google, etc, donc c'est des incohérences, mais c'est vrai que le discours est sur de l'Open Source. Et ça je pense que ça a été bien intégré, cette notion d'aller plutôt vers de l'Open Source que du propriétaire, mais après les chercheurs font aussi avec ce qu'ils peuvent. Par exemple, il y a des outils communautaires, comme FileMaker, qui est beaucoup utilisé par les archéologues, et comme ils utilisent tous cela, ils sont formés avec cet outil, ils continuent à l'utiliser. Parce qu'il n'y a pas forcément l'équivalent en Open Source ». (Marie, ingénieure de recherche novembre 2020)

S'adapter à des demandes diverses et construire un accompagnement au cas par cas

Si les membres des réseaux d'appui tendent à faire évoluer les pratiques en promouvant ce qu'ils définissent comme des modèles de « bonnes pratiques », leur relation avec les chercheurs.euses se construit avant tout **à l'initiative de ces derniers.ères, sur la base de sollicitations fondées sur une approche opérationnelle de l'équipement numérique de leur recherche**, et non pas tant sur cet objectif d'évolution de leurs propres pratiques de gestion des données. Marie explique qu'« il y a une grande richesse dans les demandes », et il est difficile de les résumer. Anne, une autre ingénieure en humanités numériques explique que les demandes vont d'une « simple sollicitation qui peut être facilement solutionnée en un échange » (ouvrir l'accès à un outil ciblé par exemple), à des formes d'accompagnement plus complexes et « de plus long terme ». Dans ce contexte, leur homologue dans une autre MSH, Martine souligne la nécessaire polyvalence de ses compétences, pour s'adapter à la fois à la diversité de **la nature des demandes** (types de besoin), mais également à **la diversité des spécificités disciplinaires** qui s'inscrivent dans le champ élargi des sciences humaines et sociales :

« Je suis plus impliquée sur certains projets, comme des projets de création de bases de données par exemple. [...] On a des projets de toutes les disciplines, c'est très varié : il y a des littéraires, des juristes, des archéologues, des économistes, géographie, histoire, c'est très divers » (Martine, correspondante d'une MSH, novembre 2020).

Dans cette configuration, venir en appui au.à la chercheur.euse à l'origine d'une sollicitation implique donc **une phase d'échange** et une démarche au cas par cas, pour caractériser les besoins, qui sont rarement en pleine adéquation avec l'offre existante. L'accompagnement, ou l'appui aux chercheurs.euses s'inscrivent donc dans **un processus de collaboration, ou de co-construction**, qui implique un des négociations entre les attentes des chercheurs, et les ajustements des services et outils pouvant répondre à ces besoins. Martine décrit explicitement cette phase de tâtonnements, et d'ajustements réciproques entre la caractérisation de la demande (du besoin) et la recherche de solutions :

« Dans tous les cas, la première réponse, ça va être de discuter vraiment pour comprendre vraiment ce que veut le chercheur. Même si ça paraît très clair, ça ne l'est pas toujours. Donc on intervient aussi sous forme de conseils sur des choses qui ne sont pas forcément demandées à l'origine. Il y a systématiquement une phase d'échanges avec eux pour bien cadrer leur demande, même s'ils ont une connaissance assez précise de ce qu'ils recherchent, on se rend très vite compte, presque toujours... À toutes les premières réunions, je pose des questions souvent basiques, par exemple : à qui il veulent s'adresser, quel va être leur public [pour leur site internet], et ça c'est des questions que les chercheurs ne se posent jamais, ou rarement. Et qui amènent forcément à d'autres discussions, réflexions. [...] L'exemple classique, souvent c'est qu'ils veulent faire un site internet, et quand on demande à qui ils veulent s'adresser : « au grand public », et en fait ce sont des projets de recherche très pointus, et pas destinés au grand public. Et voilà, à partir de là on va poser des questions qui vont amener à une réflexion plus générale sur leurs projets ».
(Martine, correspondante d'une MSH, novembre 2020)

Si la démarche d'accompagnement est initialement centrée sur les aspects techniques et les outils, elle se déploie donc dans une dimension plus large qui nécessite de saisir la teneur des besoins numériques en lien avec les objectifs du projet, ainsi que les spécificités des données concernées telles que produites par les chercheurs.euses, pour ajuster le niveau et la forme d'accompagnement nécessaires. **Elle implique une approche au cas par cas, nécessairement coûteuse en temps pour les parties-prenantes impliquées.** Une situation parfois problématique point de vue de nos interlocuteurs.trices, qui expliquent qu'ils.elles doivent rester vigilant.e.s sur l'intensité de leur implication pour chaque accompagnement, qui ne peut se transformer en partenariat approfondi du fait de la multiplicité des interlocuteurs et des demandes, ce malgré l'intérêt et/ou la nécessité que ce partenariat peut revêtir. Marie témoigne ainsi d'une expérience d'implication très approfondie dans le cadre d'un projet qu'elle a suivi sur plusieurs années et qui l'a passionnée, mais elle explique avoir « vite compris [qu'elle] ne pourrait pas s'investir à ce niveau pour tous les projets [qu'elle] suit », d'autant plus dans un contexte où elle reçoit de plus en plus de demandes (Marie, ingénieure en humanités numériques dans une MSH, novembre 2020). Cela pose structurellement le problème de l'échelle à laquelle se positionne l'accompagnement, avec un écart entre une tendance actuelle à renforcer les capacités de structures *supra* (externes aux unités de recherche) et pluridisciplinaires (comme les MSH) dans une logique de mutualisation des compétences, alors que les besoins se situent à des échelles internes et spécialisées, proches des praticien.ne.s. Un écart que les personnes interrogées tentent de réduire en négociant et en ajustant leur accompagnement, comme en témoigne cette ingénieure :

“D'ailleurs, les chercheurs sollicitent souvent un accompagnement de proximité, à l'échelle du laboratoire, ce qui n'est pas la dynamique observée dans les créations de postes ces dernières années. Nous essayons dans tous les cas de fournir cet accompagnement personnalisé, mais il y a toujours cette question d'échelle irréductible, on ne peut pas proposer le même service, selon que l'on s'adresse à 50 ou 1000 chercheurs”. (Anne, ingénieure en humanités numériques dans une MSH)

Peu de demandes portant sur des matériaux ethnographiques

Un constat ressort des témoignages recueillis dans notre enquête : les matériaux ethnographiques sont peu représentés parmi les demandes réceptionnées et suivies par les interlocuteurs.trices interrogé.e.s

concernant la gestion des données de recherche et les besoins d'appui numérique associés. Ce, malgré le fait que l'enquête ait particulièrement ciblé des interlocuteurs.trices au plus proche de la recherche qualitative (en procédant par effets de réseaux), dans des MSH susceptibles d'intégrer une composante ethnologique ou *a minima* des recherches impliquant une démarche ethnographique. Il n'est pas possible de conclure de ce constat que peu de demandes ou besoins en termes d'appui numérique sont associés à ce type de matériaux, notamment parce que l'on pourrait supposer qu'elles soient traitées ailleurs, par exemple en interne aux unités de recherches, ou encore que les chercheurs ignorent l'existence de cette ressource externe. Néanmoins, cela nous informe sur le fait que peu de demandes s'orientent vers ce type de ressource mutualisée que constituent les MSH. Notre enquête a permis de recenser deux exemples particuliers, qui sont de natures différentes, et qui intègrent une problématique de gestion de "données" ethnographiques.

Le premier est celui d'une sociologue travaillant à l'élaboration d'un Plan de gestion des données dans le cadre d'un projet de recherche financé, et s'interrogeant plus particulièrement sur la gestion de données sensibles et personnelles et l'adoption d'un dispositif numérique adapté à ces particularités. Il s'agit donc d'une démarche qui s'inscrit dans un objectif de structuration et de sécurisation des données de recherche. Le second exemple est celui de l'accompagnement d'un collectif de chercheurs.euses à la création d'un portail unique de diffusion de matériaux ethnographiques, dans le cadre d'une collaboration avec les communautés sources où ces matériaux de recherche ont été collectés. Dans ce cadre, la démarche a plus particulièrement porté sur la mise en œuvre d'un équipement numérique adapté qui permette la diffusion de données, par ailleurs déjà largement rendues publiques en ligne *via* d'autres canaux de diffusion, à partir d'un portail d'entrée unique, afin qu'elles soient accessibles aux communautés d'origine qui n'étaient pas en mesure de consulter ces contenus dans la configuration éclatée qui préexistait. Il s'agit, en l'occurrence, du seul témoignage que nous avons recueilli portant sur l'accompagnement à une forme de diffusion publique de données ethnographiques.

Ces deux témoignages contrastés, touchant d'une part à la question de la protection des données ethnographiques, et d'autre part à leur publicisation coordonnée, illustrent une dimension problématique rattachée à ce type de matériaux et qui transparaît dans ces deux processus de traitements très différents : il s'agit, dans les deux cas, d'**une procédure d'accompagnement complexe reliée à la nature hétérogène de ces matériaux, qui les rendent difficilement saisissables et caractérisables de manière globale, mais nécessitent au contraire un travail minutieux au cas par cas**. Ce travail de caractérisation fine des matériaux dans le cadre de la production d'un PGD (nécessitant de déterminer ce qui relève du cadre de la RGPD dans un corpus donné) d'un côté, ou de centralisation de la diffusion de données hétérogènes *via* un portail numérique de l'autre côté, impliquent donc des processus longs et fastidieux de définition du statut de ces données sur les plans juridique, éthique et technique, qui ne peuvent répondre à des schémas standardisés et applicables ou répliquables "clé en main". C'est sans doute, une spécificité de ces données que l'on pourrait qualifier d'ultra-qualitatives par nature, quand bien même leur traitement pourrait concerner de gros volumes.

Des formes de diffusion des données qui n'impliquent pas (encore) leur FAIRisation

Le second exemple, et le seul donc que nous avons recensé auprès de nos interlocuteurs.trices concernant une forme de publicisation de données ethnographiques, a pour spécificité de concerner des données majoritairement déjà rendues publiques par ailleurs, souvent dans le cadre de projets ou de structures porteurs d'une dimension patrimoniale (en particulier des collections numériques de musées). Marie, qui a contribué à ce projet sur plusieurs années, explique que le principal enjeu était d'harmoniser et de compléter la mise en ligne de ces contenus, en d'affiner leur documentation :

« Comme il y a une partie des données qui étaient déjà disponibles par les musées eux-mêmes, et par un des centres de recherche [impliqués], en fait l'idée c'était de proposer un point d'entrée unique pour les communautés pour accéder à ces données. Et, il y a une partie des données qui étaient inédites, donc pas du tout traitées, donc deux chercheurs ont accepté de les mettre à disposition puisqu'il s'agissait d'un retour auprès des

populations qu'ils avaient eux-mêmes filmées ou enregistrées, donc là on les a déposées dans Nakala et on les a documentées dans le formalisme qui est proposé, donc qui est interopérable, etc, donc on est en plein dans les clous de la Science Ouverte. Mais ce n'était effectivement pas tant d'être dans la Science Ouverte en termes de principes FAIR, etc, que d'ouvrir ces données et de les rendre disponibles et accessibles, et aussi manipulables, puisqu'en fait l'idée c'est que les communautés d'origine puissent aussi enrichir la documentation sur ces données, puissent aussi les commenter. Voilà, parce qu'en fait on s'est rendu compte qu'il y avait des informations qui étaient totalement partielles ou qui étaient erronées par exemple. On se rend compte de ça aussi. Donc effectivement, on n'était pas tant dans la Science Ouverte que dans la mise à disposition de ces données de manière facilitée aux communautés. [...] Donc en fait souvent c'est quand même de la valorisation qu'ils veulent faire ». (Marie, ingénieure en humanités numériques dans une MSH, novembre 2020).

Le témoignage de Marie rejoint par ailleurs le sens des autres témoignages que nous avons recueillis concernant l'accompagnement de projets de publicisation de données, qu'elles soient ethnographiques ou d'autres natures : ces projets ne prennent pas forme dans une volonté de s'inscrire dans un mouvement institué d'ouverture des données pour se conformer aux standards de la Science Ouverte, mais s'appuient davantage sur des motivations internes, propres aux projets de recherches et aux chercheurs.euses qui le mettent en oeuvre, de promouvoir d'autres formes de "valorisation" que la publication scientifique classique. Il s'agit de partager la recherche y compris en la rendant visible à différents niveaux de son élaboration, notamment à travers la publicisation de certaines données. Les outils numériques offrent donc pour cela des opportunités nouvelles dont les chercheurs se saisissent progressivement.

L'application des principes FAIR ("facile à trouver, accessible, interopérable et réutilisable"), qui définissent les fondements techniques du partage ouvert des données (et des métadonnées) au sens de l'ORD, n'est donc pas tant, de l'expérience de nos interlocuteurs.trices, un motif de sollicitation de la part des chercheurs.euses, qu'**une conséquence de cette interaction entre réseaux d'appui et réseaux de recherche**. C'est suite à une sollicitation autour des solutions de partages numériques de contenus liés à une recherche que les membres des réseaux d'appui orientent les chercheurs vers une plus grande "FAIRisation" de leurs données, pour reprendre le néologisme désormais largement employé dans cette communauté. Marie l'explique clairement dans son témoignage :

*« C'est-à-dire que la demande ce n'est pas « j'ai besoin de répondre aux principes FAIR », c'est « je voudrais valoriser, je suis d'accord pour que mes données soient totalement accessibles », ou pas d'ailleurs, là ça va dépendre de certains types de données. Mais ce n'est pas tant la Science Ouverte que le fait de vouloir les valoriser et les mettre à disposition. [...] Après il se trouve que c'est parce qu'ils ont des financements qui leur évite d'avoir ces obligations-là, de mise à disposition des données sur des entrepôts [...] Moi c'est le constat que je fais. C'est vraiment de la valorisation, et ensuite, en fait, comme on les oriente vers des outils qui rentrent dans les principes FAIR, ben voilà, ils rentrent cette politique de Science Ouverte, **mais ce n'est jamais l'argument de départ**. Ou c'est même... quand j'avais été contactée par un collègue pour répondre à un appel à projet, là on sentait bien que en fait, là c'est une analyse que je fais, c'est une interprétation, mais ils avaient besoin de sous pour ce projet-là, donc ils étaient prêts à y mettre un peu de principes FAIR, un peu de Science ouverte pour colorer ça et pour que ce soit accepté. Mais leur but c'était pas tant la Science Ouverte que la refonte de leur site web un peu vétuste. Et comme il fallait répondre à la Science Ouverte, alors on m'a contactée pour ajouter des éléments qui auraient pu plaire aux financeurs ». (Marie, ingénieure en humanités numériques dans une MSH, novembre 2020).*

Martine, ingénieure dans une autre MSH, mais qui n'a pas d'expérience de mise en partage de matériaux ethnographiques, raconte néanmoins comment l'accompagnement que son service propose depuis une dizaine d'années est largement axé sur la valorisation publique de données de recherche qui préexiste à

l'adoption des principes FAIR. Si aujourd'hui elle pousse les chercheurs.euses à adopter ces principes dans le contexte de leur avènement dans les années récentes, elle raconte que les formes "artisanales" de valorisation, et la description "maison" des données demeurent privilégiées, notamment parce qu'elles sont pensées comme plus proches des objectifs de recherche :

« Pour l'instant, nous on n'est pas encore allés vers des entrepôts de type Zenodo²⁵ ou ce genre de chose. On a plutôt le cas où les chercheurs nous demandent la mise à disposition des données sur leurs sites, mais pas sur des entrepôts institutionnels. Par exemple, ils vont faire une base de données et mettre à disposition l'intégralité de la base de données avec les fichiers. En fait, ils partagent leur base de données. [...] Nous on essaie de les pousser à FAIRiser leurs données, mais ils sont plus dans des problématiques de description maison. Parce que, dans leur logique, c'est pour avoir des données qui soient plus détaillées, et répondant plus à leurs objectifs de recherche ». (Martine, ingénieure en humanités numériques dans une MSH, novembre 2020).

Les témoignages recueillis mettent par ailleurs en lumière un autre constat relatif à ces processus de publicisation des données : lorsque les chercheurs.euses sollicitent l'appui des ingénieur.e.s, ils portent déjà **une vision relativement claire de corpus qu'ils souhaitent "valoriser"** (c'est-à-dire partager et/ou publiciser). Ainsi, de leur point de vue, l'accompagnement requis ne porte pas tant sur un corpus de données de recherche pris comme un ensemble au sein duquel il s'agirait de déterminer ce qui relève d'un statut de protection, au sens "autant fermé que nécessaire", et ce qui relève d'une ouverture par défaut ("autant que possible"), mais davantage sur des sélections de données pré-conçues par les chercheurs selon leurs propres critères et auxquelles ils associent un projet de « valorisation », (ou à l'inverse des corpus qu'ils souhaitent protéger en les stockant et en les sauvegardant sur des serveurs sécurisés). Ainsi, dans le cas des projets de publicisation des données, le processus de sélection entre ce qui est rendu "ouvert autant que possible" et "ce qui reste fermé autant que nécessaire", pour reprendre ce principe fondateur du mouvement d'ouverture des données, est principalement opéré en amont de la consultation des ingénieur.e.s, et selon des critères qui ne sont pas nécessairement portés à la connaissance de ces dernier.e.s, et ne correspondent pas nécessairement à ceux mis en avant par la Science Ouverte. L'accompagnement porte davantage sur des aspects techniques et pratiques autour de la forme et des outils les plus adaptés pour ce partage, que sur une réflexion autour de ce qui peut et devrait être rendu public et selon quels critères :

« On a toujours travaillé que sur des projets ouverts, on n'a jamais travaillé sur des projets dans lesquels il y avait un but final de laisser les données fermées. Donc elles le sont toujours une fois le projet terminé. Ça fait partie de la nature de [notre structure], le fait que ce soit ouvert et le fait qu'on utilise que des logiciels libres. Donc la question ne s'est jamais posée. [...] On n'a jamais eu de problèmes par rapport à la communicabilité des données à la fin des projets. Donc on était déjà dans cette culture-là [avant le mouvement de l'Open Data], et on ne m'a jamais posé la question, je n'ai jamais eu la sollicitation où on m'a dit : « à terme les données resteront privées ». » (Martine, ingénieure en humanités numériques dans une MSH, novembre 2020).

L'évolution des demandes : de plus en plus nombreuses et anticipées

Les témoignages recueillis mettent en exergue une double tendance concernant l'évolution des demandes des chercheurs.euses dans les années récentes : les demandes sont à la fois croissantes, et formulées de plus en plus en amont des projets de recherche, notamment dans les phases de réponses aux appels à projet.

²⁵ Initialement créé par le CERN sous le nom d'OpenAIRE Orphan Record Repository, Zenodo est un répertoire de données, d'outils et de résultats de recherche, qui constitue une référence en matière d'ouverture de la science : <https://zenodo.org/>

« A l'origine et encore aujourd'hui, sur la base des projets sur lesquels ils travaillent, ils savent qu'on va pouvoir les aider, pour faire une base de données, ou faire un site internet. Ce genre de choses, donc ils vont venir comme ça. Mais de plus en plus, ils nous associent dès la conception du projet, donc ils nous sollicitent dès qu'ils vont travailler sur la demande de financement. Donc on est sollicités de plus en plus en fonction du calendrier des appels à projets ». (Armelle, ingénieure en humanités numériques dans une MSH, décembre 2020).

Cette évolution est en partie expliquée dans les témoignages par la place croissante qu'occupe le double enjeu d'instrumentation technique et numérique des projets, et l'attention renforcée à la gestion des données numériques. Cela se combine à deux tendances observées : d'une part l'exigence croissante, dans les appels à projets, de production de plans de gestion de données et dans certains cas de mise à disposition publique des données numériques. Et d'autre part, un effet de réseau, par lequel des chercheurs.euses ayant bénéficié de l'appui des ingénieurs MSH dans la mise en œuvre du volet numérique de la gestion de leur données soit recommandent à leurs collègues de démarcher les réseaux d'appui, soit servent d'exemple concret pour ces collègues qui tendent à répliquer la procédure d'accompagnement pour leurs propres besoins, en s'inspirant des réalisations concrètes observées. **Cet effet de réseau apparaît jouer un rôle particulièrement fort au regard des enjeux de publicisation des données** : alors qu'il n'est pas aisé de se projeter dans des usages et solutions numériques perçues comme "abstraites", comportant leurs portées, limites et risques projetés, les exemples concrets jouent un rôle majeur de projection dans des usages possibles, comme le raconte Martine dans un entretien en novembre 2020. Ainsi, si les actions de sensibilisation organisées par les réseaux d'appui à destination des chercheurs (par exemple, ateliers sur les outils numériques et la gestion des données, conception d'un Plan de Gestion de données, etc), ont des retombées en termes de nombre de sollicitations, l'effet de réseau semble d'expérience jouer un rôle tout aussi prégnant sur le nombre et la nature des sollicitations reçues.

2.2.3. L'appui juridique : expertise et expériences des délégués à la protection des données (DPD)

L'appui juridique : expertise et expériences des délégués à la protection des données (DPD)

La fonction de délégué à la protection des données (DPD, ou DPO en anglais pour *Data Protection Officer*) est récente et a été créée dans le contexte de la mise en application du RGPD (Règlement européen sur la protection des données personnelles) entré en vigueur en mai 2018. La CNIL définit ainsi, le rôle d'un.e DPD:

*"Le délégué est chargé de mettre en œuvre la conformité au règlement européen sur la protection des données au sein de l'organisme qui l'a désigné s'agissant de l'ensemble des traitements mis en œuvre par cet organisme"*²⁶

Dans le contexte actuel d'une montée en puissance de l'attention à la donnée, notamment suivant un horizon d'ouverture par défaut promu par le mouvement pour une Science Ouverte, le rôle des DPD est particulièrement crucial pour accompagner les chercheurs.euses dans la caractérisation juridique et la mise en conformité des traitements applicables aux données de recherche comprenant des données à caractère personnel. Il s'agit, pour le.la DPD, "de conseiller et d'informer les responsables de traitements de leurs obligations pour l'application des réglementations en matière de protection des données à caractère personnel (cf. article 37 à 38 du RGPD). Il contrôle le respect de la réglementation, coopère avec l'autorité de contrôle, s'assure du respect de la réglementation sur la protection des personnes." (Guide INSHS-CNRS, 2021 : 14).

Dans le contexte de la promotion d'une mise en accès libre des données (numériques) de recherche comme une nouvelle norme, il s'agit donc de démêler ce qui fait et doit faire l'objet d'une protection (et comment, alors, bien gérer la protection de ces données personnelles) de ce qui peut faire l'objet d'une ouverture. La question de la protection des données à caractère personnel est cruciale pour les domaines scientifiques

²⁶ Source : <https://www.cnil.fr/fr/devenir-delegue-la-protection-des-donnees>, consulté le 25/05/2021

relevant des SHS, et en particulier l'ethnologie, où l'essentiel du matériau de recherche concerne des êtres humains et se veut demeurer particulièrement situé²⁷ : une majorité des données produites comportent donc un caractère directement ou indirectement identifiant, et correspondent de ce fait au champ d'application du RGPD (cf article 4 du RGPD).

Dans cette section, il ne s'agit pas de revenir en détails sur l'étendue de ce champ d'application et toutes les implications complexes qu'il revêt en pratique dans le contexte particulier de la mise en œuvre des recherches fondées sur la production et le traitement de données ethnographiques²⁸. Il s'agit davantage de **comprendre comment est vécue et expérimentée cette mise en application du nouveau RGPD dans la pratique des acteurs.trices qui jouent un rôle de premier plan dans l'accompagnement des communautés de recherche sur ce volet juridique, et documenter en particulier quels problèmes ils.elles rencontrent.**

La particularité de la mission des DPD que nous avons interrogés, désignés par des EPST ou des Universités, est qu'elle recouvre la mise en conformité de deux grands types de données : à la fois les **données de la recherche**, et les **données administratives concernant le fonctionnement des établissements concernés**. Le travail associé à ces deux grands types de données renvoie à des formes d'expertise et des enjeux différents. Selon les témoignages que nous avons recueillis, l'un ou l'autre de ces deux volets de leur mission peut recouvrir une portion plus ou moins grande du travail engagé, en fonction de la culture et de la sensibilité historique de l'établissement vis-à-vis de ces enjeux. Par ailleurs, les profils des quatre DPD que nous avons interrogés sont assez variés, avec pour certain.e.s un ancrage initial fort dans des compétences juridiques, et pour d'autres, un profil hybride avec une formation et une carrière dans le domaine de l'informatique et du numérique, complété par une forte sensibilité personnelle aux questions juridiques (c'est le cas par exemple pour Hervé, responsable d'une DSI et désigné DPD au moment de la création de ce statut), ou encore, des profils marqués par une formation thématique dans un domaine scientifique. Anne-Marie correspond à ce dernier profil, et explique être "ni juriste, ni informaticienne" ; mais selon elle le poste de DPD nécessite avant tout d'avoir un "profil polyvalent", avec "une bonne connaissance de l'établissement" où l'on travaille et du "droit du numérique" (Anne-Marie, DPD, novembre 2020).

Cette hétérogénéité des configurations, des profils et des missions des DPD, associée dans certains cas à une prise de fonction très récente, laisse entrevoir **le défi que représente, pour ces acteurs, la mise en oeuvre d'une expertise fine et circonstanciée autour de données à caractère personnel relevant d'une part d'un volet administratif et gestionnaire, et d'autre part de la production et de la gestion de données de recherche pluridisciplinaires**, caractérisées par nature par leur forte diversité.

Couvrir la diversité des disciplines scientifiques traitant des données à caractère personnel

Dans les entretiens que nous avons menés, nous nous sommes focalisés sur la dimension du travail des DPD touchant aux données de la recherche et aux interactions que suscite cette mission avec les communautés de chercheurs.euses. Un constat partagé est celui de l'hétérogénéité de l'origine disciplinaire des demandes: les DPD traitent des demandes issues de **tous les domaines scientifiques représentés** dans leur établissement et qui supposent le traitement de données identifiantes ou personnelles. Ils doivent donc à la fois s'adapter à cette pluralité disciplinaire, mais aussi veiller à l'articulation avec d'autres cadres juridiques applicables, le tout dans un paysage juridique évolutif :

"On travaille avec un large spectre de domaines scientifiques et de recherches impliquant la personne humaine. Il faut aussi articuler notre expertise sur les données personnelles avec d'autres cadres, le code de la santé publique par exemple. Et c'est à nous de faire la veille sur les évolutions du droit". (Anne-Marie, DPD, novembre 2020).

Cette configuration de travail implique donc un contact avec une grande hétérogénéité de types de données, et requiert une certaine souplesse pour accompagner des chercheurs.euses d'horizons divers à la caractérisation juridique de leurs données et à la mise en conformité des traitements qu'ils.elles souhaitent

²⁷ C'est-à-dire que du point de vue de la démarche heuristique, ce matériau fait sens en lien avec la connaissance du contexte et des acteurs auprès desquels il a été produit.

²⁸ Des nombreuses ressources sont disponibles sur ce sujet, notamment le récent guide produit par l'INSHS : InSHS, 2021, "Les sciences humaines et sociales et la protection des données à caractère personnel dans le contexte de la Science Ouverte : Guide pour la recherche". D'autre part, l'agenda des séminaires/webinaires, ateliers, actions d'information et de sensibilisation autour du sujet se densifie depuis 2018.

appliquer. Le témoignage de Julie, DPD interviewée en novembre 2020, rejoint à cet égard un retour d'expérience commun aux quatre entretiens que nous avons réalisés : elle explique que dans ce contexte, la procédure d'accompagnement se construit nécessairement au cas par cas, et nécessite pour elle de comprendre, *a minima* dans les grandes lignes, les enjeux recouverts par le projet de recherche dans lequel s'inscrit un traitement de données personnelles et/ou sensibles.

“On échange par email, et les chercheurs me font un récapitulatif de leur recherche. Si je n’arrive pas à comprendre les réels enjeux, on passe en visio pour se mettre d’accord. C’est vraiment du cas par cas.” (Julie, DPD, novembre 2020)

Moments et origines des sollicitations adressées aux DPD

Comme nous l’avons vu dans le cas des réseaux d’appui technique et scientifique présentés dans la section précédente, les sollicitations adressées aux DPD par les chercheurs.euses renvoient à **deux grandes temporalités de la recherche** : “certains plus **au début**, et certains quand le projet est **déjà lancé**”. (Sandra, DPD, novembre 2020)

Du point de vue des DPD, les sollicitations anticipées sont préférables à celles qui peuvent être formulées au cours d’un projet déjà engagé, notamment parce qu’il peut être difficile de faire machine arrière sur un processus de traitement non conforme déjà mis en oeuvre, comme l’explique Julie :

“Intervenir sur des projets en cours de route, c’est plus embêtant. Parce que ça peut amener à arrêter un pan de la recherche. Parce qu’on se rend compte qu’on n’est pas du tout en conformité sur tout un pan”. (Julie, DPD, novembre 2020)

A l’inverse, Julie explique que si la réflexion sur le sujet de la collecte et du traitement des données personnelles est anticipée, tout (ou presque) est réalisable. Il s’agit alors pour elle, en tant que DPD, d’accompagner les chercheurs.euses à argumenter et ajuster leurs traitements pour entrer en conformité avec la loi, pour ne pas vivre cette mise en conformité comme une entrave à la mise en oeuvre de leur projet :

“Mais le RGPD n’empêche en rien la recherche, quasiment tout est possible, le seul truc c’est que ça doit être encadré. [...] Au contraire, je trouve que le RGPD permet de mieux prendre en compte les spécificités de la recherche scientifique qu’avant. Parce qu’il y a un renversement de la charge de la preuve, parce qu’on a la possibilité de prouver qu’on est conforme par tous moyens”. (Julie, DPD, novembre 2020)

Un enjeu majeur, du point de vue des DPD, est donc de sensibiliser les chercheurs.euses à les contacter le plus en amont possible de la mise en oeuvre des programmes de recherche, notamment dans les **phases de conception**. Ils observent par ailleurs que cette tendance à recevoir des demandes anticipées se renforce, notamment dans le cadre du travail sur les réponses à des appels à projets, qui pose par ailleurs deux problèmes : celui d’une intensification des sollicitations sur certaines périodes en fonction du calendrier des appels à projet des grands financeurs ; et le fait qu’une partie du travail fourni par les DPD dans ce cadre n’aboutisse pas à des traitements réels, pour les projets non retenus par les financeurs. Ce dernier problème est toutefois nuancé dans les discours par l’idée d’une capitalisation sur la progression que ces démarches permettent autour d’une meilleure prise en charge des enjeux juridiques relatifs aux données à caractère personnel, et la possibilité de mettre en oeuvre ces projets dans d’autres temporalités, ou sur la base d’autres sources de financement. Il apparaît donc que pour les DPD, ce qui compte n’est pas tant le résultat (c’est-à-dire la sélection des projets sur lesquels ils ont travaillé aboutissant à la mise en oeuvre du traitement effectif) que l’évolution des pratiques passant par une plus grande sensibilisation et compréhension des communautés de chercheurs.euses vis-à-vis de la prise en charge ce volet juridique.

Par ailleurs, Anne-Marie, DPD dans un autre établissement, dénombre **trois grandes origines aux sollicitations** qu’elle reçoit : le retour des **actions de sensibilisation** que son service met en place, les demandes des **principaux financeurs de la recherche** (notamment en lien avec la création de plans de gestion de données), et plus récemment, des demandes liées à **des exigences formulées par les revues scientifiques**. Elle explique notamment que les deux derniers leviers ont pris une importance majeure dans les tendances actuelles de croissance des sollicitations qu’elle réceptionne :

“Il y a des sollicitations qui reviennent parce qu’on a commencé à sensibiliser. Il y a aussi plus de sollicitations de la part des financeurs, d’abord l’Europe, plus l’ANR, qui demandent une certification RGPD. Et de plus en plus de projets sont acceptés avec maintenant des work-packages éthiques. Donc il y a un gros enjeu avec les financeurs. Et une troisième source, toute neuve, ce sont des revues qui exigent que vous prouviez que vous avez une prise en compte éthique de ces questions. [...] J’ai vu s’intensifier les demandes du côté des financeurs et des revues. Ce sont des incitations qui arrivent par ce cadrage. Donc on n’est plus que dans de la prescription interne, mais on est dans une prescription qui arrive par les revues et les financeurs”. (Anne-Marie, DPD, novembre 2020).

Les DPD interrogé.e.s, et en particulier ceux.celles qui portaient la responsabilité de *correspondant informatique et liberté* avant la création du statut de DPD, observent en effet **une forte croissance des demandes dans les années récentes en particulier depuis 2018**, année d’adoption du RGPD, qui a vu s’intensifier à la fois la communication autour du sujet, et les injonctions des principaux financeurs de la recherche à se mettre en conformité avec ce cadre légal.

Pourtant, résume Sandra, « beaucoup de chercheurs ne se sentent pas concernés, parce qu’ils se disent “je ne collecte que des données qui ne sont pas personnelles”. » (Sandra, DPD, novembre 2020). Elle a donc conscience qu’une grande partie des projets de recherche sont encore menés en dehors de ce cadre, notamment par manque de connaissance sur le champ d’application de ce nouveau règlement et sur les procédures concrètes à mettre en place. Elle explique par ailleurs que les actions d’information et de sensibilisation se multipliant autour du sujet, elle envisage davantage cela comme **un processus qui se déploie progressivement**. Dans son expérience, elle remarque à ce sujet **un effet (et une nécessité) de familiarisation** avec ces nouvelles procédures, et des chercheurs qui petit à petit la consultent systématiquement pour la mise en oeuvre de nouveaux projets : « une fois qu’ils ont fait une première mise en conformité avec une inscription sur le registre [des traitements], ils ont tendance à le refaire plus facilement ensuite ».

Selon son homologue Anne-Marie, une difficulté qui explique que de nombreux projets se déroulent encore en dehors de ce cadre, réside dans **la compréhension précise de ce qu’est une donnée à caractère personnel**, et en particulier le fait qu’il ne s’agit pas uniquement de données « nominatives », mais plus largement de données « identifiantes », ce qui étend largement le champ d’application du RGPD :

« Le problème c’est pas le terme de données, mais c’est de comprendre ce que recouvre le terme “données personnelles”. Il y a une difficulté à comprendre que c’est le caractère identifiant ». (Anne-Marie, DPD, novembre 2020).

Par ailleurs, les DPD que nous avons rencontré.e.s s’accordent à dire qu’ils ne pourraient absorber l’ensemble des demandes de mise en conformité si, demain, tous les chercheurs.euses de leurs établissements en faisaient la démarche. Il existe en effet, dans les faits, un problème de moyens considérable au regard de l’application théorique de ce nouveau cadre à l’ensemble de la recherche intégrant des données à caractère personnel.

Problématiques liées aux données ethnographiques

Tous les DPD que nous avons interrogés n’ont pas nécessairement l’expérience d’avoir travaillé sur le cas de données ethnographiques. Deux d’entre eux.elles, ont plus particulièrement soulevé des problématiques associées, de leur point de vue, à ce type de données.

Un premier problème, souligné par Anne-Marie, est relié à une démarche et une méthode de travail particulièrement mobilisée par les ethnographes, et qui pose des difficultés du point de vue de la qualification des données recueillies, ainsi que du recueil du consentement. Il s’agit, pour le résumer, de toutes les données collectées en situation d’**observation, ou observation participante** :

“Les problématiques que j’ai, c’est surtout sur les techniques d’immersion, de focus group, de débats publics, des réunions par exemple. Comment on dit aux gens qu’on enquête sur eux à ce moment-là, et comment on recueille leur consentement? On est un peu sur de la

donnée sensible. Donc c'est un problème, par exemple sur l'observation, pour faire respecter le RGPD". (Anne-Marie, DPD, novembre 2020).

Cette situation de recueil de données et les traitements occasionnés par une recherche mobilisant l'observation et l'observation participante échappent en effet à une qualification juridique établie, même s'ils entrent bien dans le champ d'application du RGPD (dans la mesure où des données à caractère personnel sont produites et mobilisées). On entre, dans ce cas, typiquement dans une "zone floue" du point de vue de l'interprétation du cadre juridique applicable, alors qu'il s'agit d'une méthode particulièrement importante et structurante dans la démarche ethnographique.

Un autre problème, souligné par Julie, est celui de la **proportionnalité** des données. Il pointe, là encore, une tension entre d'une part l'application du cadre juridique, et d'autre part, le cadre épistémologique de la recherche fondée sur une démarche ethnographique et des méthodes d'enquête immersives. Il s'agit, du point de vue du RGPD, de minimiser la collecte des données personnelles de sorte qu'elles permettent de répondre strictement aux objectifs du projet. Alors que d'un point de vue scientifique, la démarche ethnographique se fonde sur une observation et une collecte élargie de données, selon une démarche inductive et ancrée :

"Ce qui peut poser problème en ethnologie c'est la proportionnalité de la donnée. Donc il faut apporter la justification que les projets d'ethnologie ont une finalité très large en fonction de la collecte. Donc en fait c'est à vous de prouver que vous avez besoin de ces données-là. Donc il faut expliquer par exemple que c'est dans le cadre d'une recherche exploratoire qui va s'affiner". (Julie, DPD, novembre 2020)

Un autre point soulevé par ces deux interlocutrices est celui de la **formalisation du recueil du consentement**, qui, dans les faits, soulève de nombreuses questions et problèmes de la part des ethnographes. Notamment, parce que le format d'un consentement éclairé écrit et formalisé n'est pas applicable dans tous les contextes culturels : se posent par exemple le problème des sociétés à tradition orale, ou encore celui de la crainte de la formalité écrite (par exemple dans des contextes d'enquête sur des pratiques clandestines ou dans des régimes totalitaires), qui modifie le rapport et la relation de confiance à l'enquêteur.trice. Le premier cas peut être solutionné par le recueil du consentement sous forme orale enregistrée, mais la difficulté demeure majeure dans le second cas, où c'est bien la question de la formalisation du consentement qui pose problème, voir risque d'entraver ou de rendre impossible la démarche d'enquête²⁹. Cette ambiguïté autour du recueil du consentement rejoint également le point soulevé en amont de situations d'observation où des données à caractère personnel peuvent être capturées. Julie souligne par ailleurs qu'il existe une confusion autour des situations dans lesquelles le consentement doit être recueilli, et celles où il n'est pas nécessaire :

« Une spécificité de la donnée ethnographique, c'est le recueil du consentement. Mais le consentement n'est pas nécessaire tout le temps, il faut faire attention, il n'est nécessaire que dans certains cas. » (Julie, DPD, novembre 2020)

Open Data et protection des données personnelles

"Des données identifiantes ne seront jamais en Open Data" (Julie, DPD, novembre 2020).

La question du rapport entre ouverture des données de recherche et protection des données à caractère personnel est ici clairement énoncée : il s'agit d'une relation d'exclusion. Néanmoins, une confusion existe en effet, au sein des communautés de chercheurs.euses, sur l'étendue de ce qui peut être rendu accessible au sein des corpus lorsque l'on s'inscrit dans un objectif d'ouverture, et *vice versa*. Cette confusion semble prendre appui sur plusieurs éléments, notamment l'adoption simultanée en 2018 du Plan national pour la Science Ouverte et du RGPD, qui a généré une émulation concomitante autour ces deux mouvements, sans que ne soient toujours exposés (ou compris) clairement leurs lignes de divergences et les points d'interaction. Une autre explication invoquée, et déjà mentionnée est reliée à une connaissance souvent sommaire de la part des communautés de chercheurs.euses de ce que recouvre très concrètement la mise

²⁹ Nous y reviendrons dans la partie 4 à travers des exemples de témoignages de chercheurs.euses qui illustrent des situations problématiques de ce point de vue.

en conformité avec le RGPD, et ce que suppose « ouvrir » des données de recherche au sens de l'ORD – deux cadres distincts qui peuvent par ailleurs être vécus comme des injonctions contradictoires. Le caractère récent et émergent de ces deux cadres renforce également cet effet de confusion, qui est d'autant plus problématique dans les communautés productrices de données ethnographiques pour qui une très grande partie des matériaux de recherche collectés sont directement ou indirectement identifiants (voir partie 4). De surcroît, il existe des situations d'exception vis-à-vis de cette ligne de partage entre donnée ouverte et donnée à caractère personnelle, comme le souligne Anne-Marie. L'équation entre ouvert et fermé n'est ainsi pas si simple, et il existe des cas dans lesquels des données à caractère personnel peuvent être ouvertes:

“Les données à caractère personnel sont protégées, ce n'est pas possible d'ouvrir des données brutes, par exemple une retranscription. Souvent il y a une incompréhension des collègues à ce niveau. Il y a [des] possibilités pour ouvrir des données à caractère personnel : [par exemple] si on a le consentement de la personne, on peut publier des documents autorisés par la loi, comme des organigrammes, ou si on fait une anonymisation complète. [...] ça pose beaucoup d'incompréhension. ça veut dire que mettre des données en open data c'est sous réserve. Par contre, c'est pas parce que ces sont des données personnelles qu'elles ne sont pas réutilisables. C'est juste qu'on restreint la réutilisation” (Anne-Marie, DPD, novembre 2020)

Elle conclut finalement sur cette prévalence d'une approche au cas par cas, y compris dans l'appréhension du rapport entre possibilité d'ouverture et protection des données à caractère personnel, qui relève à la fois du caractère spécifique de chaque projet de recherche (i.e. spécificités disciplinaires et spécificités des données), mais également de l'interprétation qui va être faite du cadre juridique au regard de chaque situation, par le DPD consulté.

« Les choses se font sans que ce soit si clair. Dans beaucoup de situations on est au cas par cas. Et on a probablement autant d'interprétations possibles que de délégués ». (Anne-Marie, DPD, novembre 2020)

La question des moyens

Un dernier point problématique est unanimement soulevé par nos quatre interlocuteurs.trices : il s'agit de la question structurelle des moyens alloués à la prise en charge de cette mise en conformité des traitements de données de la recherche. Il ressort des témoignages que nous avons collectés le sentiment partagé d'un écart important entre l'ambition portée par le cadre juridique, et les moyens attribués pour sa mise en application pratique à l'ensemble des données de recherche produites et traitées dans le contexte de la recherche publique en France. Le manque de moyens est décrit à plusieurs niveaux, à la fois à l'échelle des DPD eux-mêmes et de leurs services, dans leur capacité à fournir un accompagnement circonstancié et à la mesure du volume de demandes qu'ils.elles sont susceptibles de recevoir si toutes les données de recherche faisaient l'objet d'une mise en conformité ; mais aussi à l'échelle des chercheurs.euses, qui doivent absorber de plus en plus de contraintes professionnelles dans un contexte de réduction des moyens alloués. De fait, ce contexte tend à favoriser la réticence des communautés de chercheurs.euses à se saisir de ces nouveaux cadrages qui font évoluer leurs pratiques, et qui sont abordés dans le contexte d'une accumulation de contraintes, générant selon Anne-Marie, un sentiment de “lassitude” :

“On lutte contre le manque de moyens, et en même temps... “ [...] Ce que je ressens, c'est la lassitude des collègues, comme un marathon sans règles. Pour les chercheurs, il y a vraiment une lassitude face à l'inflation des demandes qui leur sont faites : ils doivent être bilingues, statisticiens, juristes... C'est des contraintes trop fortes et trop lourdes”. (Anne-Marie, DPD, novembre 2020).

Elle relie finalement ce sentiment de « contrainte » aussi bien aux démarches de mise en conformité avec le RGPD, qu'à celles qui relèvent de l'adoption des principes FAIR liés à l'ouverture des données. Dans son expérience en effet, elle ressent que le même mécanisme d'investissement en temps de travail sur la donnée est à l'oeuvre que l'on s'engage dans l'une ou l'autre, ou dans ces deux formes de conformations. Elles impliquent de nouvelles pratiques et formes de prise en charge des données de recherche que celles

existantes jusqu'à présent et tournées prioritairement vers l'analyse finalisée et la publication des résultats de recherche. Le constat énoncé par Anne-Marie est que travailler à rendre les données réutilisables, en appliquant les principes FAIR, "aide les chercheurs, mais pour plus tard" : selon elle, cela amène notamment "à faire du rangement", dans une logique de continuité d'activité. Mais la manière dont cela est vécu au présent est que ces tâches s'insèrent dans le contexte d'une charge courante qui est déjà très lourde, et n'apparaissent ni comme prioritaires, ni en adéquation avec les moyens alloués pour les mettre en œuvre.

Bien qu'il apparaisse être vécu comme une contrainte, notamment dans un contexte de tension autour des moyens, l'enjeu d'une meilleure prise en charge de la protection des données à caractère personnel semble néanmoins, d'après les DPD interrogés, faire l'objet d'une attention croissante de la part des communautés de chercheurs.euses. Ce constat est relié d'une part au fait qu'ils.elles observent une évolution croissante des sollicitations qui leur sont adressées, même si cette tendance est largement reliée à un facteur extérieur, à savoir les injonctions portées par les financeurs de la recherche dans le cadre des appels à projets, comme nous l'avons souligné. D'autre part, Sandra (DPD, novembre 2020) souligne également un effet plus général de sensibilisation des citoyens à ces enjeux, qu'elle associe au constat d'une société de plus en plus numérisée, où « les "data" sont partout ». Elle explique ainsi que les chercheurs, habitués eux-mêmes « à cliquer sur des boutons "j'accepte les conditions d'utilisation de mes données personnelles" », portent, dans le cadre de leurs travaux de recherche, une attention croissante à ce sujet qui touche l'ensemble de la société.

A RETENIR

- Que ce soit sur les aspects techniques (numériques), pratiques (gestion des données) ou juridiques, **l'accompagnement mis en œuvre par les réseaux d'appui à la recherche s'inscrit dans une démarche au cas par cas**, qui se construit avec le.la chercheur.euse pour prendre en compte les spécificités de sa recherche et de ses données.
- **Les réseaux d'appui sont structurés autour de compétences transversales et pluridisciplinaires, qui rendent difficile la mise en œuvre d'un accompagnement circonstancié autour des spécificités disciplinaires.**
- **A ce titre, la caractérisation juridique des données ethnographiques apparaît particulièrement épineuse.**
- **Les données ethnographiques sont relativement peu représentées parmi les types de données faisant l'objet de demandes d'accompagnement.**
- Les demandes formulées (en particulier sur le plan juridique) sont **davantage centrées sur des enjeux de protection des données** qu'en lien avec des enjeux de partage et de publicisation
- Il existe **une évolution de la quantité et de la nature des demandes formulées auprès des réseaux d'appui à la recherche sur la période récente, liée à une accentuation de l'attention institutionnelle vers les données de la recherche et à une plus grande prise en compte des questions de gestion des données numériques dans les appels à projet.** De plus, les demandes sont formulées de plus en plus en amont des projets.
- Encore relativement peu de demandes formulées auprès des réseaux d'appui scientifiques et techniques sont directement liées au mouvement de l'ouverture des données et à l'utilisation des entrepôts de données. **Les formes d'accompagnement au partage des données numériques s'inscrivent davantage en continuité avec les pratiques, motifs et formes d'ouverture préexistant à ce mouvement**, et sont pensées en relation avec les préoccupations disciplinaires et les objectifs de recherche propres à chaque programme.
- **L'engagement des communautés d'appui à la recherche dans le mouvement de l'Open Data est à la fois le produit de dynamiques descendantes et ascendantes** : elle modifie leur rôle, leur poids et leur légitimité dans l'enseignement supérieur et la recherche.
- Les dispositions juridiques et institutionnelles contraignantes sont presque systématiquement accompagnées, lors de leur porter à connaissance dans les actions de formation et de sensibilisation, d'un **effort de plaidoyer pour l'Open Science, présentée comme une opportunité pour la recherche dans son ensemble et pour les chercheurs.euses à titre individuel.**
- L'expérience des réseaux d'appui à la recherche met en évidence **le sentiment partagé que tant le mouvement d'ouverture (FAIRisation de données) que celui de mise en conformité de la gestion des données à caractère personnel sont perçus par les communautés de chercheurs.euses comme une contrainte.**
- **Cette forte appropriation des enjeux de l'Open Science par les communautés d'appui à la recherche contraste ainsi fortement avec la densité des problèmes que suscite la mise en œuvre concrète de l'Open Research Data par les chercheur.e.s en SHS.**

3. État des connaissances sur les pratiques des chercheur.e.s. qualitaticien.ne.s en France

Où en sont les communautés de recherche en SHS vis-à-vis de ce mouvement « vers » les données? L'implication d'une pluralité d'acteurs, nouveaux ou plus anciens, dans ce mouvement a renouvelé l'intérêt porté aux pratiques individuelles des professionnel.le.s qui interviennent dans leur traitement et a donné lieu à un effort significatif de collecte d'informations, largement décentralisé, par la conduite d'enquêtes aux objectifs et perspectives diversifiés.

Les enquêtes menées auprès des communautés de recherche en SHS peuvent être regroupées en deux grands ensembles. Le premier ensemble (3.1.) concerne des enquêtes à échantillon large visant principalement à évaluer le degré d'appropriation, par les intéressé.es, de leurs droits et de leurs obligations, d'une part, et le degré d'instrumentation et de compétences techniques, d'autre part. L'objectif souvent explicite de ces enquêtes est de contribuer à un état des lieux et des moyens en vue d'une mise en conformité avec les cadres juridiques et institutionnels.

Le second ensemble (3.2.) regroupe des enquêtes sur les pratiques généralement plus qualitatives, adossées à des programmes de recherche spécifiques et dont les résultats ont donné lieu à une formalisation. Au regard du nombre important de communautés qui se confrontent actuellement à la question de l'ouverture et à la protection des données de recherche, par exemple dans le contexte du lancement de projets ANR incluant une obligation de production de PGD en début de projet, ces retours d'expériences formalisés, et donc visibles, sont très peu nombreux. Nous nous appuyons ici sur 3 d'entre eux : BeQuali, Méta-projet et SAWA.

3.1. Enseignements des enquêtes récentes à échantillon large sur les pratiques des communautés de recherche en SHS

3.1.1. Panorama général

Il est encore malaisé, en 2021, d'opérer une recension systématique des enquêtes sur les pratiques de gestion des données de recherche en SHS en France. Il est très probable, en effet, que davantage d'enquêtes sont en cours ou non publiées, qu'il n'existe d'enquêtes publiées. Cela est dû, bien sûr, à la mise à l'agenda récente du sujet dans l'ensemble des EPST français, mais également au fait qu'il s'agit d'une question rarement portée, au sein des communautés intéressées, par des chercheurs.euses publiant.es. **Les enquêtes publiées sont ainsi très majoritairement inscrites dans une perspective de recherche finalisée, c'est-à-dire qu'elles participent d'un diagnostic préalable à la structuration d'une politique de gestion des données à l'échelle d'un établissement ou d'un laboratoire, fondée sur l'analyse des pratiques de "publics cibles"**. Une recension provisoire et incomplète de ces enquêtes et de leur focale est présentée dans le [TABLEAU 1](#).

Tableau 1. Recension provisoire des enquêtes sur les pratiques de gestion des données de recherche en SHS en France

Echelle	Structure porteuse	Titre	Focale	Référence
Nationale	Couperin	Pratiques de publications et d'accès ouvert des chercheurs français	Toutes disciplines Publications, une partie du questionnaire consacrée aux données	Rousseau-Hans et Ollendorf, 2020
	Réseau des Urfist	Les usages et pratiques des outils numériques chez les chercheurs en France	Toutes disciplines Outils numériques, une partie du questionnaire consacrée aux données	N.P.
	CommonData	Pratiques de gestion des données de la recherche: une nécessaire acculturation des chercheurs aux enjeux de la science ouverte?	Toutes disciplines Qualification juridique des données et rapport des chercheurs à la propriété	Amiel et al., 2018 (enquête exploratoire) N.P.
Universités (appui Urfist, SCD)	Lille 3	Les données de la recherche en SHS. Une enquête à l'Université de Lille 3	Données SHS (inspirée par les enquêtes de Strasbourg et Berlin)	Prost et Schöpfl, 2015
	Rennes 2	Données de la recherche en SHS. Pratiques, représentations et attentes des chercheurs: une enquête à l'Université Rennes 2	Données SHS (inspirée par Lille)	Serres et al., 2017
	Bordeaux Montaigne	Les données de la recherche à l'Université Bordeaux Montaigne: synthèse d'une enquête qualitative auprès des chercheurs	Données SHS (inspirée par Rennes et Lille)	Duprat, 2019
	Montpellier 3	Usages et pratiques en lien avec les données de recherche. Une enquête menée auprès des chercheurs de l'Université Paul Valéry Montpellier 3	Données SHS (inspirée par Bordeaux, Rennes et Lille)	Dillaerts et al., 2020
	UMR AAU-Cresson	Les données de la recherche AAU-Cresson: résultats de l'enquête sur les usages des chercheurs, doctorants et ingénieurs en matière de gestion des données	Enquête interne au laboratoire, inspirée par les grandes enquêtes des Universités	Mosnier et Acquier, 2019
Internes laboratoires ou groupes de laboratoires	UMR Pacte	Enquête mixte (questionnaire et entretiens) en cours	Science ouverte et ouverture des données SHS qualitatives et quantitatives	N.P.
	SO-Maté: Sud Ouest	Enquête mixte (questionnaire et entretiens) en cours	Science ouverte et ouverture des données SHS qualitatives et quantitatives	N.P.

3.1.2. Synthèse des résultats des enquêtes à large échantillon sur les pratiques des chercheurs.euses

L'enquête du consortium Couperin *Pratiques de publication et d'accès ouvert des chercheurs français en 2019* (Rousseau-Hans, Ollendorff, et Harnais 2020), réalisée dans le cadre du Plan National pour la Science Ouverte, porte à la fois sur les pratiques et sur les opinions des chercheurs au regard du mouvement d'ouverture de la recherche, en particulier la publication. Elle comprend uniquement des questions fermées, mais a donné lieu à des commentaires extrêmement nombreux des enquêtés. Sa conception s'appuie sur des enquêtes disciplinaires antérieures, menées en interne à l'INRA et à l'INSERM. L'enquête n'est pas limitée aux SHS. 11658 professionnels de la recherche publique y ont répondu, soit environ 10% de la communauté. Les répondants relevant de la catégorie "Lettres et sciences humaines" ou de la catégorie "Droit, économie, politique, gestion" représentent 27% de l'échantillon (3159 réponses).

L'essentiel des résultats exploités et publiés concerne le rapport à l'édition des chercheurs.euses (*Open Access*), l'un des volets de l'enquête est consacré à l'accessibilité des données de la recherche, de plus en plus fréquemment exigée par les éditeurs scientifiques. Bien que les chercheurs.euses en SHS soient *a priori* moins concernés par ce type d'exigence, l'enquête montre que les différences disciplinaires sont peu marquées dans les réponses: 20% des répondants jugent que cette pratique est inutile, voire constitue un frein à la publication. Par ailleurs, les commentaires font apparaître une crainte de "captation" des matériaux de recherche, y compris à des fins marchandes. Les répondants sont en revanche globalement favorables au dépôt des articles en archive ouverte sur les plateformes publiques.

L'enquête *Pratiques de gestion des données de la recherche* est réalisée dans le cadre du projet de recherche *CommonData*. Des résultats provisoires, issus d'un échantillon de chercheurs.euses localisé.e.s dans la région de Montpellier (n=158), ont fait l'objet d'une première publication en 2018 (Amiel et al. 2018). La particularité de cette enquête est de se concentrer sur la caractérisation et l'évaluation de la conformité des pratiques de gestion des données sur le plan juridique, tout en s'intéressant au rapport qu'entretiennent les chercheur.e.s avec la *propriété* de leurs données. L'enquête met donc explicitement de côté la dimension scientifique des pratiques. 30% des 158 répondant.e.s à l'enquête étaient issu.e.s des SHS. De façon globale, les auteurs du rapport d'enquête estiment que "les réponses (des chercheurs.euses) traduisent une méconnaissance des concepts politico-juridiques à l'œuvre pour la classification des données, pourtant nécessaire à leur gestion raisonnée." (Amiel et al. 2018: 158). Les auteurs insistent sur les différences disciplinaires assez fortes qui émergent de l'analyse des réponses : l'enquête confirme par exemple **le caractère plus souvent individuel et plus rarement délégué de la collecte des données dans ces disciplines**, par contraste notamment avec ce qui se pratique dans le domaine des sciences biophysiques et mathématiques. Des différences sont également observables en ce qui concerne le "sentiment de propriété" des chercheurs.euses vis-à-vis des données qu'ils.elles produisent: interrogées sur le sort qu'elles réserveraient à leurs données si elles quittaient leur laboratoire, 77% des personnes enquêtées en SHS déclarent qu'elles emporteraient leurs données avec elles, et les ¼ de celles-ci qu'elles en laisseraient une copie à leur laboratoire d'origine - ce qui est sensiblement supérieur à ce que déclarent par exemple celles qui sont issues des sciences de la vie. Les réponses traduisent un primat accordé à la "légitimité auctoriale" sur la légitimité du financeur des recherches, ainsi qu'une nette corrélation entre la quantité de travail que demande la collecte des données et le sentiment de propriété personnelle qui leur est attaché. Par ailleurs, le statut public semble très majoritairement associé par les répondant.e.s à la publication d'articles et non à la matière première qui en nourrit l'écriture. L'enquête exploratoire *CommonData* est donc riche d'enseignements, même si la focale retenue pour analyser les résultats conduit les auteurs à retenir *in fine* comme conclusions marquantes "la méconnaissance de concepts qui se trouvent à la croisée du droit et de la politique, l'absence de représentation des enjeux au-delà de la simple conservation des données pour la poursuite de leurs recherches par les chercheurs ou leur valorisation économique, ainsi que les contradictions entre les pratiques et les représentations" (P.167). Les auteurs s'interrogent, dans ce contexte, sur la réelle capacité des pouvoirs publics de contraindre les pratiques de communautés de recherche qui fonctionnent en général sur des principes d'auto-régulation...

L'enquête de l'Université Lille 3 (Prost et Schöpfel 2015) constitue sans doute la première du genre en France, bien qu'elle soit présentée comme prenant entre autre appui sur une démarche similaire menée à l'Université de Strasbourg (non publiée) : comme celles qui vont s'en inspirer par la suite, elle s'inscrit dans une perspective de recherche finalisée, visant à la fois à dresser un état des lieux des pratiques, et à éclairer l'élaboration d'une stratégie et d'une politique de gestion des données à l'échelle de l'Université. Cette perspective oriente la façon dont les données de l'enquête sont recueillies : l'hypothèse sous-jacente est qu'il existe des "besoins" et des "attentes" des communautés de praticiens en matière d'information et d'équipement, et les questions de recherche sont plutôt formulées comme des préoccupations organisationnelles et opérationnelles : "comment conserver ses données, comment les signaler et mettre à disposition d'autres chercheurs, comment faire le lien avec les publications, comment les intégrer à une politique de libre accès à l'information scientifique?" (p.3). Toutes les disciplines représentées à Lille 3 sont concernées (FIGURE 3), c'est-à-dire l'ensemble des sciences humaines et sociales. Le questionnaire a été renseigné par 270 répondants, soit 15% de la population cible. L'échantillon final met en lumière l'hétérogénéité des statuts des professionnels de la recherche, en particulier le fait que 4 répondant.e.s sur 10 ne disposent pas d'un emploi stable dans la recherche et que la plupart cumulent des activités d'enseignement et de recherche - situation logique dans un contexte universitaire.

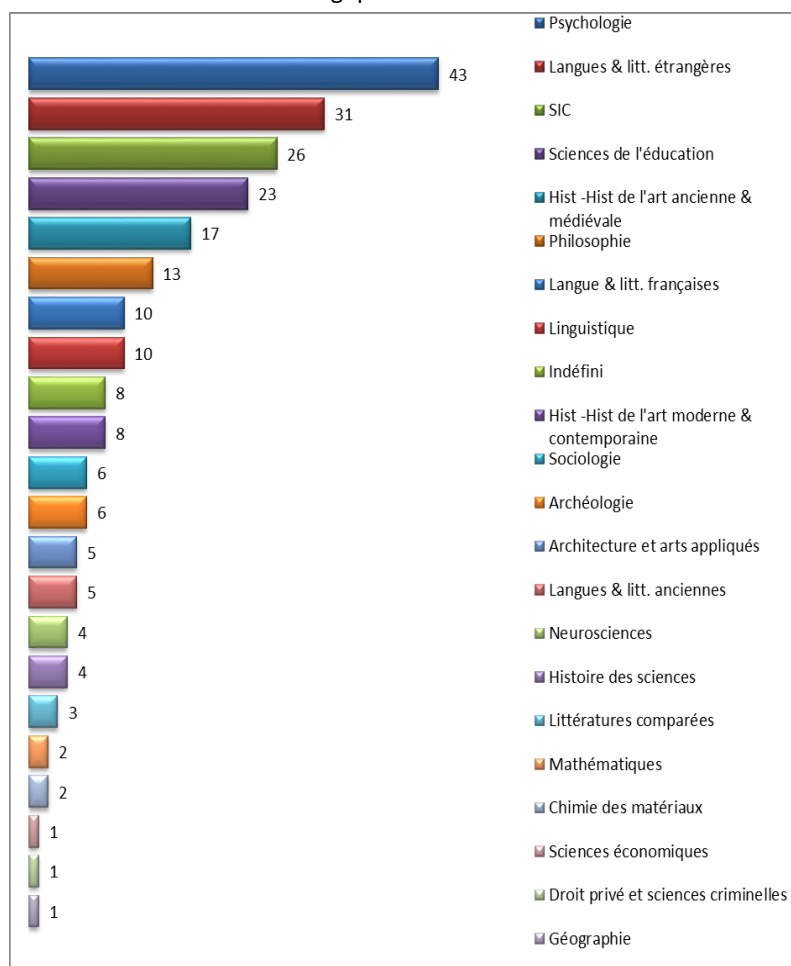


Figure 3. Les disciplines représentées dans l'enquête *Les données de la recherche en SHS de l'université Lille 3* (source: Prost & Schöpfel, 2015)

Un très large ensemble de disciplines sont représentées dans l'enquête. L'anthropologie ou l'ethnologie ne sont pas identifiées par les 217 répondants à cette question comme des disciplines de rattachement, 6 d'entre eux mentionnent la sociologie. La psychologie, la littérature, les langues représentent plus du tiers des réponses.

Sur un plan conceptuel, l'enquête s'appuie sur la distinction entre "données sources" et "données résultats" retenue par une enquête menée en 2013 à l'Université Humboldt (Berlin) (Simukovic, Kindling, et Schirmbacher 2013) pour documenter les pratiques d'archivage, de partage et d'ouverture des données de recherche. En ce qui concerne les données sources, les résultats font ressortir la très grande hétérogénéité des données mobilisées dans les recherches en SHS, qu'elle s'appréhende à l'échelle de l'ensemble des répondants ou à l'échelle individuelle, et la prédominance (FIGURE 4). Il s'agit, comme le relèvent les auteurs, principalement de "small data, voire de(s) dark data" (p.10) peu standardisées et peu accessibles.

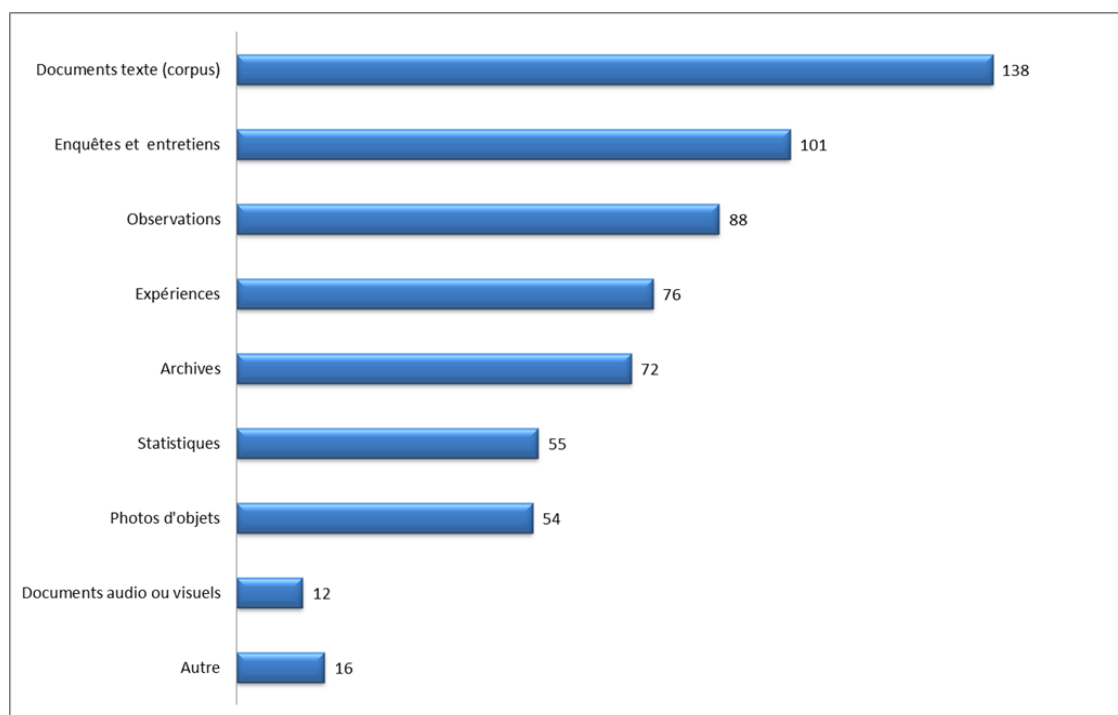


Figure 4. Données sources mobilisées par les répondants à l'enquête *Les données de la recherche en SHS de l'université Lille 3* (source: Prost & Schöpfel, 2015)

La majorité des répondants déclare ne pas partager ses données. Les 36% qui en ont fait l'expérience évoquent plutôt un partage de données faisant l'objet d'un traitement intensif, les rapprochant de "données résultats" (archive ouverte, publication d'un article). Le cadre de partage des "données sources" semble circonscrit au partage entre collègues impliqués dans un même projet de recherche. Pour les auteurs, l'absence de pratique de partage est liée à la disponibilité de l'information et d'outils adaptés: 30% des répondants déclarent ainsi manquer de connaissances et d'informations concernant le partage des données de la recherche. Ceux qui se déclarent ouvertement réticents sont peu nombreux (5 à 10%).

L'un des résultats qui nous intéresse particulièrement dans le contexte de notre propre enquête (PARDOQ) est le fait que les besoins d'appui identifiés par les répondants sont en premier lieu de nature juridique (50%), dans une moindre mesure technique (40%) et éthique (24%).

Par ailleurs, les pratiques de stockage et d'archivage semblent plus artisanales, moins maîtrisées et moins organisées à l'échelle collective qu'elles n'apparaissent chez les chercheurs.euses berlinois.e.s. Ce résultat peut être relié au caractère précoce de l'enquête : il fournit un point de référence avec les enquêtes ultérieures, qui ont été réalisées à des stades plus avancés du déploiement de l'offre institutionnelle de services de stockage sécurisé.

L'enquête menée auprès des enseignants-chercheurs de l'Université Rennes 2 *Données de la recherche en SHS : Pratiques, représentations et attentes des chercheurs* (Serres et al. 2017) s'inspire largement de cette initiative. Elle a été lancée par l'URFIST de Rennes, avec le soutien des instances de l'Université, de son

service de documentation et de la MSH Bretagne. Il s'agit, comme à Lille 3, d'une enquête par questionnaire (n=143), avec un fort taux de réponse pour ce type d'enquête, proche de 30% de la population cible. Une vingtaine de répondants ont par ailleurs accepté un entretien d'une trentaine de minutes, en complément. La méthodologie, comme les résultats de cette enquête sont très convergents avec ceux de l'enquête menée à Lille 3.

La diversité des types de données "sources" est très importante, qu'elle s'apprécie à l'échelle de l'ensemble des répondant.e.s, à l'échelle des répondant.e.s d'un même laboratoire ou à l'échelle individuelle (ainsi, 83% d'entre eux.elles en manipulent plus d'un, et 42% plus de quatre). Les trois types de données les plus citées sont les données textuelles (citées par 72% des répondants), les données d'enquête (53%) et les données d'observation (43%). Un tiers des répondants déclarent par ailleurs travailler également à partir de données images ou multimédia.

Les auteurs de l'étude relèvent à la fois des lignes de partage entre le domaine Arts-lettres-langues d'une part, et les Sciences sociales de l'autre, et une difficulté à dégager des tendances disciplinaires. Les "données résultats", dans les deux cas mais avec des proportions variables, sont avant tout textuelles, et dans une moindre mesure graphiques, puis statistiques. Ils soulignent également la perplexité des répondants face à la notion de "donnée" et à la distinction entre "source" et "résultat", omniprésentes dans le questionnaire. Ce malaise offre un contraste saisissant avec leur capacité à décrire en entretien très précisément les opérations de transformation successives qu'ils mettent en œuvre face à un matériau de recherche (Serres et al. 2017 : 7). La valeur attachée aux données de recherche s'exprime, dans les entretiens, principalement en terme de temps de travail et de pénibilité: c'est le travail du chercheur sur le matériau, mais aussi le savoir-faire du chercheur, qui lui confèrent à ses yeux une valeur et qui construisent un rapport d'attachement spécifique.

Comme le soulignent les auteurs :

"Paradoxalement, l'attachement aux données s'exprime davantage en creux, dans la perte, leitmotiv récurrent des entretiens, qu'elle soit réelle avec « des expériences douloureuses de perte des données » (incidents matériels, obsolescence des supports et formats de sauvegarde) ou fantasmée (vol). Mais autant la perte de la donnée « vive » et active est ressentie vigoureusement et brutalement (un « incendie »), autant la perte de la donnée « dormante » de projets de recherche antérieurs est constatée mais vécue sans drame. La masse et la diversité des données suscitent aussi quelques commentaires. Ce foisonnement est diversement perçu, suivant le point de vue où l'enquêté se place : producteur, auteur... Les chercheurs se satisfont généralement de leur méthode personnelle, et ne semblent pas en demande." (op.cit. p.8)

Ce constat rejoint celui que les auteurs formulent au regard de l'intérêt relativement faible pour les données en tant que "problématique largement partagée et débattue" exprimé dans les entretiens : en 2017, à Rennes, **les répondants ne semblent pas considérer le sujet comme brûlant, ce que les auteurs relient à l'inexistence d'un débat au sein des laboratoires de SHS ou des instances universitaires** - qu'ils appellent par ailleurs de leurs vœux. Les auteurs de l'étude apprécient en effet ces résultats plutôt sur le plan de l'appropriation (souhaitée) des pratiques d'ouverture, soulignant l'ampleur des changements à accomplir et à accompagner pour les communautés d'appui à la recherche, dont ils font partie.

L'enquête à Rennes 2 de Serres et al. approfondit trois volets des pratiques: les pratiques de stockage et d'archivage, les pratiques de partage et de diffusion, les éventuels besoins et attentes.

Sur la question du stockage et de l'archivage, l'étude montre une faible utilisation des services de stockage mis à disposition par les institutions (Université ou TGIR), utilisés seulement par 10% des répondant.e.s. Les répondant.e.s soulignent presque unanimement l'absence de politique d'archivage organisée à l'échelle de leur laboratoire, politique dont l'émergence semble directement reliée à l'urgence d'héberger un très gros volume de données sur une infrastructure *ad hoc* - problème qui ne se pose que ponctuellement en SHS. Cette absence de politique à l'échelle du laboratoire renvoie les chercheurs.euses à des pratiques

personnelles de stockage et d'archivage, très hétérogènes, y compris lorsqu'il s'agit d'archivage de longue durée sur des données "froides". Si la majorité des personnes interrogées déclarent documenter leurs données, elles le font en général à l'aide d'une documentation personnelle (95%) et sans s'appuyer sur un standard de métadonnées (utilisé uniquement par 11% d'entre elles).

Le paysage que dessine l'enquête en matière de partage et de diffusion est particulièrement contrasté : 75% des répondants n'ont jamais déposé leurs données de recherche en libre accès (qu'il s'agisse de données "sources" ou de données "résultats"), et plus de la moitié d'entre eux n'ont jamais téléchargé (réutilisé) les données produites par d'autres chercheurs ni partagé avec d'autres les leurs propres. Les motivations au partage les plus citées relèvent de la mise en visibilité de leurs recherches (65%) et de leur souhait d'établir de nouvelles coopérations scientifiques (54%). Les incitations institutionnelles sont identifiées par un tiers d'entre eux comme importantes. Les freins au partage les plus fréquemment cités sont plus spécifiques aux communautés enquêtées, puisque les freins juridiques liés à la nature des données sont évoqués par 54% des répondants, le manque de temps arrivant en seconde position. Serres et al. proposent une répartition de ces freins en 2 ensembles: les freins externes et internes, ces derniers étant davantage "liés aux perceptions, aux représentations personnelles des enseignants chercheurs" et constituant, selon les auteurs, l'essentiel des obstacles à l'ouverture des données. Cette catégorisation conduit les auteurs à énoncer les tensions entre pratiques constatées et opinions énoncées comme une contradiction: "des pratiques de libre accès minoritaires, qui s'expliquent par différentes perceptions négatives de la diffusion des données, mais qui sont contredites par une position de principe favorable au partage des données." (*op.cit.* p.14)

Comme l'enquête lilloise dont elle confirme et affine les conclusions, l'enquête menée à Rennes 2 est riche en données. Mais elle montre aussi à quel point le travail interprétatif achoppe sur la difficulté à dépasser une lecture normative (les pratiques ne sont pas satisfaisantes), très liée au contexte : y a-t-il vraiment paradoxe ou contradiction? Les "freins" ou les "obstacles" se situent-ils du côté des producteurs de données? Ou alors s'agit-il de constater l'écart entre les représentations et problématiques des enquêteurs et des enquêtés en matière de gestion des données?

Sur les deux dernières années, un dispositif d'enquête proche a été retenu par les universités de Bordeaux-Montaigne (Duprat 2019) et de Montpellier 3-Paul Valéry (Dillaerts et al. 2020). L'enquête bordelaise, qui bénéficie de l'antériorité des études lilloise et rennaise, confère aux entretiens qualitatifs une plus grande importance que ces dernières. Sur le fond, les résultats sont convergents. A l'UBM, un volet spécifique consacré au partage et à l'archivage des données intéresse plusieurs chercheurs.euses, en particulier dans le cadre de programmes de recherche spécifiques, a été introduit : les réponses montrent, une fois encore, la grande hétérogénéité des pratiques et le développement du recours à des plateformes de stockage et de partage en ligne, en général privées, pratiqués par 46% des répondant.e.s, même si beaucoup disent souhaiter s'en affranchir. Duprat (2019) attribue cette pratique aux limites techniques des outils et plateformes institutionnels, ainsi qu'à leur faible notoriété au sein des communautés. Les entretiens font apparaître, en parallèle, un haut niveau de préoccupation et d'insatisfaction des répondant.e.s vis-à-vis de leurs propres pratiques (voir notamment Duprat 2019 : 15). L'un des apports de l'enquête bordelaise est aussi de mettre en évidence le rapport au partage des données de la recherche, pensé par les répondants comme assez spécifique à leur(s) champ(s) disciplinaire(s): aucun.e d'entre eux-elles ne se déclare par principe opposé.e au partage des données. Mais la moitié des répondant.e.s soulignent spontanément le caractère très individuel du processus de recherche, par contraste avec ce qui peut se pratiquer par exemple en sciences biophysiques³⁰: le partage n'est donc pas, pour eux, un impératif incontournable qu'il leur faut organiser et rationaliser.

³⁰ On peut relever à cet égard que les pratiques de publication collective sont également beaucoup plus circonscrites en SHS : ainsi (chiffres 2018), plus de 50% des publications en SHS des chercheurs.euses du CNRS ne comportent qu'un.e seul.e auteur.e, proportion qui atteint 90% dans certaines disciplines, comme l'histoire ou l'anthropologie (Dassa 2019), cette dernière discipline étant parfois décrite comme la plus "artisanale" et "bricolée" des sciences sociales (Olivier de Sardan 2004).

L'enquête montpelliéraine (n=81), la plus récente, apporte des éclairages supplémentaires sur les conditions concrètes de la rencontre entre les personnes interrogées et cet impératif, pratique ou normatif, de se pencher sur la gestion de leurs données de recherche. Ainsi, si 9 personnes enquêtées sur 10 se déclarent intéressées par "les problématiques liées aux données de recherche (stockage, mise à disposition et partage)", seule une personne sur trois a déjà porté un projet de recherche financé. Cette enquête aborde également, de façon inédite, la question des données "sensibles"³¹: 79% des répondant.es ne se sentent pas concerné.es. Les autres ont livré des précisions évoquant à la fois la nature des données, les conditions de leur recueil et la prise en compte de cadres juridiques et déontologiques existants (FIGURE 5).

Répondants	Données sensibles évoquées et citées par les répondants
Répondant 1	« Dans le domaine de la prison »
Répondant 2	« Vie privée des Interessés ou de tierces personnes, éventuellement incriminations de tierces personnes. »
Répondant 3	« données patrimoniales (archéologiques : données sont détruites au fur et à mesure qu'on les étudie, donc fort enjeu autour de leur enregistrement pérenne) »
Répondant 4	« Données personnelles, médicales»
Répondant 5	« anonymat et pratique de drogues, déviance, exclusion»
Répondant 6	« Données personnelles et privées (mais je ne suis pas sûr que cela corresponde à des « données sensibles»)»
Répondant 7	« archives de renseignements généraux »
Répondant 8	« données cliniques (patients) »
Répondant 9	« entretiens avec des professionnels liés à des clauses de confidentialité »
Répondant 10	« entretiens avec des acteurs identifiables et potentiellement menacés »
Répondant 11	« intimité familiale, drame familiaux »
Répondant 12	« Données personnelles »
Répondant 13	« discours »
Répondant 14	« Données sur le fonctionnement interne des entreprises, mécanismes internes à chaque entreprise de gouvernance et de régulation des ressources humaines »
Répondant 15	« santé »

Figure 5. Les "données sensibles" évoquées par les chercheurs.euses de Montpellier-Paul Valéry (Source: Dillaerts, 2020, p.24)

Mais ces enquêtes plus récentes mettent en évidence l'émergence de politiques coordonnées de gestion des données à l'échelle des laboratoires, d'une sensibilisation et d'un soutien technique renforcés, qui ne se traduisent pas (encore?) par une évolution des pratiques, toujours très individualisées, de stockage et d'archivage. Au regard de ce qui précède, la question se pose donc de savoir si cette "politique de l'offre" est susceptible de susciter un "besoin" qui n'est pas *a priori* fortement exprimé ni identifié par ses destinataires, donc un changement de comportement.

Comme les enquêtes menées à l'échelle des Universités présentées dans la section précédente, dont elle reprend la méthodologie (questionnaire, ici administré en ligne, et entretiens complémentaires), l'enquête **AAU-CRESSON** (Mosnier, Acquier, et Dom 2019) s'inscrit dans une perspective de recherche finalisée liée à l'évolution du contexte législatif et réglementaire en matière de gestion des données. Elle est cependant construite à l'échelle d'un laboratoire et exploite davantage, de ce fait, les éléments recueillis dans le cadre d'entretiens semi-directifs, réalisés auprès de 18 personnes, soit 1/3 de l'effectif de l'équipe CRESSON (Centre de Recherche sur l'Espace Sonore et l'environnement urbain) de l'UMR Ambiances, Architectures, Urbanités

³¹ Les conditions dans lesquelles cette catégorie a été présentée ou explicitée auprès des enquêté.es ne sont pas précisées.

(AAU). L'enquête en ligne a, elle, été administrée à l'échelle des deux équipes du laboratoire (CRENAU et CRESSON). L'enquête vise ainsi à améliorer la gestion du "cycle de vie" des données et à "sensibiliser l'équipe pour préparer une future politique de gestion des données" (p.2). **De nombreux laboratoires ont pu initier ce type d'enquête au cours des trois dernières années, généralement sous l'impulsion de la direction ou des équipes d'appui à la recherche. Toutefois, peu en ont publié les résultats.** L'intérêt de l'enquête CRESSON réside également dans le fait que l'équipe CRESSON est une équipe interdisciplinaire historiquement fédérée autour de l'objet son, mais dont les matériaux de recherche se sont diversifiés au cours du temps. Les répondant.es sont, pour 5 d'entre eux-elles, des sociologues. Les résultats de l'enquête sont globalement très convergents avec ceux réalisés par les Universités de Lille, Rennes, Montpellier et Bordeaux : l'hétérogénéité des corpus n'est pas moins grande à l'échelle de cette communauté de recherche de taille limitée; les matériaux textuels coexistent avec les matériaux iconographiques et sonores, mais les entretiens et observations sont plus fréquemment cités que dans les autres enquêtes, ce qui s'explique par l'ancrage de l'équipe dans l'analyse de l'expérience vécue des espaces urbains. **Plus de 7 répondant.es sur 10 déclarent soutenir les valeurs de l'archive ouverte, mais peu déclarent partager des données au-delà des petits collectifs qui collaborent de façon très étroite, au quotidien.**

De ces différentes enquêtes menées auprès des communautés SHS au sens large, il ressort que la discipline ne semble pas un prédicteur significatif des pratiques, qu'il s'agisse du type de données collectées, des pratiques d'archivage, de la maîtrise des techniques, du rapport à l'ouverture³². A cette échelle, **c'est plutôt la grande diversité des sources et l'absence de convention partagée en matière de gestion des matériaux de recherche qui semble caractériser ces communautés.** Au point que les enquêtes montrent, en creux, que les conventions professionnelles qui les structurent ne semblent pas "datacentrées". Dans ce contexte, le partage des données apparaît comme une pratique artisanale, limitée à des circonstances précises et à des collectifs de petite taille, prenant certes appui sur le développement d'outils numériques devenus foisonnants, mais sans que leurs usages ne s'inscrivent dans une logique de gestion planifiée.

De façon plus diffuse dans les enquêtes apparaît en effet le fait que d'une part, **les valeurs de la science ouverte semblent largement partagées, mais que cette ouverture n'est pas associée par les répondant.es à une exigence spécifique de gestion de leurs données de recherche, ni à une perspective de publication de celles-ci.** En revanche, la méfiance vis-à-vis du mouvement d'ouverture des données, bien qu'elle concerne une minorité des personnes interrogées, se retrouve dans la plupart d'entre elles. Même si les méthodologies retenues ne permettent pas d'analyser de façon approfondie ce point, il nous paraît de ce fait important de relever que la transformation des pratiques de gestion des données ne peuvent être analysées sans prendre en compte d'une part, les changements plus larges affectant les conditions pratiques de réalisation des recherches, et d'autre part, le rapport éventuellement critique que les chercheurs.euses entretiennent avec les institutions qui portent ces incitations à la transformation des pratiques. Dit autrement, les dimensions politiques et épistémologiques de cette question constituent un point aveugle de ces enquêtes.

³² Cette difficulté à discriminer les pratiques entre disciplines est aussi liée aux limites méthodologiques des enquêtes, en particulier leur échantillonnage (voir à ce sujet: Dillaerts et al. 2020 : 3). Certaines disciplines sont par ailleurs très peu représentées parmi les répondant.es.

3.2. Enseignements d'enquêtes qualitatives sur des projets de recherche en SHS ayant intégré un objectif de mise en partage de corpus

Les enquêtes présentées dans la section précédente comportent pour certaines un volet qualitatif, mais il est peu développé au regard des résultats issus de l'exploitation des questionnaires. Qui plus est, ces enquêtes portent sur des échantillons hétérogènes issus de communautés de recherche en SHS déployant des méthodes variées, au sein desquels les sociologues et ethnologues "qualitatifs" sont peu représentés et les expériences de partage de données limitées. Dans cette section, nous nous attacherons à restituer les résultats des retours d'expérience, encore peu nombreux, de communautés de recherche qui se sont impliquées dans des démarches d'ouverture. Ces expériences ont en commun de s'être construites à l'échelle de programmes de recherche spécifiques.

3.2.1. Une typologie des projets

Les projets de recherche impliquant des chercheurs.euses en SHS à partir desquels se sont déployées des réflexions sur la gestion des données qualitatives s'inscrivent dans plusieurs traditions et cadres institutionnels. On peut distinguer 3 ensembles à partir des expériences publiées.

Le premier ensemble est constitué par **des projets "datacentrés" qui s'inscrivent dans la consolidation de pratiques de mutualisation en direction des communautés académiques**. Il est illustré dans ce rapport par le projet de banque d'enquêtes qualitatives BeQuali, porté par le CDSP. Ces projets ont en commun d'être construits à partir d'infrastructures dont la pérennisation est recherchée: la conception de projets financés peut ainsi être décrite comme un effort d'animation scientifique et de mobilisation de ressources en vue d'accroître la robustesse, la visibilité et l'assise de celles-ci.

Le second ensemble rassemble **des projets de recherche dans lesquels la réflexion sur les données, leur partage et leur protection est plutôt une propriété émergente à partir d'un dispositif de recherche impliquant la réutilisation au sein d'une communauté de recherche** - comme dans le cas, développé ci-dessous, du *Méta-projet* mis en oeuvre par une équipe de l'UMR CITERES.

Le troisième ensemble est constitué par **des projets qui sont à la fois "datacentrés" et situés, autour d'un objectif spécifique de construction et d'appropriation de connaissances par un public extra-académique**. Cette configuration est ici illustrée par le projet partenarial SAWA, porté par le LESC.

Les projets relevant de ces trois ensembles ont en commun d'avoir été engagés avant l'évolution du cadre législatif : la façon dont ils ont problématisé la question de la protection, de l'ouverture et de la gestion des données est ainsi particulièrement éclairante, en ce qu'elle témoigne des **formes d'engagement des chercheurs.euses et des équipes d'appui à partir de dynamiques largement endogènes**. Par ailleurs, les collectifs qui s'impliquent dans ces projets récents et parfois encore en cours, ont évidemment été pris dans et confrontés à ces changements de contexte, sur lesquels ils portent, dans le cadre de leurs publications, un regard informé par leur dense pratique.

Il existe très certainement un quatrième ensemble en voie d'émergence, que l'on pourrait caractériser par son niveau d'engagement dans le mouvement d'ouverture des données : il se distingue du premier en ce que les projets qui s'en réclament ne sont pas forcément adossés à des infrastructures numériques pérennes. Le projet *CommonData - Les données de la recherche : un commun scientifique?*, né en 2016 et financé par la MSH Sud, pourrait en constituer un exemple précurseur³³. En lien, principalement, avec l'appel à projets "flash" Science Ouverte porté par l'Agence nationale de la recherche ont émergé, depuis 2019, de nouveaux projets de recherche d'une ampleur en général plus importante, parfois articulés avec ces dynamiques de recherche préexistantes, prenant en compte les évolutions les plus récentes de la gestion

³³ Voir : <https://commondatafr.wordpress.com/>

des données, mais qui n'ont pas encore donné lieu à publication. C'est le cas, pour la recherche en sciences humaines et sociales, de trois d'entre eux:

- le projet SO-DRIIHM (Impulser la science ouverte au sein du Dispositif de Recherche Interdisciplinaire sur les Interactions Hommes-Milieus : co-design d'une e-infrastructure intégrant les principes FAIR)³⁴
- le projet Anthrophen (Les frontières des données anthropologiques)³⁵
- le projet UPMET (Changer d'échelle pour améliorer la qualité des métadonnées afin d'augmenter leur réutilisation par les sciences sociales)

3.2.2. BeQuali (UMS 828 Centre de données socio-politiques) : les enquêtes qualitatives à l'épreuve de la bancarisation

BeQuali est sans doute l'expérience française la plus aboutie de mise en partage de données d'enquêtes qualitatives. La spécificité de l'approche retenue est de **prendre l'enquête comme unité de mise en partage**, ce qui implique un travail considérable à l'échelle de toute une équipe, sur environ un an, pour rendre l'enquête publiable (ce qui signifie ici, la déposer dans une base de données). La publication est donc ici réalisée *ex post*, sur la base du volontariat, par les chercheurs.euses, essentiellement en fin de carrière, dans une démarche d'archivage et de transmission de leurs travaux, et lorsqu'ils sont sortis des circuits de la concurrence et où les enjeux professionnels liés à la protection de leurs travaux sont moins forts (Cadorel et al. 2019). Cette "bancarisation" répond à des objectifs à la fois patrimoniaux et pédagogiques, dans le domaine de l'apprentissage des méthodes de recherche qualitative en sciences sociales. En 2020, 17 enquêtes ont été bancarisées et l'année précédente, 340 étudiants, doctorants et stagiaires les ont utilisées. Le site Internet du projet ne donne pas de chiffres pour ce qui concerne les usages en recherche, mais en livre quelques exemples³⁶.

La bancarisation est réalisée par des documentalistes et archivistes du programme (donc extérieurs à l'enquête), ce qui suppose donc un long travail "d'enquête sur l'enquête", réalisé par l'équipe BeQuali. Cette enquête inclut des entretiens et de nombreux contacts avec le.la chercheur.euse dépositaire pour arbitrer cette recherche constante d'un équilibre entre "protéger les enquêtés" et les "chercheurs" et "donner de la matière à réutiliser" :

"Et pour y arriver on travaille avec le chercheur car l'intérêt de beQuali est que l'enquête soit réutilisable et réutilisée donc si on masque tout, on perd aussi la richesse des données". (S. Cadorel, in: Cadorel et al, 2019: 5)

Le projet est développé depuis 10 ans, avec des moyens assez conséquents sur certaines périodes, et se présente comme précurseur en France sur la mise en partage des données qualitatives et les réflexions sur les analyses secondaires. L'équipe a une politique et des pratiques assez avancées sur leur cadre de diffusion, qui se limite à des porteurs de projets de recherche ou l'enseignement, notamment des méthodologies d'enquête, qui selon leur retour d'expérience est la forme de réutilisation la plus fréquente. Elle vérifie le sérieux des réutilisations et conditionne fortement la mise à disposition des données d'enquête, même anonymisées et nettoyées.

La question de la lourdeur du processus de traitement des données pour permettre leur mise en partage est fortement soulignée par les porteurs du projet : ils insistent sur le caractère lent, méticuleux, voire "casse-tête" des tâches d'anonymisation, mais surtout le tri de ce qui peut et doit être partagé. Une question

³⁴ Voir : <https://anr.fr/fr/lanr-et-la-recherche/engagements-et-valeurs/la-science-ouverte/les-projets-laureats-de-lappel-flash-science-ouverte/projet-so-driihm/>

³⁵ Voir : <https://anr.fr/fr/lanr-et-la-recherche/engagements-et-valeurs/la-science-ouverte/les-projets-laureats-de-lappel-flash-science-ouverte/projet-anthropen/>

³⁶ Voir: <https://bequali.fr/fr/> . Consulté le 15 avril 2021.

essentielle se pose à eux: à quel moment arrêter, jusqu'où aller dans ce travail de fourmi? Par extension, cette réflexion pose le problème de ce qui est partagé, à travers cette notion de recherche d'équilibre citée plus tôt et qui imprègne le discours des cinq membres du projet interrogés par la revue Tracés en 2019. Se pose par exemple la question du statut des notes manuscrites des chercheurs, en marge des entretiens, observations ou autres archives : qu'en faire? les garder ou les effacer? Là encore, chaque décision est longuement discutée - d'ailleurs leur travail d'enquête sur l'enquête est toujours mené en binôme, permettant d'avancer sur ces arbitrages.

“On doit trouver un équilibre entre l’anonymisation et la protection de la richesse des données et à chaque fois c’est à remettre en place, à actualiser en situation. Et je sais que pour cette enquête-là, plein de fois, on basculait d’un côté ou de l’autre. Moi je disais : « Ça, il faut l’anonymiser », puis en discutant dans le binôme ou l’équipe, on se disait : « Si on fait ça c’est absurde : personne ne pourra jamais retrouver de qui on parle parce que c’est vieux mais par contre si on enlève cette information-là, c’est dommage ».” (ibid)

La perte d'information au cours du processus de sauvegarde paraît inéluctable; cependant, qui peut décider *in abstracto* de l'intérêt de telle ou telle information à la réutilisation? Dans beQuali, le dilemme est en partie résolu par le fait que l'équipe a défini un cadre de réutilisation sécurisant (un contrôle sur qui réutilise et pourquoi). Ses pratiques ont toutefois évolué au cours du temps :

“Au départ on était axé sur un type, la réanalyse, donc il y avait une idée très précise de ce que l’on devait faire avec les matériaux. Petit à petit on a élargi les usages et on s’est dit : « Tous ces matériaux, ça peut servir et ce sont les utilisateurs qui verront pour quoi ».” (ibid.)

Le retour d'expérience de 10 ans met en évidence une problématique importante sur la hiérarchisation des données (et de leur valeur), en fonction de ce qui oriente ou motive la stratégie de partage des données : beQuali part d'une logique d'exploitation scientifique de ces données, en regardant l'ensemble de ce qui a été produit, et en essayant de rendre accessible et intelligible ce qui peut l'être. Cette approche déplace quelque peu les finalités recherchées et affichées de la Science Ouverte, qui prône en priorité l'ouverture des données qui sont à l'appui des publications, et donc pas forcément de toutes les données d'enquête. Autrement dit, c'est une autre vision de la science comme bien commun qui appuie la démarche de beQuali, dans laquelle la dimension patrimoniale (la sauvegarde de ce qui risque de disparaître) l'emporte sur la dimension de disponibilité formelle (la mutualisation de ce qui fonde les énoncés scientifiques):

“Je trouve que ça pose une vraie question mais c’est toute la politique ministérielle de la science ouverte qui pose cette question. Plutôt que de savoir comment les données sont produites se pose la question de celles qui vont rester. À beQuali on garde ce qui a été produit en essayant de l’organiser pour que ce soit intelligible mais dans toutes les plateformes en lien avec les sciences ouvertes on va demander de déposer les données en lien avec les publications. Mais qu’est-ce qu’on va faire du reste ? Le plan national pour la science ouverte se focalise beaucoup sur les données qui viennent en appui aux publications alors que dans les archives on va plutôt partir du producteur et voir tout ce qu’il a produit en rapport avec un objet donné. Ce sont deux logiques inversées. Je trouve que ça pose moins la question de comment les choses vont être produites que de ce qui va rester”. (ibidem)

Cette conception inclusive n'empêche pas que se crée une hiérarchie entre les données (ou éléments) publiables et celles qui ne le sont pas, et qui restent dans une sorte de boîte noire : par exemple, on privilégie la publication d'entretiens, qui constituent une certaine unité claire, par rapport à des observations qui sont consignées de manière diffuse et parfois désorganisée, pas forcément uniformément. Cela repose à BeQuali l'éternelle question de ce qu'est - ou pas - une donnée :

“La définition des données est un vaste débat [...] c’est compliqué, c’est trop polysémique. Une note de synthèse est-ce que c’est une donnée, sachant que c’est aussi l’enregistrement

d'une activité de recherche ? Pour nous il y a un noyau dur des données dans une enquête qui sont les observations, les entretiens, les matériaux de terrain écrits et après, ce qui est autour et qu'on a du mal à nommer : des métadonnées, des données périphériques ? On rentre dans des questions de vocabulaire qui sont compliquées. Pour certaines personnes, les données de l'enquête, ce sont les transcriptions d'entretiens ou les carnets de terrain. Nous, on va prendre tous les documents qui attestent de la préparation du terrain et en aval, tout ce qui relève de l'analyse et a mené à des publications ou des valorisations, quelles qu'elles soient, de l'enquête." (ibid.)

La série d'entretiens BeQuali se termine sur une réflexion sur la notion de "donnée", qui montre à quel point la pratique intensive du partage et la recherche de réutilisations constituent plutôt un facteur de complexification des questions d'ouverture et des catégories et concepts qui l'appuient, qu'elles ne les clarifient. L'équipe souligne d'ailleurs qu'elle contourne le problème en évitant soigneusement d'utiliser le terme de donnée, préférant celui de "banque d'enquêtes". La donnée se définit plutôt par la forme d'utilisation ou de réutilisation dont elle fera l'objet: si on fait de l'analyse secondaire, on est dans une définition plus restrictive, que pour des réutilisations en sociologie des sciences ou à des fins pédagogiques, où tout le matériel, y compris les micro-notes en marge, est susceptible de *faire donnée*.

Autrement dit, la pratique de beQuali montre que la qualification d'un matériau en "donnée" est conventionnelle, et que la stabilité des conventions qui lui confèrent ce statut n'est effective qu'au sein de communautés de recherche fortement structurées autour d'une épistémologie commune. Elle témoigne également du caractère peu soluble des matériaux de la recherche qualitative dans un système de bancarisation qui implique un certain degré de standardisation: de ce fait la "routinisation" des pratiques d'ouverture implique une renégociation presque permanente de ces conventions - c'est-à-dire des arbitrages situés, prenant en compte les cadres et projections dans la réutilisation future.

3.2.3. Méta-projet (UMR Cités, TERritoires, Environnement et Sociétés - CITERES) : éprouver les limites de l'analyse secondaire

La perspective portée au sein de "Méta-projet" (voir: Huyghe, Cailly, et Oppenchain 2018a; 2018b) est différente. Porté par le laboratoire CITERES, le Méta-Projet poursuit un double objectif, épistémologique et de recherche finalisée: il s'agit d'élucider les conditions éthiques, juridiques et pratiques rendant possible l'analyse secondaire d'un corpus d'entretiens (n=250) et de suivis GPS (n=100) réalisés par des chercheuses de CITERES à l'occasion d'enquêtes antérieures portant sur les mobilités, datant d'une dizaine d'années. Ce corpus a donc pour caractéristique d'avoir été constitué d'une part, sans intégrer une perspective de réutilisation, d'autre part, sous un régime juridique différent. A l'échelle du laboratoire, l'enjeu est de constituer ce que Huyghes et al. (2018b) nomment un "fond commun", c'est-à-dire capitaliser, agréger et éventuellement faire dialoguer des données anciennes et plus récentes. C'est, aussi, une mutualisation des données au sein d'un collectif de recherche relativement stable qui est ici recherchée : ainsi, les réutilisateurs potentiels intéressés par une analyse secondaire sont aussi des "chercheurs primaires", selon la terminologie retenue dans le Méta-projet, dans le sens où ils ont toutes et tous produit une partie des données.

La réflexion sur le processus de réutilisation a conduit l'équipe à distinguer trois niveaux d'ouverture croissants :

- le partage au sein du programme lui-même
- le partage au sein d'une communauté scientifique élargie
- le libre accès (*open data*)

Les conditions paraissent au collectif du Méta-projet plutôt favorables à une analyse secondaire, puisque les questions de recherche s'inscrivent en continuité avec celles qui prévalaient lors des précédentes enquêtes, et que d'autre part, le collectif de recherche est stable (tout chercheur secondaire est également un chercheur primaire).

Cependant, les données concernées interdisent de fait le partage aux trois niveaux : les entretiens incluent des informations à caractère personnel et le consentement des enquêtés.e.s n'a pas été formellement recueilli à l'époque. L'un des grands intérêts de ce retour d'expérience est de **rendre compte de façon très transparente des péripéties associées à la recherche d'une mise en conformité *ex post* avec le cadre légal qui permettrait de rendre les données réutilisables.**

La première démarche a consisté à rechercher les enquêtés.e.s pour recueillir leur consentement, activité lourde et posant des difficultés pratiques pas toujours surmontables.

La seconde démarche relevait du traitement de la donnée pour la "purger" des informations permettant l'identification des personnes. Ont ainsi été testées différentes pratiques comme le brouillage, la randomisation, l'anonymisation, sans qu'aucune d'entre elles n'évite une perte significative d'informations limitant l'intérêt d'une remobilisation de ce corpus dans le cadre d'une analyse secondaire.

Dans ce contexte, à chaque niveau d'ouverture identifié correspondent des enjeux spécifiques appelant des réponses pratiques différentes. Le partage au sein de l'équipe projet (niveau 1) laisse la place à des pratiques informelles reposant sur une relation de confiance. A ce niveau, c'est plutôt la plus-value d'une analyse secondaire que le retour d'expérience sur le Méta-projet questionne (Huyghe, Cailly, et Oppenheim 2018b). Dès le niveau 2 (partage au sein d'une communauté scientifique élargie), la formalisation d'engagements garantissant le respect de la confidentialité se révèle nécessaire, dans la mesure où l'autorégulation des usages au sein d'un collectif pérenne n'est plus assurée. Ainsi, les chercheurs.euses auteur.e.s des données réutilisées conservent un droit de regard sur la réutilisation: ils.elles doivent être informé.e.s de celle-ci, cité.e.s, voire être considéré.e.s comme coauteur.e.s d'une éventuelle publication. Comme dans le cadre de beQuali, l'équipe du Méta-projet a donc élaboré une charte de réutilisation. Enfin, le niveau 3 s'avère, dans la pratique, inatteignable dans le cadre du Méta-projet : l'impossibilité de réguler la réutilisation, qui constitue pour l'équipe du projet une forme d'impératif catégorique lié à la protection des enquêtés et des collecteurs originels, les conduit *in fine* à renoncer à une ouverture libre des données.

3.2.4. SAWA (Laboratoire d'Ethnologie et de Sociologie Comparative) : de la restitution comme impératif moral à l'accompagnement de l'appropriation collective des données

Le projet SAWA (Savoirs autochtones wayana-apalaï (Guyane) – Une nouvelle approche de la restitution et ses implications sur les formes de transmission) s'inscrit dans une perspective à la fois classique en ethnologie de constitution, d'animation et de documentation de collections d'artefacts, d'archives orales et iconographiques, et plus novatrice, en s'efforçant de mobiliser les outils numériques au service d'une recherche collaborative avec les experts autochtones et d'une réflexion sur les pratiques de restitution, que sa coordinatrice a synthétisées dans le cadre de l'introduction qu'elle a donnée à un numéro spécial du Journal de la Société des Américanistes intitulé *Restitution numérique, réappropriations amérindiennes* (Vapnarsky 2020).

Le projet SAWA se situe ainsi au cœur des enjeux politiques et épistémologiques contemporains de la recherche en anthropologie, ainsi que des articulations spécifiques que la recherche dans ce champ implique avec la numérisation, la bancarisation et le partage des données (Vapnarsky 2019b). Il a aussi pour spécificité de faire figure de "modèle" en France dans le champ des études patrimoniales, en raison de l'effort de co-construction et de réduction des asymétries entre d'une part, les membres de communautés Wayana et Apalaï impliqués, et d'autre part les chercheurs.euses et ingénieur.e.s du CNRS, du Musée du Quai Branly-Jacques Chirac et du Musée des Cultures Guyanaises (Vapnarsky 2019a). Ils sont ainsi largement intervenus dans la conception du portail WATAU³⁷, musée virtuel dont ils ont contribué à choisir les contenus, dans les formes de restitutions associées, ainsi que dans les choix linguistiques, interagissant avec un large spectre

³⁷ www.watau.fr

de métiers et d'expertises (Camargo et al. 2020). Les retours d'expérience qui en sont issus sont donc particulièrement précieux à analyser dans le cadre du présent rapport.

La première dimension qu'explore ce retour d'expérience est l'importance des outils numériques dans l'intensification des revendications et des initiatives de retour des données concernant les savoirs traditionnels et, plus largement, des connaissances dont les communautés autochtones ont rendu l'émergence possible. Ce processus est parfois ambivalent, dans la mesure où la numérisation peut jouer un rôle de substitution à une restitution matérielle des objets. Il n'en reste pas moins que ces projets restent en général animés par ce que V. Vapnarsky (2020) identifie comme des "considérations éthiques postcoloniales" ou une volonté de "compensation symbolique", contribuant à l'*empowerment* des "communautés sources" des données en question.

Les formes d'auctorialité et de propriété intellectuelle collective associées à ces matériaux sont de plus en plus reconnues par le droit international, ce qui constitue un défi aux procédures d'archivage et les met à l'épreuve de l'interculturalité :

"Acteurs et observateurs peuvent en perdre le sens de l'amont et de l'aval. L'ethnologue en quête de fonds, l'archiviste en quête de droits, le documentaliste en quête d'auteurs, l'enregistré en quête de passeur... Chacun, finalement, cherche sa source, et les retours sont plutôt des allers-retours. Par ailleurs, notre fascination pour le temps long et la préservation de ces vestiges est loin d'être universellement partagée (Erikson et Vapnarsky sous presse). Ce qui est archive et conservation, ici, peut ailleurs avoir une valeur d'usage limitée et changeante dans le temps, en devenant monnaie d'échange, objet de curiosité, sujet de conflit, étendard politique, munition, secret à cacher, matière à transformer, support rituel de secondes funérailles... ou tout simplement gadget insignifiant. Ce qui était pensé comme trace de savoir se révèle tantôt inerte, tantôt principe hyperactif dans la création aussi bien de relations sociales que de nouvelles formes esthétiques, épistémiques et communicationnelles." (Vapnarsky 2020: 4)

Dans ce contexte, un projet comme SAWA, très fortement soutenu institutionnellement et financièrement, met à jour les nombreuses opportunités et les tensions spécifiques que crée la publication des données au format numérique. Parmi celles-ci, l'instabilité des formats numériques par rapport aux formats analogiques, pourtant a priori plus vulnérables. Cette "obsolescence constitutive" du numérique implique "des infrastructures qui puissent continûment financer des processus autant de conservation que de formation aux – toujours – nouvelles technologies." (Vapnarsky 2020 : 6). En ce sens, il ne permet le partage et ne facilite l'accès aux données pour les populations autochtones que de façon conditionnelle, si des moyens d'appui technique, de traduction et d'animation quasi-permanents sont mobilisés, ce qui est rarement le cas. En ce sens, le dédoublement du numérique par l'analogique et les échanges directs restent difficilement dépassables.

La restitution peut ainsi être considérée, dans les exemples qui servent d'appui à V. Vapnarsky dans son analyse, comme un transport de matériau dans lequel le statut de celui-ci se trouve en permanence transformé par sa reproduction, sa mise en série, sa mise en circulation. Ce, **sans que les conditions d'énonciation de celui-ci, qui jouent dans le cadre de la recherche qualitative en sciences sociales et en particulier de l'ethnographie le rôle de métadonnées incontournables permettant d'associer à un discours, un son, une séquence, un sens spécifique, lui restent attachés.** La réaction de certains groupes autochtones face à des matériaux numérisés (chants, mythes) qui leur sont restitués et qu'ils ne reconnaissent pas, illustre à quel point ces transformations opèrent.

Pour autant, la complexité, le coût et les limites de ces dispositifs n'invalident pas leur intérêt, ce dont témoigne les formes multiples, y compris politiques, par lesquelles les communautés recontextualisent, se ressaisissent des matériaux en question et leur attribuent une valeur en situation. Pour les ethnologues, l'une des difficultés les plus soulignées concerne le risque de standardisation et de "congélation" des matériaux, les versions enregistrées et restituées pouvant devenir des références au sein des communautés d'origine comme des communautés académiques, statut obtenu de façon quelque peu hasardeuse car

dépendante des aléas de la collecte et de la numérisation. Mais cette difficulté est davantage amplifiée qu'elle n'apparaît avec la configuration qui prévaut dans SAWA, et pose de façon particulièrement aiguë des questions d'une portée très générique :

“Quelle que soit l'explication retenue, le dilemme que pose Cohn³⁸ n'en demeure pas moins capital : faut-il ou non tout restituer ? Quels aspects du passé doivent-ils être récupérés et lesquels doivent-ils être oubliés ou effacés ? Et qui doit en décider ? L'enregistrement qu'un grand-père défunt aurait été horrifié qu'une femme puisse entendre, peut-il être aujourd'hui remis à sa petite-fille, devenue dans l'entremise une jeune intellectuelle féministe ? La possibilité de créer des copies ad infinitum intensifie les effets des choix, alors que les injonctions éthiques et juridiques complexifient les décisions.” (Vapnarsky, 2020: 12)



Figure 6. Lancement du portail WATAU dans les communautés (© Sara Tandar, 2020 ; source: Camargo et al., 2020)

³⁸ Voir : Cohn 2020.

4. Apports d'une approche ethnographique des pratiques des communautés de recherche qualitative en sciences sociales

L'approche ethnographique enrichit et affine notre compréhension des pratiques, enjeux, expériences vécues et problèmes rencontrés par les communautés praticiennes de la recherche qualitative.

Nous proposons dans cette quatrième partie d'aborder les résultats de deux enquêtes complémentaires qui documentent des expériences situées, des pratiques, des "frictions" (Jaton et Vinck 2016) et des épreuves traversées par les praticien.ne.s de la recherche ethnographique qui se confrontent à un objectif de partage ou d'ouverture de leurs matériaux ethnographiques.

La première, l'**enquête "Parchemins"** (présentée en 4.1), constitue le retour d'expérience du collectif interdisciplinaire de ce programme de recherche (*Paroles et chemins de l'agriculture littorale*, 2016-2021), alliant des chercheur.euses et ingénieur.e.s en agronomie, en sciences et technologies de l'information (STI), en ergonomie, en anthropologie et en sociologie, et poursuivant un objectif énoncé dès le lancement du programme de mise en partage de données qualitatives inédites sur les relations entre agriculture et littoral en Bretagne. Deux axes de travail de ce programme ont particulièrement été consacrés à cet objectif:

- l'axe *thématique* 2 : "Produire, valoriser et **publiciser** des données qualitatives inédites" ;
- et l'axe *transversal* B : "Ethique de la recherche collaborative et de la **publication des données qualitatives** sensibles en sciences sociales".

Cette enquête s'appuie donc sur l'ethnographie d'un processus d'ouverture institué au sein d'un programme de recherche, et observé de l'intérieur grâce à une démarche proche de l'auto-ethnographie. Les résultats de l'enquête Parchemins ont fait l'objet de plusieurs communications (voir notamment: Revelin et al. 2019), ainsi que d'une publication récente (Squidant et Revelin 2021), mais le présent rapport en propose une analyse largement inédite, centrée non sur le processus d'ouverture de la recherche dans son ensemble, mais sur celui des **pratiques des ethnologues** et des **tensions qui traversent le processus de publication des données ethnographiques**. L'intérêt d'un regard approfondi sur cette configuration de recherche est double : au-delà du fait qu'il s'agisse de l'une des rares expériences de construction d'une infrastructure de publication de données incluant des matériaux ethnographiques sans s'y limiter, les travaux existants sur le sujet insistent en effet sur le rôle important des infrastructures de publication de données au sein des programmes de recherche interdisciplinaires dans la diffusion de cultures de la donnée au sein des SHS, ainsi que dans l'hybridation avec des normes épistémiques issues des sciences biophysiques (voir: Millerand 2011).

La seconde enquête présentée dans cette partie - dite "**l'enquête PARDOQ**" (présentée en 4.2 et 4.3) - fait ici référence au second volet³⁹ de cette enquête centré sur une série de 13 entretiens semi-dirigés menés auprès de **chercheur.euses** dont les travaux reposent partiellement ou entièrement sur des matériaux ethnographiques. Ce panel diversifié d'interlocuteurs.trices comprend plusieurs générations de chercheur.euses, ayant différents statuts et travaillant dans divers environnements et configurations de recherche, détaillés dans la partie 4.2 qui décrit le dispositif d'enquête et la diversité des profils enquêtés.

³⁹ Le premier volet, concernant les entretiens menés auprès des réseaux d'appui est présenté en partie 2.2. de ce rapport

4.1. Pistes ouvertes par l'enquête Parchemins

Les modalités de réalisation de l'enquête "Parchemins" sur les pratiques d'ouverture de la recherche et de partages des données la place à la frontière d'une **auto-ethnographie collective**. Elle est en effet principalement portée par une membre (ingénieure de recherche en anthropologie) de l'équipe du programme, également membre du sous-groupe de travail concepteur de l'instrumentation technique de l'ouverture des données Parchemins (Squivalent et Revelin 2021). L'enquête allie ainsi une démarche de participation observante, la collecte de traces du processus vécu, et 40 entretiens réalisés en 2019 auprès de l'ensemble des membres et partenaires du programme dans le cadre d'un stage de Master portant sur la sociologie des activités d'intermédiation (Beurier 2019). Elle s'appuie également sur une **démarche réflexive concertée et explicitée** dès le début du projet de recherche, qui a permis à l'ensemble des membres de l'équipe scientifique de prêter attention de façon consciente et de témoigner, tout au long de celui-ci, de la façon dont ils s'approprièrent, faisaient avec, transformaient éventuellement leurs pratiques, au contact d'une démarche d'ouverture des données de recherche - et du sens qu'ils donnaient à ce processus.

L'auto-ethnographie de l'expérience de mise en partage et d'ouverture des données au sein du programme Parchemins a fait émerger différents constats, qui contribuent à une réflexion plus profonde sur l'enjeu d'ouverture des matériaux de recherche ethnographiques.

4.1.1. La déconstruction de l'évidence du "cycle de la donnée"

Les supports de présentation du processus d'ouverture des données de la recherche mobilisent de façon presque systématique l'image du "cycle" de vie de la donnée, en général présenté comme un "modèle" conceptuel de référence pour la gestion des données et le processus de publicisation (FIGURE 7): la récursivité se produit au moment de la phase de "réutilisation", qui rompt la linéarité du processus d'ouverture et lui donne son sens, de façon littérale comme figurée. La donnée "renaît" par la réutilisation qui peut en être faite par d'autres. La base de ce cycle est avant tout technique, et tout au long de cycle transparaît toujours sa traduction numérique à des fins d'ouverture, afin d'en faciliter la mise en œuvre et *in fine* assurer l'accès et la réutilisation.

Derrière la présentation en "cycle", c'est pour le ou la "producteur.trice" des données un processus linéaire de travail qui est donc en fait décrit, car d'une part la réutilisation engage d'autres praticien.ne.s, d'autre part ce ne sont pas les tâches concrètes qui sont représentées.

- ☆ Objectifs
- ① Introduction
- > Contexte et enjeux
- > Les données de la recherche
- ▼ Bonnes pratiques de gestion
 - 📄 Le cycle de vie des données
 - 📄 Le plan de gestion des données
 - 📄 Des bonnes pratiques
 - > Testez vos connaissances
- > Décrire les données
- > Publier et valoriser les données
- 🕒 Conclusion

Le cycle de vie des données



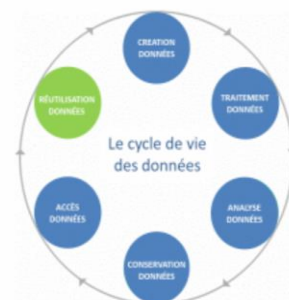
Le cycle de vie des données de recherche (Research data lifecycle) décrit le processus d'utilisation des données de leur création à la publication et à leur réutilisation ultérieure.

Il existe plusieurs représentations du cycle de vie des données de la recherche.

Le centre d'archives de données de la recherche, UK Data Archive, fournit un modèle de cycle de vie des données sur lequel les chercheurs peuvent s'appuyer pour gérer leurs données.

Ce modèle définit les six étapes suivantes :

- **Création ou collecte** des données (creating data) ;
- **Traitement** des données (processing data) ;
- **Analyse** des données (analysing data) ;
- **Conservation** des données (preserving data) ;
- **Accès** aux données (giving access to data / data discovery) ;
- **Réutilisation** des données (reusing data).



Cycle de vie des données de la recherche (adapté de UK Data Archive)



Figure 7. Extrait d'un support pédagogique publié par l'INIST sur son site Internet⁴⁰

Cette référence au cycle de vie de la donnée est mobilisée par les ingénier.e.s de l'équipe Parchemins au démarrage de celui-ci, qui correspond à un moment de co-conception d'un système d'information *ad hoc*, entre l'équipe d'ingénierie et les chercheurs.euses du projet, en particulier l'équipe de 3 ethnologues en charge de réaliser les enquêtes ethnographiques sur les 5 sites d'étude du programme.

L'objectif de partage interne des données pour produire de la connaissance interdisciplinaire et de publicisation permettant la réutilisation par une diversité d'utilisateurs étant partagé et consensuel au sein de l'équipe, les ethnologues du projet Parchemins s'engagent de façon confiante et résolue dans ce processus et ne sont pas *a priori* critiques de cette forme de conceptualisation. Elles adoptent le vocabulaire qui l'accompagne, archivent avec soin leurs données sur des espaces sécurisés, avec une nomenclature précise, définie collectivement.

Cependant, émergent rapidement et de façon dispersée des questions chaque fois plus épineuses : qu'est-ce qui, dans mon matériau, "fait donnée" ? Puis-je et dois-je tout partager ? Comment savoir ce qui sera utile à d'autres, ce qui "mérite" d'être publié ? Comme l'élaboration du système d'information permettant la publication se déroule simultanément avec la collecte de terrain elle-même, **les enquêtrices indiquent par leurs doutes et leurs questions qu'elles ne collectent pas leurs matériaux de la même façon que si la perspective de leur publication était absente**. S'invite dans la phase de collecte non plus seulement la question de la pertinence du matériau recueilli, mais également celle de son caractère publiable ou partageable:

« Pour moi en fait, rendre public, c'est à tous les moments de la recherche une préoccupation qui transforme les pratiques. Chaque moment, chaque volet on va dire plutôt du travail de recherche est transformé par ce processus de publicisation. C'est : en quoi le fait de rendre public fait que je vais faire comme si ou comme ça à tel moment? Et

⁴⁰ Source : https://www.inist.fr/wp-content/uploads/donnees/co/module_Donnees_recherche_22.html . Consulté le 1er avril 2021.

cette question, tout le temps, elle change tout. Elle change la façon dont tu collectes, elle change ce que tu collectes, elle change ton rapport à tes enquêtes, elle change la façon dont tu vas recenser tes données, les travailler, les documenter, elle change les formes que peuvent prendre la restitution, elle change tout quoi. » (Ethnologue du programme Parchemins, 2019)

Ce qui est vrai de la “phase” de collecte est ainsi, pour cette ethnologue, également vrai des “phases” supposées ultérieures de traitement, d’analyse et d’archivage. Ce constat s’impose rapidement au sein de l’équipe scientifique: **la réutilisation ne peut être considérée uniquement comme la phase ultime d’un processus par un.e chercheur.euse qui travaille sur du matériau qualitatif, mais comme une perspective qui modifie structurellement la nature des corpus qui servent de base à la construction de connaissances**, dès lors qu’il est considéré que l’accès doit être la règle et la protection l’exception.

4.1.2. L’explicitation de l’indétachabilité du triptyque : “contexte-donnée-observateur” et la question de la “violence” faite aux données

Le contexte interdisciplinaire du projet rendait à un certain degré indispensable la construction d’un langage commun entre chercheurs.euses pratiquant la recherche qualitative en sciences sociales et chercheurs.euses en sciences biophysiques. Mais l’objectif de mise en partage des données au travers d’un dispositif *ad hoc* a renforcé cette nécessité d’explicitation de l’épistémologie sous-jacente aux données ethnographiques, dans la mesure où la conception-même d’une chaîne de publication adaptée impliquait un degré élevé de familiarité entre l’ingénieur développant les outils et les productrices de “données”.

L’explicitation des formes d’existence de la “donnée” ethnographique

Au début du programme Parchemins, la géomatique et les langages de programmation informatique des bases de données agronomiques et environnementales semblaient aux uns, tout aussi étrangers que la démarche ethnographique et la collecte de données qualitatives aux autres. Si l’effort est à double - voire multiple - sens, nous nous arrêterons plus particulièrement ici sur les questions que pose l’explicitation de la démarche ethnographique en tant que productrice de « données »⁴¹ de recherche.

Ce qui semblait évident aux ethnologues du programme l’était bien moins pour leurs collaborateurs.trices agronomes, géomaticien.ne.s, chercheurs.euses en sciences de l’information, ou gestionnaires de base de données : mais sur quoi s’appuient donc les recherches des ethnologues ? Quelles sont leurs « données » ? Les premiers échanges oraux ont permis d’éclaircir la question, d’amorcer une intercompréhension et de comprendre que le « terrain » et l’enquête tiennent une place de premier plan dans la démarche scientifique des ethnologues.

Mais concrètement, à quoi ressemble une « donnée » issue d’une enquête ethnographique ? Telle était la question que posaient les interlocuteurs et interlocutrices des ethnologues au sein de l’équipe du projet Parchemins, toutes et tous habitués à côtoyer ou gérer de grands corpus de données spatialisées de formes et formats divers, et dont la standardisation numérique confère un caractère explicite qui leur paraissait facile à capturer ou à se représenter.

L’effort de traduction, ici centré sur la question des matériaux sur lesquels s’appuie la recherche, s’est déployé tout au long du programme et il n’est sans doute jamais pleinement achevé. Dès le départ, il est passé par un souci d’explicitation, qui s’est imposé comme une condition *sine qua non* du travail collaboratif. Outre comprendre sur quoi les ethnologues travaillaient pour co-construire et instruire collectivement les questions de recherche du programme, il fallait aussi définir précisément ce qui pourrait et allait être partagé, à quoi cela ressemblerait et comment l’équipe scientifique allait pouvoir l’utiliser collectivement.

⁴¹ Le terme « données » figure ici entre guillemets car il s’agit dans ce contexte précis d’une catégorie émiqque propre à certains chercheurs, mais par forcément partagée par tous. C’est un point important de ce qui est discuté dans ce rapport.

L'hétérogénéité à tiroirs des données ethnographiques

C'est dans ce contexte qu'une ethnologue du projet a posé une première pierre au commencement de celui-ci : un document décrivant dans des termes simples et explicites ce que sont les données ethnographiques (FIGURE 8). D'abord d'usage interne, ce document s'est transformé en billet sur le site internet de Parchemins, se voulant lui-même transparent et pédagogique quant aux activités et travaux menés par l'équipe scientifique pour les rendre accessibles à toute personne s'intéressant à l'agriculture en zone littorale et en particulier à ses interlocuteurs sur le terrain.

L'auteure de ce document pose d'emblée les spécificités de ces « données » : elles sont hétérogènes ; elles passent par de multiples filtres que sont les sens de l'ethnologue, avant d'être consignées dans des formats divers à travers lesquels ce dernier restitue sa collecte de « morceaux de réel » :



Qu'est ce qu'une donnée ethnographique?

10 mars 2017 | Ecrit par parchemins

Les données ethnographiques : des données hétérogènes

L'ethnographie est une **méthode de recherche descriptive et analytique « intensive »** qui privilégie l'immersion, l'observation, la description fine du terrain et la production de données essentiellement qualitatives. Cette méthode ne constitue pas l'unique mode de production des données en ethnologie mais demeure essentielle et dominante dans la discipline.

Ces données sont produites par l'ethnologue. Pour rendre compte de ce qu'il voit, entend, sent, goûte, touche dans le cadre d'une recherche, **il produit des données très hétérogènes** qui lui permettent de rendre analysables les « **morceaux de réel** » qu'il choisit de collecter (Olivier de Sardan, 1995).

Figure 8. Extrait de la brève "Qu'est-ce qu'une donnée ethnographique" sur le site agricultuelittorale.fr, consultée le 12/10/2020.

Dans cette brève, l'auteure s'attache par ailleurs à rendre le plus tangible possible ce type de données. Elle les présente à travers leur manifestation matérielle, c'est-à-dire leurs formats, qui sont de trois types : sonores, textuels et iconographiques, avec des hybridations possibles. Son effort d'explicitation tend ainsi à présenter les formes les plus élémentaires que ce matériau peut prendre. Ce qui n'est pas chose aisée car ce matériau est par nature complexe : sa spécificité est à la fois d'incorporer les « filtres » de la perception du chercheur (Mohia 2008), et d'être toujours entremêlé à un contexte complexe et épais qui rend particulièrement difficile, voire impossible sa réduction dans des formes élémentaires qui pourraient être indépendantes à la fois de ce contexte et les unes des autres.

L'auteure évoque par exemple le cas des entretiens, que l'on peut considérer comme une « donnée » au sens où ils forment une unité : le texte retranscrit d'un échange avec un ou plusieurs interlocuteurs. Elle les présente comme une « source centrale d'une enquête ethnographique », qui va bien au-delà du simple discours retranscrit de la personne, et est de surcroît enrichi de métadonnées. Un matériau qui n'est pas seulement recueilli (enregistré) mais aussi traité (*i.e.* réécouté, retranscrit) et élaboré (*i.e.* documenté de

métadonnées sur le contexte de collecte et le contenu) pour constituer une « matière première de l'ethnographie » dont on constate qu'elle est loin de pouvoir être qualifiée de « brute ».

Le poids des arbitrages

La culture interdisciplinaire des ethnologues impliquées dans Parchemins (elles sont toutes anthropologues de l'environnement, habituées à la collaboration avec les sciences “dures” et parfois issues d'une double formation en sciences sociales et sciences biophysiques) les conduit à ne pas mettre systématiquement en avant les singularités des données produites dans leur discipline pour expliquer le fait que les pratiques de partage y sont peu répandues:

« On peut faire un parallèle intéressant avec les données naturalistes. Qui sorties de leur contexte n'ont parfois aucun sens. Enfin, peuvent amener à des contresens. Elles ont besoin, ouais bah... le parallèle il est fort en fait. En fait c'est très intéressant parce que euh, c'est un peu pareil. » (Ethnologue du programme Parchemins, 2019)

« Les données ethnographiques, c'est des données qui en fait d'habitude ne se partagent pas du tout, qui sont propres à un enquêteur. Il arrive que ce soit partagé, mais c'est quand même vraiment rare. » (Ethnologue du programme Parchemins, 2019)

Elles attribuent plutôt ce phénomène à un *ethos* des ethnologues, pour des raisons qui se rapportent, pour le dire rapidement, au processus de décentrement caractéristique de la démarche ethnographique et à l'habitude de vivre et d'accepter la solitude qui en résulte (voir par exemple à ce sujet: Augé 2013; Gaillard 1993)

Malgré l'énergie et l'enthousiasme qu'elles déploient tout au long du projet pour ouvrir la “fabrique” de l'ethnographie à l'ensemble du collectif de recherche, le processus d'ouverture des données n'aura pas en lui-même contribué à atténuer cette solitude : **un dispositif collectif de publication des données, même issu d'un processus de co-construction attentif à l'épistémologie et aux contraintes pratiques des “collecteurs”, est plutôt générateur d'une problématisation accentuée des pratiques que d'une routinisation de celles-ci:**

« Tu vois, on est obligé de faire des choses, on est obligé de trancher en permanence, c'est des micro-choix, tu vois. » (Ethnologue du programme Parchemins, 2019)

De fait, l'existence d'un temps et d'une expertise d'ingénierie dédiés, très certainement beaucoup plus développés dans le cadre du projet Parchemins que dans la moyenne des projets de recherche, ne conduit pas à une mutualisation ni à un allègement de la charge pour les ethnologues qui alimentent le catalogue de données publiées. La citation précédente montre que **le processus d'ouverture, au-delà du traitement spécifique du matériau qu'il demande, est générateur d'arbitrages permanents pour le ou la chercheur.euse, qui ne peuvent être différés dans le temps ni réalisés “en situation”, en connaissance de cause, c'est-à-dire au moment de l'utilisation des données pour une publication, un partage ciblé etc.**

Face à ces “micro-choix” permanents associés d'une part à une forte incertitude sur les réutilisations futures, et d'autre part au sentiment de perte de contrôle et au risque de perte d'intégrité de la “donnée”, deux réflexes contradictoires en apparence ont été observés et identifiés par les ethnologues du programme lors des ateliers réflexifs réguliers auxquelles elles participaient. Le premier est **l'encyclopédisme** : pour maintenir ouvertes les potentialités attachées à des matériaux de recherche dont les usages futurs possibles, par soi-même comme par d'autres, ne sont pas connus, et pour éviter une discrimination entre eux perçue comme arbitraire, tous les matériaux collectés font l'objet d'une numérisation et d'un formatage. Le second est **l'hyper-protection** : les données sont publiées au sens formel du terme, c'est-à-dire qu'elles sont archivées et documentées en respectant les standards *ad hoc*, mais associées à des licences qui en empêchent pratiquement la réutilisation, voire l'identification, par des tiers. Elles sont, selon la terminologie employée au sein du programme, “publiées sous statut privé”, ce qui permet éventuellement d'assouplir le processus d'anonymisation et d'éviter des pertes d'informations qui altèreraient par trop, pour les

ethnologues, la qualité des données, voire obèreraient les possibilités d'une réutilisation⁴². Cette façon de faire peut ainsi être, de façon un peu brutale, résumée par l'idée d'ouvrir une route dans des conditions difficiles, pour en interdire l'accès dès qu'elle est achevée : l'effort réalisé permet de signaler que la route existe, de la faire exister sur la carte (les métadonnées sont publiques), mais pas de l'emprunter (les données ne sont pas accessibles).

Ces deux tendances peuvent être analysées comme **deux formes de résistance au processus de réduction** du matériau de recherche qu'implique sa mise au format, et qui s'oppose à la perspective holiste que défendent les ethnologues du projet. Il s'agit, autrement dit, **d'éviter de se trouver face à un choix entre partage avec perte de qualité de la donnée, et invisibilisation des matériaux.**

4.1.3. La dépendance aux instruments et ses effets

Plus le processus avançait et le système d'information servant de support à la publication se précisait, plus l'équipe d'ethnologues s'est trouvée confrontée au caractère chronophage du double processus de standardisation et de numérisation qui accompagne la démarche d'ouverture. De leur côté, les deux ingénieurs concepteurs de la chaîne de publication tentaient de concilier, avec peine, la prise en compte de la nature des matériaux collectés et du confort des enquêtrices d'une part, les fonctionnalités d'outils configurés au départ pour héberger et documenter des jeux de données plus stables, moins hétérogènes d'autre part. Il semblait s'agir ainsi en permanence de faire rentrer "des ronds dans des carrés" au "chaussepied", pour reprendre des expressions entendues dans l'équipe: malgré les efforts considérables de co-conception et d'attention à l'utilisateur réalisés par les ingénieurs, des fonctionnalités qui se révélaient progressivement comme centrales pour les enquêtrices n'étaient pas prises en charge par les outils disponibles et nécessitaient des développements informatiques *ad hoc*, au risque de conduire les utilisatrices à revenir sur les données déjà saisies (parfois en grand nombre).

« Il y a eu plein de versions de comment on nomme les fichiers par exemple, donc j'ai dû renommer mes fichiers plein de fois, enfin... Il y avait plein de données et des trucs ultra bateau, enfin tu collectes de la littérature grise, ça veut dire qu'il faut tout scanner et du coup, plus tu dois en collecter puis tu dois traîner et du coup tu ne t'en sors pas... ça demandait une rigueur dans l'organisation que je n'ai pas. Du coup, ça m'obligeait à me faire violence sur pas mal de trucs ». (Stagiaire, Parchemins, 2019)

Le travail austère, systématique et pointilleux de saisie des données et des fiches de métadonnées, de vérification des corpus, d'anonymisation quand il le faut, semble ainsi se substituer à un rapport au matériau qui apparaît, *ex post* et par contraste, plus incarné, plus habité, plus porteur de sens.

"C'est des infrastructures qui sont quand même extrêmement lourdes, mais en même temps peut-être qu'il faut en passer par là. Mais en même temps c'est paradoxal, il y a ça dans Parchemins, il y a des infrastructures très lourdes qui sont censées porter beaucoup de légèreté et donc je ne sais pas... Je trouve que... Je ne sais pas comment... je n'ai pas de conclusion à tirer de ça, c'est juste un regard que j'ai sur la manière dont ce projet fonctionne". (Ethnologue, Parchemins, 2019)

L'un des effets de ce processus a été le découragement, temporaire ou plus durable, face au poids qu'occupe dans l'activité quotidienne une instrumentation technique instable et difficile à apprivoiser, et dont les exigences finissent, malgré tout, par s'imposer et par modifier très substantiellement le rapport aux matériaux de recherche que l'on peut construire sur un temps contraint.

⁴² Millerand (2011) fournit un exemple concret de cas où des données ethnographiques versées à un catalogue de données interdisciplinaire sur les zones polaires (Polar Data Catalogue) ont été tellement "épurées", en particulier par la suppression de toute forme d'indication géographique, que dans la pratique toute analyse croisée ou comparée ultérieure est devenue impossible.

Une autre dimension très frappante pour les ethnologues impliqués dans le projet Parchemins a été la façon dont la chaîne technique de publication, associée aux impératifs de protection des données personnelles et à la mise en œuvre du RGPD, conduisait à modifier la façon dont les enquêtes se donnaient à lire à l'extérieur. En effet, les matériaux considérés comme les plus essentiels et caractéristiques du travail ethnographique, à savoir les entretiens informels (donc sans déclaration préalable, sans recueil écrit de consentement), les observations et notes de terrain qui peuvent en résulter, ne pouvaient être rendus publics, alors que d'autres sources jugées en général plus secondaires, comme les archives et documents écrits, pouvaient être rendues publiques plus aisément. Si les enquêtrices exprimaient une sensibilité très forte à la nécessité de protéger la vie privée et l'anonymat de leurs informateurs et informatrices, il n'en restait pas moins préoccupant pour elles de constater que **le processus de publication invisibilisait dans la pratique l'essentiel de leur matériau, et donc la spécificité de leur travail et de leur expertise**. De fait, la conjonction des deux impératifs d'ouverture et de protection créent un effet de distorsion dont les ethnologues ont pu faire brutalement l'expérience dans le projet Parchemins, mais qui dépasse largement le cadre du projet et qui confère une dimension à la fois épistémologique et politique à l'investissement individuel dans ce type de démarche - pouvant donner lieu à des positionnements de principe d'adhésion ou de rejet. Cette charge est très perceptible dans la façon dont les enquêtrices relatent leur expérience dans le projet et l'associent à des questionnements plus globaux sur la pratique ethnographique et sa reconnaissance comme démarche scientifique :

« Il y avait aussi un intérêt pour quelque part peut-être redonner ou conserver une place à l'ethnographie dans un monde en train de se numériser quoi si on veut. Parce que ce sont des données qui ne sont pas solubles dans des bases de données justement. Des connaissances ou des données ou des informations... Et donc à partir de là, elles sont impartageables et donc elles n'existent pas. Et donc le métier qu'on fait n'existe pas non plus. Et donc nous allons disparaître! (rires) » (Ethnologue Parchemins, 2019)

“Et moi c'est par rapport à ça que je me mets en décalage dans ma pratique aujourd'hui quoi, je ne sais pas comment je vais le faire, mais, ce refroidissement. Mais donc en ce qui me concerne ça remet même en question mon, mon... comment dire ? Mon implication dans une démarche qui peut être qualifiée de scientifique. Et je l'assume complètement. » (Ethnologue Parchemins, 2019)

Cette seconde citation met particulièrement en exergue **l'importance de distinguer refus de l'ouverture et résistance à la standardisation** pour analyser le rapport des ethnologues au processus de publication de leurs matériaux de recherche:

“C'est un effet d'assèchement, d'assèchement de la pulsation de vie qu'il y a dans des données ethnographiques et c'est pour ça qu'on aime faire ce métier aussi quoi, c'est que c'est vivant. Et là, forcément, on va faire rentrer ça dans une base de données. Ça demande une aridification qui est encore plus forte que celle que tu fais de toute façon vivre à tes données quand tu les analyses toi-même, tout seul. Je ne suis pas pour être tout seul dans son coin, je constate que la mise en partage, la mise en partage du coup, elle passe par un assèchement très fort et par du coup une approche de la parole et de toutes ces informations qui ont circulé au départ et qui les amène vers quelque chose qui est extrêmement froid, quoi.» (Ethnologue Parchemins)

Ainsi, le poids de la “technique”, des “outils” peut apparaître comme contradictoire avec les finalités-mêmes qui justifient, dans le cadre du projet, leur existence : œuvrent-ils vraiment à rendre *accessibles* au plus grand nombre les données de la recherche s'ils se substituent à des interactions humaines qui pourraient remplir une fonction analogue, et qui pourraient apparaître comme un déploiement plus fructueux, plus nourrissant du savoir-faire de l'ethnologue et de la qualité des connaissances qu'il ou elle co-produit avec ses enquêtés?

La mise en évidence d'asymétries de position vis-à-vis du processus d'ouverture

Le parcours biographique et le type de relation qu'elles entretiennent aux terrains d'enquête placent cependant les enquêtrices dans des situations différenciées vis-à-vis de cette démarche de publication: l'enquête Parchemins montre ainsi qu'il est **particulièrement difficile à des chercheurs.euses en contrat à durée déterminée, qui ne peuvent se projeter dans une relation de long terme au terrain ethnographique qu'ils ou elles investissent dans le cadre d'un projet lui-même à durée déterminée, de (re-)donner un sens à cet investissement de forme** considérable et très contraint : en effet, le cadre de collaboration "projet" transforme l'enjeu de capitalisation et d'archivage associé au processus de publication. D'une part, l'enjeu de partage de données en vue de leur analyse croisée, en interne à l'équipe scientifique, apparaît rapidement comme difficilement traitable dans la temporalité limitée de celui-ci (dans le cas de Parchemins, 4 années), si la conception de la chaîne de traitement ou même le processus de documentation des données lui-même sont trop consommateurs en temps, à moins de construire un nouveau projet sur une thématique proche. D'autre part, l'enjeu de capitalisation à l'échelle individuelle, souligné par la seule ethnologue du projet à être certaine d'investir les terrains d'enquête de Parchemins dans le temps long, par ailleurs seule titulaire de la recherche de l'équipe et coordinatrice scientifique du projet, se pose différemment pour les autres membres de l'équipe: pourront-elles, quand bien même elles le souhaiteraient, revenir sur ces terrains d'enquête? Si oui, sont-ce les thématiques et les angles qui ont contribué à configurer et à cadrer le processus de publication (par exemple, les choix d'indexation propres au projet) qu'elles souhaiteront mobiliser ou approfondir?

Ce résultat montre à quel point **le fait que les praticien.nes de la recherche qualitative se trouvent dans des situations structurellement inégalitaires a un impact sur les formes possibles d'appropriation du processus de publication des données, en raison de l'investissement en temps et donc des arbitrages qu'il requiert**, la précarité jouant ici un rôle important dans la capacité à se projeter dans une ligne (voire dans une possibilité) de recherche de moyen terme.

Ces deux éléments (la **précarité structurelle** et les **cultures de l'ouverture par l'intersubjectivation**) invitent à prendre en considération la trajectoire biographique des chercheurs.euses et à prêter une attention soutenue au contexte dans lequel ils ou elles exercent leur activité.

4.1.4. L'intelligence des pratiques: pistes pour une approche de l'ouverture des données intégrant les enjeux épistémologiques et éthiques

Dans la pratique, la mise en évidence de cette tension et la prise de conscience de la portée pratique et épistémologique de l'affirmation d'une intention d'ouverture des données donne lieu à la recherche d'ajustements. Dans le cadre de l'expérimentation d'ouverture menée dans Parchemins, ces ajustements ancrés dans la pratique collective de la recherche se sont révélés, au cours de l'enquête, comme particulièrement performatifs, dans le sens où ils ont permis de faire avancer à la fois la démarche d'ouverture et la qualité des connaissances produites par le collectif de recherche. Ces ajustements sont de trois ordres.

Le détournement des fonctionnalités des instruments

Les outils de publication, qu'ils soient portés par les institutions de recherche elles-mêmes (comme la suite de publication proposée par la TGIR Huma-Num) ou issus, comme dans le cas de Parchemins, de développements "maison" fondés sur l'*Open Source*, sont conçus à partir de mondes et d'expériences éloignés de la pratique ethnographique (Squidant et Revelin 2021). Ils sont par ailleurs orientés plutôt vers la protection de la propriété intellectuelle des auteur.e.s de la donnée publiée, davantage que vers la protection des enquêt.e.s.

L'équipe scientifique a d'abord utilisé les espaces textuels destinés à la documentation des métadonnées pour décrire le contenu de données qui soit ne pouvaient être interprétées sans une guidance du collecteur,

soit étaient trop denses et complexes pour constituer une “unité élémentaire” directement réutilisable (par exemple, un entretien ethnographique peu dirigé de plusieurs heures).

Dans le premier cas, le **risque principal était d'appauvrir la donnée en la publiant**, ou autrement dit de perdre dans le processus de publication les connaissances et interprétations qui y sont incorporées et **de faire d'une donnée de recherche une donnée formellement partagée, mais sans signification**. La production iconographique était ici particulièrement concernée, en particulier deux situations emblématiques : la première concernait les photographies de paysages prises par les enquêtrices, auxquelles elles associaient une lecture experte que la simple publication, y compris associée à des métadonnées formelles détaillées, ne permettait pas de partager. La deuxième situation, également très représentée dans le projet, était celle des archives très techniques et/ou très situées dont les enjeux ne sont pas évidents à identifier pour des non spécialistes, comme les délibérations municipales anciennes. Si le statut public de ces données est garanti par la loi, dans la pratique leur accessibilité est très limitée et leur interprétation implique en général une expertise spécifique.

Pour résoudre cette difficulté, l'équipe scientifique a détourné les rubriques de description formelles des métadonnées pour y ajouter des “métadonnées de contenu”, c'est-à-dire inclure des textes parfois développés permettant d'accompagner la réutilisation (FIGURE 9).



Figure 9. Accompagner la réutilisation des données de recherche en incorporant les lectures expertes dans les fiches de métadonnées : l'exemple des lectures fines du paysage

Par cette démarche, l'équipe de recherche a donc cherché à substituer une perspective substantielle à une perspective procédurale, dans laquelle l'ouverture serait appréciée uniquement par des critères formels et non par les possibilités concrètes de réutilisation à des fins de production de connaissances pour le plus grand nombre.

Dans le second cas, qui concerne principalement les entretiens, le risque était différent : l'ouverture et la protection des enquêtés s'y formulaient comme deux objectifs *a priori* inconciliables, pour des raisons à la fois éthiques et juridiques. Si le procédé de description des "métadonnées de contenu" utilisé pour les autres données fonctionnait très bien pour un partage restreint, en revanche il ne permettait pas pour autant de rendre la donnée publique. L'équipe de recherche a contourné ce problème et pris en compte la protection des enquêtés de plusieurs façons. D'abord, les métadonnées ont été rendues publiques dès que possible, à l'inverse de la donnée qui restait inaccessible. Cela impliquait d'abord de pouvoir dissocier le statut de la donnée et celui de la métadonnée; cette dissociation, essentielle à la conciliation des objectifs d'ouverture et de protection, est encore rarement envisagée dans toutes ses implications, alors que dans la pratique c'est bien cette dissociation qui rend possible la cohabitation, dans un système de publication de type "catalogue", la coexistence de données relevant d'épistémologies et de statuts hétérogènes (Millerand 2011). Elle a dû faire l'objet d'un développement informatique *ad hoc*. Ensuite, les "métadonnées de contenu" ont fait l'objet d'un traitement spécifique très fin, visant à anonymiser les entretiens, tout en permettant aux réutilisateurs potentiels d'identifier très précisément ce qui, dans les sujets abordés au cours de cet entretien, pouvait les intéresser (FIGURE 10).

Entretien avec un-e futur-e exploitant-e sur Brehat

Abonnés
0

Organisation

Parchemins
Paroles et chemins de l'agriculture littorale est un projet interdisciplinaire qui vise à améliorer la connaissance et la compréhension des transformations contemporaines de... [lire plus](#)

Social
Google+
Twitter

Jeu de données **Thèmes** Flux d'Activité

Entretien avec un-e futur-e exploitant-e sur Brehat

Entretien avec un-e futur-e exploitant-e sur Brehat, le 16 mai 2018 à Lannion. (0'-10') parcours vers la reprise de la ferme, formation, difficultés propres à Bréhat, position de la municipalité sur la question, adaptation de la structure et du modèle agricole au contexte social et géographique bréhatin, valorisation et commercialisation, (17'-28') parcours sur l'île avant la reprise de ferme, formation. Ambiguïté des rapports touristes/résidents secondaires – résidents à l'année, question du foncier, investissement dans les dynamiques associatives locales, (28'-40') statut des terres familiales sur l'île, conflits avec les résidents liés aux pratiques agricoles, rapports de force politiques. (41'-50') importance du tissu associatif bréhatin et redéploiement récent, place des néo-insulaires, question de l'identité bréhatine. 50-1'00 rapport aux touristes

Données et ressources

Entretien avec un-e futur-e ... (Unauthorized) Connexion or S'inscrire to access resource
Entretien avec un-e futur-e exploitant-e sur Brehat, le 16 mai 2018 à...

Entretien avec un-e futur-e ... (Unauthorized) Connexion or S'inscrire to access resource
Entretien avec un-e futur-e exploitant-e sur Brehat, le 16 mai 2018 à...

1962 a 1992 1992 a 2010 agriculture agriculture biologique agriculture et tour... agriculture et urba... agriculture insulaire animaux autres apres 2010 associations attachement au lieu avant 1945 bovins allaitants brehat caprins changement des syst... changements dans la... changements ecologi... changements sociaux changements techniques commercialisation commune littorale commune retro-litto... conflits d usage conflits de voisinage culture cultures fourrageres deprise agricole developpement agricole economie touristique elevage encadrement juridique evolution du rappor... foncier

Figure 10. Extrait du catalogue data.agriculturelittorale.fr : exemple de fiche de métadonnée concernant un entretien ethnographique

Il n'en restait pas moins que la plupart des métadonnées d'entretiens restaient sous statut privé, et qu'ainsi, à l'instar des données les moins détachables de leur producteur.trice et de leur contexte de production, l'essentiel du matériau de recherche qui fondait les connaissances produites par les enquêtrices restaient invisibles dans le catalogue présentant les données du projet. L'équipe de recherche a donc conçu un outil *ad hoc* en changeant l'échelle de description des données et en faisant de **la campagne ethnographique une unité élémentaire associée à une fiche de métadonnées**. Ces "fiches campagne" permettent de présenter l'ensemble des données collectées de façon à la fois beaucoup plus synthétique et beaucoup plus

complète que ne peuvent le faire les jeux de données, et de rendre visible leur existence sans restriction. Elles permettent aussi de reconstituer l'intégrité de la démarche ethnographique, tout en facilitant les échanges des réutilisateurs potentiels avec l'enquêteur.trice ([FIGURE 11](#)). Elles permettent, enfin, de faire exister des terrains réalisés en amont de l'adoption des nouvelles dispositions encadrant la publication des données de recherche.

Jeu de données
Thèmes
Flux d'Activité

Campagne de collecte de données ethnographiques - Presqu'île de Rhuys

Fiche synthétique décrivant une campagne de collecte de données ethnographiques menée dans la Presqu'île de Rhuys entre février 2017 et février 2018, dans le cadre du programme de recherche Parchemins (Paroles et chemins de l'agriculture littorale – 2016-2020). Les données collectées portent sur le thème des relations entre agriculture et littoral en région Bretagne. Cette fiche rappelle les objectifs, lieux et durée de la campagne, et caractérise l'ensemble des matériaux de recherche collectés en précisant lesquels sont rendus publics sur le catalogue data.agriculturelittorale.fr ou via d'autres médias en ligne, et lesquels sont conservés par l'enquêteur.trice principal.e.

Données et ressources

 [RH-CCDE-0102017-01022018_Campagne_de_collecte_d ...](#) Explorer ▾

Ethnographie
agriculture
circuits courts
ethnographie
littoral
paysage
tourisme

Info additionnelle

Champ	Valeur
Type de données	Campagne de collecte
Date collecte	
Date saisie métadonnées	2020-11-24 00:00:00
Terrain	Presqu'île de Rhuys
Lieu	
Informateur ou auteur	
Durée	
Langue	fr
Collecteur	Marine Legrand
Email collecteur	marine.m.legrand@posteo.net
Rédacteur	Florence Revelin
Email rédacteur	florence.revelin@mnhn.fr
contributeur	
Email contributeur	
Référence	RH-CCDE-0102017-01022018_Campagne_de_collecte_de_donnees_ethnographiques_sur_l_agriculture_en_Presqu_ile_de_Rhuys
Relation	

Figure 11. Rendre les terrains ethnographiques publics : un exemple de fiche de métadonnées d'une "campagne" Parchemins⁴³

⁴³ Pour accéder à la donnée "fiche campagne" correspondante : http://data.agriculturelittorale.fr/dataset/58244bef-24c6-4bff-a2df-ce8bc683f9ca/resource/d654c426-f5b9-4308-9f87-f91b6ab1ddb0/download/rh-ccde-0102017-01022018_campagne_de_collecte_de_donnees_ethnographiques_sur_l_agriculture_en_pr.pdf

La constitution de l'ouverture des données en question de recherche interdisciplinaire

Le deuxième ajustement issu de la pratique s'est déployé suivant une temporalité plus longue : il s'est agi, pour l'équipe de recherche confrontée aux tensions et arbitrages répétés suscités par la démarche, pourtant choisie, d'ouverture des données de la recherche, de "réouvrir l'ouverture", c'est-à-dire de faire de celle-ci une question de recherche à part entière. Cela impliquait en particulier de prendre au sérieux, sans s'y enfermer, les questions techniques et juridiques que pose l'ouverture des données, et de comprendre où se niche, et comment se déplace, la construction du savoir scientifique dans un tel processus.

Elle est d'abord passée par un travail de la qualification de la donnée ethnographique qui a progressivement intégré la question de sa "détachabilité" du contexte de production. Les données étaient ainsi techniquement qualifiées par leur format (texte, image...), mais également sur un plan plus épistémologique par la possibilité d'en maintenir la qualité au cours du processus de publication, et les conditions pratiques dans lesquelles ce maintien était possible.

Elle s'est ensuite matérialisée par le fait de replacer la publication des données qualitatives de recherche dans un ensemble plus large d'initiatives susceptibles d'en renforcer la pertinence et le sens. Par exemple, en s'interrogeant sur l'accompagnement des réutilisations par des publics non experts, comme les étudiant.e.s, les élu.e.s, les bénévoles d'associations, les agriculteurs.trices. Mais aussi, en travaillant à d'autres formes d'ouverture, moins standardisées et plus orientées vers la médiation.

Conserver aux matériaux ethnographiques leur "chaleur", leurs propriétés spécifiques, lorsque et parce qu'elles résistent à la standardisation, y compris lorsqu'on les oblige à s'y plier : c'est à cette condition que l'expérimentation a pu être menée à son terme et la démarche portée collectivement, sans être reléguée au rang de mal nécessaire, de concession faite de mauvaise grâce aux équipes d'appui à la recherche ou d'intrusion illégitime d'une logique gestionnaire dans le processus de recherche (voir à ce sujet: Beurrier 2019; Revelin et al. 2019; Levain et al. 2021 : 64-81).

4.2 L'enquête PARDOQ : Dispositif d'enquête et profils des chercheurs interrogés

4.2.1. Identifier des profils variés pour recueillir une diversité d'expériences

De façon complémentaire aux enquêtes à large échantillon menées dans un contexte universitaire, et aux retours d'expérience plus ciblés émanant de différent.es chercheurs.euses et collectifs de recherche, les pistes ouvertes par l'enquête Parchemins ont fait émerger deux hypothèses sur lesquelles s'appuie le programme PARDOQ pour fonder le dispositif d'enquête auprès des chercheurs.euses dont les travaux mobilisent des matériaux ethnographiques. Pour les désigner, nous utiliserons ici le terme générique de "praticien.ne.s" qui englobe la diversité des profils et statuts des interlocuteurs.trices auquel.le.s nous faisons référence dans cette partie.

La première hypothèse est que le rapport des praticien.ne.s à la mise en partage de leurs matériaux de recherche est **façonné par les contextes et configurations scientifiques dans lesquelles ils.elles travaillent**. Des rapports différenciés sont ainsi susceptibles d'émerger selon que le.la praticien.ne appartient à un collectif avec une forte culture disciplinaire en anthropologie, où le partage des matériaux de recherche est *a priori* peu envisagé en amont des phases de formalisation de résultats, ou, au contraire un collectif avec une forte culture interdisciplinaire avec d'autres SHS et sciences biophysiques, où d'autres rapports à la "donnée" et "au partage" sont susceptibles de faire évoluer la posture du.de la praticien.ne. De même, les configurations de recherche (plutôt individuelle ou collective, dans des collectifs disciplinaires ou interdisciplinaires, sur projet financé ou non) et **le statut du praticien** (doctorant, chercheur ou enseignant-chercheur, titulaire ou non titulaire) sont également susceptibles d'influencer les rapports aux matériaux de recherche et à leur partage, notamment en lien avec la possibilité de se projeter ou non dans une perspective de long terme.

La mise à l'épreuve de cette première hypothèse a plaidé en faveur d'une approche par des collectifs de recherche porteurs de cultures spécifiques. Nous avons sélectionné trois types de collectifs différents, au sein desquels travaillent des chercheurs.euses produisant des matériaux ethnographiques :

- des collectifs en anthropologie avec une **forte culture de l'interdisciplinarité avec des sciences biophysiques**, et un contexte de **gestion de collections muséales** ;
- des collectifs en sociologie qualitative avec **une forte culture critique appuyée sur les études sociales des sciences et des techniques**;
- des collectifs de recherche avec **une interdisciplinarité entre SHS**.

La seconde hypothèse est celle d'**un effet générationnel** lié à la socialisation précoce ou non aux usages du numérique dans la formation et la carrière des praticien.ne.s. Cette hypothèse suggère que les plus jeunes générations sont susceptibles d'avoir des pratiques de gestion de leurs matériaux de recherche plus nativement numériques et une sensibilité plus forte aux enjeux de gestion et de partage dans le paysage numérique. Nous avons ciblé diverses générations de praticien.nes au sein des différents types de collectifs, en incluant notamment des chercheurs.euses retraités (honoraires).

4.2.2. Le corpus d'interlocuteurs.trices

Sur la base de ces hypothèses, nous avons contacté plusieurs interlocuteur.trice.s en automne 2020 au sein de chacun des types de collectifs retenus (correspondant principalement à l'échelle d'unités de recherche), en utilisant nos réseaux professionnels et en veillant à couvrir une diversité de statuts et une diversité générationnelle. La majorité sont anthropologues ou sociologues, avec des domaines de spécialisation variés (tels que l'ethnoécologie, l'anthropologie de la santé ou la sociologie du travail par exemple). Nous avons également cherché à identifier et contacter à la fois des praticien.ne.s connus ou réputé.e.s pour leur implication en faveur du partage de leurs matériaux de recherche sous d'autres formes que la publication académique formelle, ou au contraire, *a priori* peu enclins à ce type de pratique (sans pouvoir préjuger du

fondement réel ou non de cet *a priori*). L'enjeu était d'accéder à une diversité de motifs à l'appui des pratiques de partage ou au contraire de sanctuarisation (recherchée ou spontanée) des matériaux de recherche, suivant l'idée que comprendre et interroger ces rapports hétérogènes au partage des "données" nécessite de s'intéresser aussi à celles et ceux qui, *a priori*, ne partagent pas.

Les évolutions de l'enquête (réponses aux sollicitations, identification de nouveaux interlocuteurs.trices par effets de réseaux) ont construit un panel asymétrique de 10 praticien.nes (certains collectifs sont plus représentés que d'autres), et riche de la diversité explicitement recherchée. Ce panel a été complété par la participation de 2 autres praticien.ne.s à l'enquête, qui ont spontanément répondu à un appel ouvert lancé sur le réseau professionnel Maté-SHS, davantage orienté vers le recrutement d'interlocuteurs.trices au sein de réseau d'appui à la recherche. L'une est chercheuse (doctorante) en histoire, l'autre est postdoctorante en psychologie de la santé.

Tableau 2. Profils des praticien.n.es interrogé.e.s en fonction des critères de diversité établis pour l'enquête

Les catégories génériques proposées ici pour dessiner les tendances du panel d'interlocuteurs.trices sont formulées par les auteures du rapport sur la base des statuts officiels des praticien.ne.s interrogé.e.s. Elles ne prétendent pas refléter finement la manière dont ces dernier.e.s se définissent. Nous avons, en complément des statuts, renseigné le niveau d'expérience selon l'échelle suivante, en nous référant à l'année de titularisation : « début de carrière » correspond à une titularisation <5 ans, « milieu de carrière » une titularisation entre 5 et 20 ans et fin de carrière, au-delà de 20 ans. La notion de génération est difficile à représenter car les personnes n'ont pas été interrogées précisément sur leur âge. Toutefois, dans notre échantillon, les praticien.ne.s titulaires en début de carrière ou sans poste ont entre 30 et 40 ans, en milieu de carrière entre 40 et 60 ans, et plus de 60 ans pour ceux en fin de carrière ou honoraire. Pour les chercheurs.euses non titulaires, par souci de fidélité à leur situation réelle (puisque les non titulaires ne sont pas nécessairement jeunes), nous avons indiqué si l'expérience de recherche postérieure à l'obtention de leur doctorat est inférieure ou supérieure à 5 ans.

Types de collectifs d'appartenance	Statuts
Collectifs de recherche en anthropologie avec une forte culture de l'interdisciplinarité avec des sciences biophysiques et un contexte de gestion de collections muséales	Un chercheur non titulaire (doctorat <5ans) Une enseignante-chercheure titulaire en milieu de carrière Un chercheur titulaire en milieu de carrière Un professeur titulaire en fin de carrière Un chercheur titulaire honoraire
Collectifs de recherche en sociologie avec une culture critique appuyée sur les études sociales des sciences et des techniques	Une chercheuse non titulaire (doctorat >5ans) Une jeune chercheure titulaire en début de carrière Un chercheur titulaire en milieu de carrière
Collectifs de recherche avec une interdisciplinarité en SHS.	Une enseignante-chercheure titulaire en début de carrière Une chercheuse non titulaire (doctorat > 5ans)
Autres	Une doctorante (en histoire) Une chercheuse non titulaire en psychologie (doctorat <5ans)

4.2.3. Thèmes des entretiens semi-dirigés et posture inclusive

Les entretiens semi-dirigés, réalisés entre mi-septembre et décembre 2020, par téléphone ou en visioconférence en raison du contexte pandémique Covid-19, ont abordé **plusieurs thématiques** :

- les thème(s), terrain(s) et principales méthodes de recherche de l'interlocuteur.trice,
- les principaux matériaux de recherche produits ou mobilisés et leur gestion concrète,
- les pratiques liées au recueil de consentements et leur formalisation,
- les pratiques et expériences de mise en partage de matériaux de recherche sous d'autres formes que la publication scientifique des résultats de recherche,
- les outils et infrastructures numériques utilisés pour gérer les corpus de recherche,
- le rapport aux incitations institutionnelles à ouvrir les données de la recherche et la manière dont la personne est touchée ou non par ce mouvement de l'*Open Data*,
- les problèmes concrets rencontrés ou projetés et les questions soulevées par l'ouverture des corpus ethnographiques.

Au cours des entretiens, une vigilance particulière était portée à la formulation des questions et des relances, de sorte d'**éviter, autant que possible, toute imposition de problématisation et de catégories externes**. Au contraire, nous avons adopté une posture ouverte et inclusive qui porte une attention forte aux catégories des praticien.ne.s. Par exemple, l'emploi du terme "donnée" était soigneusement évité, en particulier au début des échanges, afin de permettre aux interlocuteurs.trices d'exprimer spontanément leurs propres catégories et éléments de langage pour évoquer ce qu'ils.elles produisent au cours de leurs recherches. S'il n'était pas spontanément évoqué, le terme pouvait alors revenir plus tard dans les échanges, soit sous la forme de son usage dans des catégories institutionnelles (par exemple lorsqu'était évoqué la manière dont le.la praticienne était touchée ou non par le "mouvement d'ouverture des données" ou 'd'Open Data' en tant que catégorie de la Science Ouverte), ou encore pour préciser, comme relance en fin d'entretien, si les personnes pensent ou interrogent l'utilisation ou la mise à distance de cette catégorie : "parlez-vous spontanément de données pour évoquer les matériaux de recherche sur lesquels vous travaillez?"

Cette attention à ne pas imposer de cadrage fort dans les entretiens s'inscrit dans le choix d'adopter une posture ouverte destinée à décrire des pratiques des chercheurs.euses, c'est-à-dire dans une démarche ethnographique qui laisse émerger la manière dont les interlocuteurs pensent et vivent des situations, y compris à travers l'énonciation de cette expérience dans leurs propres catégories de langage (catégories *émic*). Cette forme ouverte de recueil d'expérience prend donc en compte l'explicitation de pratiques qui pourraient *a priori* paraître hors champ, comme par exemple des pratiques de partage que l'on pourrait qualifier d'artisanales et très éloignées de la bancarisation de données sur des entrepôts dédiés. L'enquête laisse ainsi émerger d'autres formes de partages que celles entendues au sens du mouvement de l'*Open Data*, et qui peuvent être reliées à des visions et des valeurs plus ou moins divergentes de ce mouvement.

Enfin, l'enquêtrice est elle-même identifiée comme ethnologue par ses interlocuteurs.trices, puisqu'elle se présente ainsi (comme ingénieure de recherche en anthropologie) lors de la prise de contact et au début de l'entretien. Cela influence la nature de la relation d'entretien en générant à la fois une expression assez libre, qui prend appui sur un socle commun de connaissances, de références et de pratiques liées à la démarche de recherche dans cette discipline ; et génère dans le même temps une part d'implicite.

4.3 Résultats de l'enquête PARDOQ menée auprès des chercheurs.euses

4.3.1. Diversité des rapports à la "donnée"

Comment les praticien.ne.s désignent-ils.elles ce qu'ils.elles collectent ?

Une attention particulière a été accordée au cours des entretiens à la manière dont les praticien.ne.s ethnographes interrogé.e.s désignent ce qu'ils.elles collectent: il s'agit en particulier de comprendre si la catégorie de "donnée", structurante dans le mouvement de "l'ouverture des données", fait sens dans ce contexte. On peut regrouper les témoignages recueillis selon trois types de rapports à cette catégorie:

- A. Celles et ceux qui **ne mobilisent pas spontanément la catégorie de "donnée" et privilégient d'autres catégories** pour désigner leurs matériaux de recherche, sans produire de discours particulier sur leurs catégories émic.
- B. Celles et ceux qui **ne mobilisent pas spontanément la catégorie de "donnée" et ont un discours critique par rapport à cette catégorie**, qu'ils mettent explicitement à distance.
- C. Celles et ceux **s'approprient cette catégorie de "données" et en font un usage assez courant**, principalement en lien avec une pratique de l'interdisciplinarité, mais qui produisent un discours réflexif sur cet usage et l'appliquent rarement à ce qu'ils produisent dans le cadre d'une démarche ethnographique.

Le "**groupe A**" est composé des chercheurs.euses qui ont évoqué leurs pratiques et matériaux de recherche dans leurs termes, en les désignant spontanément par des notions regroupées dans le nuage de mots proposé en illustration (FIGURE 12), sans mobiliser la catégorie de "donnée". Par exemple : des "entretiens semi-dirigés", des "observations", des "photographies", des "archives", des "films", des "cartes participatives", etc. Lorsque nous abordons au cours de l'entretien leur rapport à l'usage du terme "donnée" : "parlez-vous spontanément de données?", ces interlocuteurs répondent négativement et précisent les catégories plus génériques qu'ils préfèrent utiliser, mais sans produire de discours particulier sur l'évitement ou la mise à distance de la notion de "donnée". Elle apparaît simplement comme extérieure à leurs catégories émic et leur langage courant pour désigner ce qu'ils.elles produisent. Ils.elles évoquent alors leurs "matériaux" ou "matériaux ethnographiques", leurs "corpus", des "récits" ou leur "terrain" par exemple, pour désigner génériquement les productions de l'enquête.

Le "**groupe B**" rejoint en partie cette posture, dans la mesure où les témoignages recueillis privilégient des catégories qui n'intègrent pas la notion de données. Mais pour ce second ensemble d'interlocuteurs.trices, une dimension réflexive et critique est invoquée, plus ou moins fermement, pour justifier un évitement volontaire de cette catégorie. C'est le cas par exemple d'Amandine, enseignante-chercheuse en socio-anthropologie, qui hésite au moment où nous l'interrogeons sur son emploi du terme "donnée", et finalement le récuse :

« Je parle de données mais ça ne me plait pas? On ne me les donne pas, on me les partage ou on me les prête, mais je n'ai pas de termes plus adaptés... Mes résultats? Je ne l'utilise pas tant que ça en fait, je vais dire plutôt "dans mes entretiens" ou "dans mes observations" ». (Amandine, enseignante-chercheuse titulaire, octobre 2021)

Plus critique, Jacques, chercheur en sociologie avec un fort intérêt pour les sciences et les techniques, déconstruit cette catégorie qui selon lui ne fait pas sens. Il la proscrit de son vocabulaire :

« Je préfère parler de corpus documentaires, de cas, ou de terrain. Mais surtout pas de données. Si par hasard j'utilise ce terme, je mets toujours des guillemets. Je ne considère pas du tout que ce sont des données, ce sont des choses que l'on fabrique et c'est toujours retraité ». (Jacques, sociologue, chercheur titulaire, octobre 2020)

Marie, jeune chercheuse ayant suivi une formation interdisciplinaire en sciences biophysiques et sciences humaines et sociales, souligne que de son point de vue, la catégorie de "données", qu'elle attribue plutôt

au monde des sciences biophysiques qui produisent des matériaux standardisés, n'est pas adaptée pour évoquer ce qu'elle considère comme le caractère "vivant" d'un matériau ethnographique. Elle associe la "donnée" à une dimension "mathématique", "figée" dans un "tableau", et qui correspond à un cadre épistémologique étranger à celui mobilisé dans une démarche ethnographique, et à une catégorie qui n'est pas propre au vocabulaire des communautés SHS.

*« Non je ne parle pas de données, parce que pour moi **c'est vraiment un terme de sciences dures**. Je parle de **matériaux**, et de **sources** comme les historiens. Donnée c'est trop mathématique : ce serait le stade entre la source et le matériau en fait. Si je faisais un tableau avec tous mes interviewés et toutes mes variables démographiques par exemple, là je parlerais de donnée oui. Pour moi les données c'est... C'est un truc que j'avais eu dans ma formation [en biologie], la différence entre données, connaissances etc., et pour moi les données c'est déjà une forme organisée de ton matériau, c'est le stade d'après en fait. **Pour moi ça implique forcément une formalisation**. Tu vois dans la formation en phylogénie, quand tu observes ton fossile et que tu le décris dans ton tableau avec toutes tes variables, ça c'est des données. Mais un entretien, ce n'est pas une donnée pour moi. **C'est trop brut, c'est trop vivant pour qu'on l'appelle donnée. Pour moi dans la donnée il y a un truc de mort, c'est figé pour l'éternité**. Bon après un entretien, quand il est retranscrit il est retranscrit, mais je trouve que ça ne convient pas... Mais je ne saurais pas expliquer. Pour moi une donnée c'est du vocabulaire qui appartient à ma formation de biologiste. Et ce que je recueille dans le cadre de mes enquêtes ethno ça ne ressemble pas du tout à ça, c'est pour ça que je ne les appelle pas comme ça". (Marie, chercheuse non titulaire, octobre 2020)*

On pourrait positionner cette chercheuse à mi-chemin entre le "groupe B" et le "**groupe C**", car son point de vue est en partie partagé par ce dernier groupe qui rassemble des praticiens travaillant communément dans des configurations de recherche interdisciplinaires et collaborant avec des chercheurs.euses en sciences biophysiques ou sciences de la santé dans le cas de notre panel d'interviewé.e.s. Ils.elles mobilisent facilement la catégorie de "données" qui semble être pleinement intégrée à leur vocabulaire, mais principalement en tant que catégorie qui appartient à d'autres domaines disciplinaires, comme le souligne la précédente citation. Cette appropriation est reliée au contexte particulier des configurations de recherche interdisciplinaires qui mettent en frictions des cultures épistémiques diverses (Knorr Cetina, 1999) et hybrident les catégories de langage. Pierre est un jeune anthropologue largement formé à l'interdisciplinarité, notamment au cours de son doctorat réalisé dans le cadre d'un programme ERC. Il travaille actuellement pour un programme de recherche qui implique des chercheurs.euses en écologie, épidémiologie et virologie. Il explique comment il a progressivement intégré "les données" dans son vocabulaire scientifique, comme catégorie-frontière (au sens d'un objet-frontière) pour échanger avec les chercheurs.euses d'autres horizons disciplinaires. Il souligne toutefois son malaise face au manque de réflexivité associé selon lui à l'usage de cette catégorie, qu'il considère finalement comme peu adaptée à la démarche qualitative d'un chercheur anthropologue :

*« En fait c'est un mot que je me suis mis à utiliser **pour me faire comprendre de tout le monde je crois**. Dans un contexte interdisciplinaire le mot « data » ou « donnée », ça **passé auprès de tout le monde**. En tous cas, c'est sûr, plus pour des virologues, épidémiologistes, écologues, etc. Je trouve que c'est le terme passe-partout. Après je n'en ai pas forcément un mieux, depuis le temps que j'utilise ça, depuis ma thèse où vraiment on allait collecter les données quoi. On allait sur le terrain récupérer les données qui existaient déjà en soi. En tous cas dans l'approche quantitative qu'on avait à l'époque, c'était ça. Il n'y avait pas de remise en question du contexte de collecte de la donnée, de l'influence que peut avoir le chercheur, etc, c'était : la donnée existe en soi, on va juste la repêcher. Moi j'avais cette perception-là, c'était la donnée existe en soi, sans nous. **Ce avec quoi je n'étais pas forcément d'accord, parce que, j'étais peut-être le plus réflexif, et parce que j'étais peut-être le plus anthropologue entre guillemets du groupe, quand***

d'autres venaient plus de sciences environnementales ou sociologie». (Pierre, chercheur non titulaire, octobre 2020)

Un autre témoignage rejoint cette idée d'appropriation du terme donnée par effet d'emprunt de vocabulaire et de nécessité de construire un langage commun, dans une configuration de recherche et de collaborations interdisciplinaires, où les sciences exactes sont majoritaires :

*“Je parle de données. Le contexte dans lequel je travaille, c'est un contexte plutôt... C'est très interdisciplinaire. Et je suis amenée à travailler avec des épidémiologistes, des économistes, etc. **Donc c'est le langage. Donc si je veux me faire comprendre il vaut mieux parler de données.** C'est plutôt une question de jargon.” (Aurélié, chercheuse non titulaire en psychologie sociale et santé publique, octobre 2020)*

Il apparaît ainsi dans les témoignages un effet de domination de ce terme vis-à-vis d'autres catégories génériques de désignation des matériaux de recherche : l'emprunt, qui se produit préférentiellement dans les contextes interdisciplinaires avec des sciences dites “exactes”, apparaît ici à sens unique. Même si la catégorie de “donnée” est critiquée pour son inadéquation à désigner ce qui est produit dans une démarche ethnographique et inductive, elle est appropriée par soucis de collaboration et de symétrisation avec d'autres types de matériaux de recherche produits sur la base d'épistémologies différentes, et de simplification du vocabulaire, et ce, malgré les critiques existantes et relativement fortes vis-à-vis du déficit de réflexivité induit par l'usage de ce terme.

A RETENIR :

- **La catégorie de “données” n'est généralement pas spontanément mobilisée par les praticien.ne.s** dont les travaux s'appuient en partie sur une approche ethnographique : **elle ne correspond pas aux catégories *emic* partagées en SHS** pour désigner ce type de matériaux de recherche.
- **L'appropriation de cette catégorie se produit de manière privilégiée dans des configurations de recherche interdisciplinaires qui confrontent et hybrident des cultures épistémiques diverses** : la “donnée” apparaît alors comme une catégorie empruntée à des domaines scientifiques reposant sur des épistémologies où la connaissance scientifique se fonde sur des séries d'informations standardisées et reproductibles. **Son usage repose principalement sur des motifs d'intercompréhension et de symétrisation** de la valeur heuristique des matériaux de recherche issus d'une démarche qualitative et inductive.
- Qu'elle soit appropriée ou mise à distance, voire proscrite, **cette catégorie fait l'objet d'un discours critique**, en particulier vis à vis de son inopérabilité dans un contexte où le rôle et la subjectivité du.de la chercheur.euse est assumée comme consubstantielle de la construction des matériaux de recherche et fait l'objet d'une forte réflexivité embarquée.
- L'usage ou non du terme “donnée” est **un indicateur intéressant pour caractériser et différencier les pratiques et les *ethos*** de chercheurs.euses.

La catégorie de “terrain” : reflet de l'intrication entre contexte, sujets et thèmes de recherche dans la production de matériaux ethnographiques

Pour introduire les entretiens, les personnes interrogées étaient invitées à présenter librement leur activité de recherche, présente et passée, suivant cette question large : « Pour commencer, pouvez-vous vous présenter rapidement : sur quoi travaillez-vous et quelles sont vos principales méthodes ? ».

Libre à chacun.e de développer son propos, chronologiquement, thématiquement, par projets, par terrain, par communauté ou groupe social étudié, etc. Cette entrée en matière volontairement généraliste nous donne à lire comment ces praticien.ne.s présentent leur travail de recherche, et décrivent comment y est

incorporée la production de leurs matériaux de recherche par l'intermédiaire des méthodes utilisées. Une première étape qui permettra d'explicitier en détail, dans un second temps, comment chacun.e construit ses matériaux de recherche, les qualifie, les catégorise, les travaille, les documente, les hiérarchise, et éventuellement les partage.

Trois interlocuteurs.trices engagent ainsi leur présentation, en donnant, de diverses manières, une place de premier plan au "terrain" comme descripteur de leurs recherches:

« Bon, en général ma présentation je la fais d'abord par terrains. Par terrains géographiques. Donc le terrain princeps, c'est le Sénégal oriental où j'ai fait ma thèse. Le deuxième terrain, c'était en anthropologie urbaine à Dakar. [Il poursuit et énumère ainsi les différents terrains et thèmes de recherche abordés au cours de sa carrière] ». (Jean, chercheur honoraire, octobre 2020)

« Moi j'ai été amenée à travailler sur un terrain pendant ma thèse de doctorat, et sur des terrains tout autres par la suite en post-doctorat" [Elle présente ensuite les groupes et thématiques abordées dans les deux cas] ». (Isabelle, enseignante-chercheure titulaire en milieu de carrière, octobre 2020)

« La plus grande partie de ma vie, j'ai travaillé en Afrique centrale, essentiellement dans des communautés de chasseurs-cueilleurs, et après élargi à des communautés d'agriculteurs avec lesquelles les chasseurs-cueilleurs vivent et échangent. [...] Dans les années plus récentes, j'ai travaillé surtout avec des doctorants et masterants, mais j'ai arrêté d'aller sur le terrain, donc là je n'ai pas eu de recherche personnelle, c'était plutôt de l'accompagnement et de la visite de mes doctorants, donc je n'ai plus de terrain personnel depuis assez longtemps ». (Thomas, professeur titulaire en fin de carrière, octobre 2020)

Ces extraits illustrent une tendance que l'on retrouve chez une grande partie des personnes interrogées, en particulier les anthropologues : le "terrain" constitue une entrée privilégiée pour évoquer leur activité de recherche, et plus largement leur champ de spécialisation en tant que chercheurs.euses en sciences sociales. Pour ces interlocuteurs.trices, la production de matériaux empiriques n'est pas tant reliée en priorité à leurs questions ou thématiques de recherche, qu'à leur(s) "terrain(s)", évoqué(s) comme **un ensemble : une catégorie qui englobe le contexte ou les contextes géographiques et les aires culturelles étudiées, les groupes sociaux et les interlocuteurs auxquels ils s'intéressent et avec qui ils interagissent, et les thématiques qu'ils abordent**. Le "terrain" est par ailleurs une notion majeure en anthropologie, support de l'empirisme fondateur de la discipline. Ici abordé comme une catégorie générique, il désigne ainsi à la fois le lieu, les sujets et l'objet de l'enquête, mais aussi le processus à travers lequel le chercheur documente ce triptyque contexte-sujet-thème lorsqu'il fabrique des matériaux qui capturent des "morceaux de réel" tel que le suppose la démarche ethnographique (Olivier de Sardan 1995). Dans les témoignages recueillis, cette catégorie générique du terrain, communément utilisée dans le discours des interlocuteurs.trices anthropologues, préfigure l'intrication complexe dans laquelle s'insère la construction des matériaux ethnographiques. Une intrication déterminante pour l'activité de recherche, car, pour reprendre la dernière citation d'un professeur en anthropologie, c'est selon lui par le terrain - dans le sens d'aller soi-même sur le terrain - que se constitue la contribution scientifique originale du chercheur anthropologue : "mais j'ai arrêté d'aller sur le terrain, **donc là je n'ai pas eu de recherche personnelle**".

D'autres, en particulier Perrine, une sociologue, nuancent cette centralité du terrain dans leur rapport à leur activité de recherche. En fonction des contextes dans lesquels on travaille, du degré d'immersion qu'impliquent les méthodes qu'on privilégie, du niveau de proximité avec les sujets ou situations d'enquête, les discours révèlent une mise à distance de la catégorie de « terrain » pour penser et décrire ses pratiques. Perrine, dans un effort réflexif de caractérisation de sa pratique, considère comme abusif l'usage de ce terme pour évoquer son travail. Après s'être présentée en abordant ses thématiques et questions de recherche comme champs de spécialisation, elle explique :

« Le mot « terrain » je l'évite ; ça donne une vision un peu caricaturale de ce que je fais. Le mot terrain qui est issu de « faire du terrain », en anthropologie [...] J'entends bien ce truc de dire « terrain »... J'ai fait ma thèse au CIRAD donc il y a toute cette culture de dire « terrain », moi je trouve qu'aujourd'hui je fais des entretiens avec des avocats qui habitent dans le 16ème et ce genre de choses. Je trouve que le mot terrain est un peu dépassé par rapport à tout ça, donc je ne l'utilise pas ». (Perrine, chercheuse titulaire, octobre 2020)

Ce que souligne en creux ce témoignage, c'est aussi la reconnaissance et l'admission, dans les communautés praticiennes de l'anthropologie, de ne pas évoquer finement la complexité des matériaux produits lorsque l'on évoque la fabrication de sa recherche. La catégorie générique du terrain, pour évoquer la pratique de "faire du terrain", ou "aller sur le terrain" pour enquêter, permet d'aborder l'activité de recherche sans passer nécessairement par une explicitation fine des méthodes mobilisées ou décomposer la complexité des corpus produits. Dans son approche personnelle et son discours, cette sociologue tend au contraire à décomposer d'emblée cette association entre thème de recherche, groupe étudié ou sujets de l'enquête, méthode utilisée et types de matériaux produits.

Pour les autres praticien.ne.s interrogés, et en particulier celles et ceux qui s'identifient aux méthodes classiques de l'anthropologie, c'est dans un second temps de leur témoignage que se révèle ce qui se produit sur le terrain. Il est donc nécessaire d'entrer dans un niveau plus fin d'énonciation des pratiques qui composent "le terrain" pour comprendre comment les matériaux de recherche y prennent forme, et quelles formes ils prennent.

A RETENIR :

- Le "terrain" est évoqué comme une **unité élémentaire holistique et intégrative** (le cadre épistémologique dans lequel chaque praticien.ne se situe est rendu explicite dans la façon de se présenter) et comme "définition" de soi.
- Les seuls discours d'enquête.e.s qui s'éloignent de ce modèle sont énoncés par des **sociologues** chez qui le poids de cette unité élémentaire semble soit moindre, soit récusé.

Expliciter sa démarche ethnographique : mise en évidence des liens qui unissent le.la praticien.ne à ses matériaux de recherche

Un temps long des témoignages recueillis dans notre enquête est consacré à expliciter finement les pratiques de fabrication des matériaux de recherche. Chaque interlocuteur.trice développe comment il.elle construit et documente le recueil d'informations, en s'appuyant sur **diverses méthodes qui alimentent sa démarche ethnographique**. Les descriptions de ce processus de fabrication de matériaux ethnographiques font apparaître **une diversité de liens qui relie le.la chercheur.euse aux matériaux** qu'il.elle produit, et les rendent plus ou moins indissociables.

Le premier niveau de cette explicitation est celui de l'inventaire des méthodes mobilisées par chacun et chacune. Leur caractère assez "classique" ou "ordinaire" est régulièrement souligné, comme un implicite, presque comme une évidence qui ne nécessiterait pas d'explications particulières pour comprendre ce à quoi se réfère un.e ethnologue lorsqu'il.elle évoque ses méthodes d'enquête, comment il.elle s'y prend, et ce que ces méthodes permettent de produire (peut-être en partie parce qu'ils.elles s'adressent à une enquêtrice elle-même ethnologue) :

« Ma méthode c'est la méthode classique des ethnologues, ce qu'on appelle toujours ordinairement l'observation participante ». (Thomas, professeur titulaire, octobre 2020)

« Moi, c'est une anthropologie, enfin une ethnographie très classique : j'ai procédé la plupart du temps à des entretiens, à de l'observation participante. Je m'inscris dans une ligne qui n'est pas très originale! ». (Isabelle, enseignante-chercheuse titulaire, octobre 2020)

« C'est comme tout le monde : observation participante, flottante, boule de neige, causeries en situation ... » (Jean, chercheur honoraire, octobre 2020)

Pour entrer dans la matérialité de ces pratiques "classiques" de recherche et ce qu'elles permettent de produire, chaque praticien.ne est invité.e à décrire plus finement la manière dont il.elle construit concrètement son ethnographie. Décrire la **matérialité des corpus** produits apparaît alors dans les discours comme indissociable de la description de la **démarche d'enquête** dans laquelle la fabrication de ressources documentaires se produit, et qui implique au premier chef l'implication personnelle de l'enquêteur.trice dans la création d'une relation avec des interlocuteurs. C'est-à-dire que la matérialité des corpus produits incorpore nécessairement une **dimension immatérielle** déterminante, parce qu'elle conditionne la production de ces matériaux, qui devient plus ou moins indissociable de ceux-ci. C'est, pour reprendre les mots de Daniel Cefaï, le propre de "l'engagement ethnographique" (Cefaï, 2010), qui fait de l'ethnographie, plus qu'une méthode, "un art de mener l'enquête" (*idid.*). Jean, chercheur honoraire, illustre bien cette idée, en évoquant avec humour et provocation les conditions de sa mise en oeuvre - se faire "accepter" et créer des liens de confiance :

« Je dis souvent que le travail d'ethnographe sur le terrain c'est de taper l'incruste, et je dis aussi souvent que c'est une relation opportuniste, puisqu'on va au bout du monde pour être très gentils avec les gens, pour qu'ils nous acceptent et qu'ils causent ». (Jean, chercheur honoraire, octobre 2020)

Thomas, un autre interlocuteur résume cette même démarche, indispensable pour accéder, jusqu'où il est possible, aux informations que l'ethnographe souhaite documenter :

« Ca veut dire en gros aller habiter chez des gens qui ne s'y attendent pas, se faire accepter, passer son temps à regarder ce qu'ils font, de temps en temps en bricolant avec eux, et puis leur poser des questions le plus souvent possible jusqu'au point de les emmerder et qu'ils t'envoient balader. Mon prof. il disait : "jusqu'où on peut aller trop loin". Donc c'est à peu près ça le quotidien de ce type de recherche ». (Thomas, professeur titulaire, octobre 2020)

Pierre, jeune chercheur travaillant dans le cadre de projets collectifs et interdisciplinaires entre sciences sociales, sciences environnementales et sciences de la santé, décrit très explicitement comment se construisent progressivement ses "données" qualitatives, en lien avec son expérience du terrain et l'élaboration de sa pensée. Il montre ainsi comment ses matériaux de recherche incorporent selon lui, une dimension immatérielle qu'il ne sait nommer, mais qui est inséparable de son ethnographie :

*« Je pense que, **inconsciemment, les observations et la vie sur le terrain affectent beaucoup la perception de ce que je vais penser et ce que je vais dire.** Dans le sens où je n'aurais pas le même regard sur mes entretiens si j'analysais les entretiens qui avaient été collectés par quelqu'un d'autre. Dans le sens où, **le fait de vivre sur le terrain ne serait-ce qu'un mois, décuple pas mal la vision que tu as d'une donnée qualitative, dans le sens où c'est assez inséparable finalement de la vie sur le terrain, des observations conscientes ou inconscientes que tu fais, et comment ça te marque et t'imprègne dans des choses que tu vois,** et ça affecte en fait ta manière de lire les données. Je ne sais pas comment l'expliquer. Il y a forcément **une part d'interprétation** des données, qui sont une retranscription d'entretiens semi-directifs, du fait qu'on a passé du temps et qu'on a observé des choses. Sans que ce soit forcément des observations qu'on a retranscrites dans un carnet. Mais en tous cas, les souvenirs de ce qu'on a vu, ou des intuitions. Quelque chose qui ne se consigne pas mais qui se passe dans ton cerveau. C'est une sorte de troisième...une sorte de donnée mais qui n'en est pas, qui est : comment tu t'es imprégné du sujet sur le terrain et comment ça imprègne ton regard. Ce n'est pas une donnée en soi, mais ça te donne la capacité d'analyser tes données, comme si les données d'entretiens retranscrites sur un document word, il y a ça, mais il y a aussi quelque chose de plus qui*

est: comment tu as collecté ces données, dans quel cadre, qu'est-ce que tu as vu sur le terrain ». (Pierre, chercheur non titulaire, octobre 2020)

Trois particularités ressortent des témoignages des praticien.ne.s décrivant leur démarche d'enquête ethnographique et qui marque leur rapport à leurs matériaux de recherche : d'abord ils.elles **abordent une posture, des pratiques et des gestes courants** dans la vie professionnelle des praticien.ne.s (c'est-à-dire souvent incorporés dans leurs pratiques de recherche, sans qu'ils ne nécessitent d'être constamment pensés), mais également très importants car **constitutifs de l'originalité de leurs contributions scientifiques** en tant que ce recueil de matériaux uniques qui se produit dans l'enquête ethnographique reflète leur champ de spécialisation, comme supports documentaires de leur "terrain" de recherche. Cet attribut représente une charge importante dans ces "données", car il implique à la fois un rapport individuel relié à l'engagement qui y est incorporé (cf la dimension immatérielle évoquée ci-dessus), mais aussi une forme de mise à nu lorsqu'il s'agit d'envisager de rendre visible ce processus de production (soit en le décrivant dans un entretien, soit en envisageant le partage de ces matériaux). Ainsi, ces gestes et pratiques incorporent également **une forme d'intimité embarquée dans ce travail** à plusieurs niveaux. Une majorité des praticien.ne.s interrogé.e.s me confient s'être rarement exprimés à ce degré de finesse (et de matérialité) sur le sujet. Habituellement, ils évoquent davantage des résultats d'enquête - qui prennent certes appui sur des descriptions, des extraits, des situations, ou des éléments très précis qu'ils ont collectés et sélectionnés - que le détail de tout ce qui la compose. Par exemple, rares sont ceux.celles qui ont déjà expliqué en détails comment ils consignent leurs observations, sur quels supports, ce qu'ils écrivent, à quel moment, etc. Certains soulignent à la fin de leur entretien que l'exercice les a "intéressés". Ils s'y sont déjà livrés de manière générale, ou l'ont abordé en enseignement, mais sans entrer dans la complexité du détail, qui soulève pour eux de nouvelles interrogations - par exemple le fait d'avoir peu remis en question "sa façon de faire" pendant sa carrière comme le soulignent un professeur titulaire et une chercheuse non titulaire. Parler de ce rapport direct, pratique et très concret à la collecte et la production de matériaux de recherche, c'est en effet sortir de l'individualité qui prévaut dans le travail d'enquête ethnographique, et d'une forme de pudeur sur tout ce qui est embarqué dans cet engagement ethnographique. C'est aussi décomposer la "rigueur" (du qualitatif) de la description ethnographique évoquée par Jean-Pierre Olivier de Sardan (2008), ou ouvrir "la boîte noire" des méthodes qualitatives et empiriques pour reprendre une image souvent évoquée dans les contextes de recherche et enseignement en anthropologie. C'est enfin évoquer **sa personnalité et sa subjectivité**, nécessairement embarqués dans ces processus de construction de matériaux de recherche qui, à la fois, assument une dimension interprétative, et reposent largement sur la construction "réussie" de relations interpersonnelles de confiance avec les enquêté.e.s, qui engagent le.la praticien.ne en tant que chercheur.euse et en tant que personne. Ainsi, les matériaux, ou "données" produits incorporent nécessairement ces dimensions, qui scellent des liens à la fois forts et de diverses natures entre le.la praticien.ne et son matériau.

A RETENIR :

- Il existe parmi les personnes enquêtées un **standard professionnel fortement partagé et énoncé, défini comme une démarche** (la démarche ethnographique) **d'avantage que comme une méthode.**
- Ce standard professionnel donne naissance à **des matériaux qui sont, eux, très peu standardisés.**
- Cela est communément admis : il existe une conscience partagée de l'indétachabilité entre les matériaux de recherche et leur contexte de production, qui repose sur l'idiosyncrasie propre aux matériaux ethnographiques : c'est-à-dire une **forme de singularité irréductible, liée au fait que la situation qui a donné naissance à ce matériau n'est pas reproductible, y compris par l'enquêteur lui-même.**

Des matériaux “entremêlés” et “évolutifs” : le matériau ethnographique comme antithèse de la donnée élémentaire stabilisée?

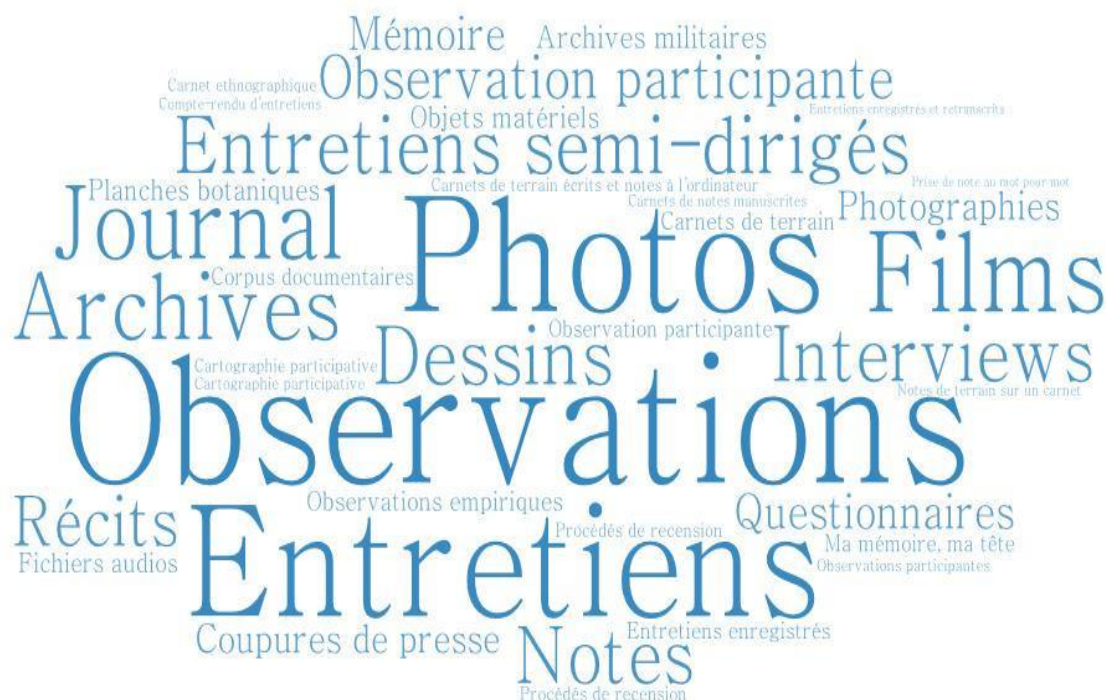


Figure 12. Nuage représentant l'ensemble des matériaux de recherche cités par les 13 personnes interrogées.

La taille de la police est proportionnelle au nombre d'occurrences.

On constate, à travers le nuage de mots présenté en [FIGURE 12](#), la diversité des termes employés par les personnes interrogées pour énoncer les types de matériaux produits durant les phases d'enquête décrites précédemment. Il ressort de ce nuage que **diverses terminologies peuvent désigner un même type de matériau**, qui est pensé ou adapté plus ou moins singulièrement dans les pratiques individuelles, loin de former un standard. C'est le cas par exemple pour les observations - nous y reviendrons - qui apparaissent comme un type de matériau à la fois majeur (commun à l'ensemble des répondants) et multiforme (chacun.e a sa manière de les consigner). L'observation est évoquée à travers des modes de caractérisations différents, par exemple en référence à la posture du chercheur plus ou moins intégré à son observation (observation participante ou observation “simple”), ou en référence à la forme prise par ce type de matériau de recherche : des “notes”, un “journal” ou “carnet” de terrain, ou dans certains cas des films (observation filmée), comme l'évoque un chercheur qui explique que la vidéo est selon lui le mode de documentation le plus complet et fidèle à ce que le chercheur peut observer en tant que personne, à travers ses différents sens.

Si l'on devait regrouper ces différents types de matériaux cités dans des catégories génériques qui synthétisent les principaux matériaux produits, constitués en ressources documentaires, et exploités par les personnes interrogées, nous pourrions proposer les catégories suivantes :

- des entretiens
- des observations
- des vidéos
- des photographies
- des archives

- de la littérature grise
- des recensions (procédés systématiques d'inventaire de noms de techniques ou de plantes par exemple)
- des dessins
- des questionnaires
- des planches botaniques
- des artefacts (objets matériels)

Chacune de ces catégories renvoie à des méthodes particulières du point de vue du type d'informations collectées et des manières de collecter ces informations, ainsi que des formats documentaires pris par ces matériaux, qui sont souvent tous reliés à une même réalité observée ou à une question de recherche commune. Notre plongée dans la matérialité de ces corpus et des pratiques qui les composent met en évidence une double particularité : leur **caractère entremêlé (entre eux) et évolutif**. Nous allons ainsi voir que chacune de ces catégories se résume difficilement à une somme d'unités élémentaires indépendantes qui formerait un corpus homogène, stabilisé et standardisé.

Formes d'entrelacements

Les matériaux ethnographiques présentent la caractéristique d'être **entremêlés** d'abord, parce que l'enquête ethnographique, conçue comme un ensemble cohérent, se compose généralement de différents types de matériaux constitués en ressources documentaires (cités ci-dessus), qui sont **reliés ou combinés les uns aux autres, en particulier par leurs contenus**.

Ainsi, la frontière qui sépare les différentes formes prises par la collecte d'informations n'est pas toujours nette et ne se limite pas, par exemple, à la question des supports sur lesquels sont consignées ces informations. Au contraire, en fonction des pratiques individuelles, ces différentes ressources peuvent avoir des liens très forts entre elles et s'interpénétrer. C'est souvent le cas entre les entretiens et les observations par exemple. Plusieurs praticien.ne.s décrivent ainsi comment le contenu informationnel d'un entretien incorpore les observations qu'ils.elles consignent, soit en y faisant référence, soit directement en hybridant le contenu d'un entretien avec la description de leurs observations. Leurs contenus sont de ce fait conçus comme complémentaires, et parfois inséparables.

Un chercheur qui reprend volontiers le terme de "donnée" pour désigner l'ensemble des éléments informationnels qui composent la réalité, ou la situation observée dans le cadre de son enquête, explique ainsi que "tout fait donnée", au sens où tout est informatif pour lui. Ce "tout" est capté par plusieurs supports et se mélange, en particulier dans les notes prises sur son carnet de terrain:

« Les données, elles sont partout, c'est autant dans la manière dont les gens marchent, se curent le nez, ouvrent la porte de leur voiture, que leur discours. Il y a tout là-dedans. La manière dont ils adressent la parole à un subordonné ou un étranger, tout ça, ça rentre en ligne de compte, évidemment. C'est ça qui te permet de dire qu'il y a une sorte de mépris de classe entre tel groupe et tel groupe en Arabie Saoudite, ce n'est pas les déclarations. Les gens sont polis. Ça ne va pas apparaître dans la formulation ou dans le contenu des mots, mais c'est dans l'attitude. L'attitude c'est une donnée. Ce que je note dans mon carnet, c'est une donnée. Je ne te parle pas de tout mon travail sur les perceptions sensorielles, qui sont souvent indicibles. C'est peu dit, donc tu es obligé d'observer par la marge en fait, par la marge des discours. Donc je le documente par des photos, de vidéos, des audios, carnets de terrain, et dans les carnets de terrain il y a les retranscriptions d'entretiens, et puis les observations que j'ai faites, mes sentiments, analyses, méta-analyses, voilà, il y a tout ça. » (Emmanuel, chercheur titulaire, octobre 2020)

Le matériau de recherche dont parle Emmanuel est donc à la fois, alternativement, successivement, donnée ou métadonnée... documentant une donnée qu'on ne peut saisir, récolter ou mesurer que par fragments.

Un autre chercheur non titulaire explique que lorsqu'il réalise des entretiens informels, ses prises de notes sont documentées avec ses observations, c'est-à-dire dans le même carnet de terrain, où il note aussi chronologiquement ses réflexions, ses réactions, et décrit ce qu'il observe au quotidien. C'est le cas aussi pour Amandine, une enseignante-chercheuse qui explique noter au " mot pour mot", dans son carnet de terrain où sont également consignées ses observations, ce que lui disent les interlocuteurs qu'elle n'enregistre pas. Un autre enseignant-chercheur, qui ne travaille que sur la base d'une posture d'observation participante, inclut directement dans le récit de son observation le contenu, en substance, de ses échanges formels et informels avec ses interlocuteurs.trices. Il n'enregistre et ne consigne que très rarement ses entretiens "à part", si bien que l'ensemble de son matériau d'observation et d'entretiens se retrouve confondu sur un même support (ses carnets de terrain) où lui seul sait se repérer dans ces contenus. En complément, il utilise beaucoup le support photographique et vidéo dans sa pratique : une même situation d'enquête est donc fréquemment documentée par le biais de ces différentes méthodes et supports, qui constituent, entre eux, un réseau d'informations articulées. C'est ce que décrit également Jacques, chercheur en sociologie, qui explique comment il travaille en situation d'observation et produit un corpus hétérogène qui mêle prises de note, recueil de littérature grise et photographies, lui servant à la fois à documenter une situation et à capturer des documents de littérature grise :

« Par exemple, on accompagne la personne, on regarde son travail, on photographie ce qu'il fait, ses documents de travail, ses notes. On relève les traces matérielles de son activité ou travail pour pouvoir revenir dessus. Les photos aident beaucoup. » (Jacques, chercheur titulaire, octobre 2020)

Selon le même type de démarche, Thomas explique comment ses corpus photographiques sont directement reliés à ses entretiens et observations, dans la mesure où il documente des techniques particulières de conception ou d'utilisation d'objets du quotidien. Ainsi, dans ses corpus, le discours sur une technique est nécessairement relié à l'observation qu'il en fait, et décrite précisément dans son carnet de terrain avec le vocabulaire utilisé par ses interlocuteurs.trices et photographiée ; souvent même, il associe à ce chaînage d'informations la collecte de l'artefact concerné qu'il acquiert au cours de son enquête :

« Je suis d'une génération ancienne, c'est-à-dire que je suis d'avant les ordinateurs. Donc je n'ai jamais eu la méthode d'écrire à l'ordinateur ce que j'entends et de retranscrire les entretiens systématiquement. Donc je travaille avant tout avec du papier. Et ce que je n'ai que très rarement fait c'est enregistrer. Donc je fais de l'observation, beaucoup de photos, et puis écrire sur le cahier. [...] Et puis dernier point, et ça c'est mon éducation professionnelle, c'est que j'apporte une importance fondamentale au vocabulaire: j'ai toujours noté le plus systématiquement possible le vocabulaire, le nom des choses, le nom des animaux, des plantes, etc., et ça je le fais en phonétiques et avec les tons puisqu'en Afrique centrale c'est des langues à tons. » (Thomas, professeur d'anthropologie, octobre 2020)

Isabelle, enseignante-chercheuse, raconte à son tour comment plusieurs entretiens de son corpus sont reliés entre eux, dans une démarche de progression de son enquête où elle rencontre plusieurs fois les mêmes interlocuteurs pour enrichir, affiner ou compléter sa collecte d'informations. Elle sait alors relier ensemble cette série d'entretiens dont le sens et le contenu s'enrichissent à mesure de sa collecte, mais qui, pris indépendamment les uns des autres, ne sont que partiellement exploitables. On peut facilement imaginer que le dernier entretien d'une telle série n'est pas pleinement compréhensible sans les précédents.

L'ensemble des témoignages recueillis illustrent ainsi une diversité de configurations qui mettent en évidence comment leurs divers matériaux de recherche **dialoguent** entre eux, à divers niveaux d'intensité, les rendant parfois indissociables les uns des autres. Les praticien.ne.s abordent ainsi ces matériaux non pas comme une série d'unités documentaires indépendantes, mais comme **un ensemble qui fait sens à travers ses interconnexions**, et dont, en général lui.elle seul.e possède **la clé de lecture**. On voit en effet que ces pratiques, si elles correspondent aux canons de la démarche ethnographique, se construisent en finesse à l'échelle individuelle, et que, à la fois la manière dont les différentes ressources documentaires sont reliées

entre elles, et l'intensité de ces liens diffère selon les cas et selon les types de ressources. Ainsi, évaluer le degré de "dépendance" ou "d'indépendance", d'« indétachabilité » ou de « détachabilité » d'une ressource documentaire, c'est-à-dire la possibilité que cette ressource existe en elle-même et pour elle-même en tant qu'unité élémentaire en dehors d'un ensemble plus élargi, et que son contenu fasse sens hors ces liens, nécessite dans beaucoup de cas une connaissance approfondie de cet ensemble et de la manière dont il a été élaboré. Cela signifie que **l'enquêteur.trice est la seule instance en position de pouvoir évaluer facilement ce degré de dépendance ou de "détachabilité", et que cette évaluation se fait au cas par cas.**

A RETENIR :

- **Deux modes principaux d'existence des matériaux de recherche ethnographique coexistent :** le **corpus** et le **réseau**. Le corpus correspond à une mise en série de ressources documentaires du même type couvrant une catégorie générique de matériaux ethnographiques qui renvoie à une classification par types de matériaux. Le réseau correspond à des ressources documentaires de mêmes types reliées à des ressources documentaires d'autres types.
- La forme réseau se prête moins à une conception standardisée des données que la forme corpus : **la forme corpus est compatible avec une gestion standardisée et systématique; la forme réseau, très présente chez les ethnologues, résiste à une gestion standardisée.**

Evolution des matériaux de recherche et hétérogénéité de leur degré d'élaboration

Les matériaux ethnographiques revêtent par ailleurs un caractère **évolutif**: ils apparaissent rarement comme des contenus (ou ensembles d'informations) stabilisés, mais au contraire comme susceptibles d'évoluer, de s'enrichir ou de se réduire dans le temps, et d'exister sous plusieurs formats, avec parfois des niveaux d'élaboration divers.

Un premier niveau relié à ce caractère multiforme et évolutif, c'est la **diversité des supports possibles** pour une même information ou un même contenu. Le cas le plus classique, qui revient de diverses manières dans notre enquête est celui de l'entretien : il peut être enregistré, et ainsi constituer un document sonore, puis se doubler d'un document écrit, soit en étant intégralement ou partiellement retranscrit, soit en faisant l'objet d'une prise de notes complémentaire à l'enregistrement sonore. Un même entretien, en tant que contenu informatif, peut ainsi être multi-support et donc renvoyer à **plusieurs ressources documentaires**.

A travers ces mutations de forme, il peut aussi **évoluer du point de vue de son contenu**. Marie explique ainsi qu'elle retranscrit intégralement ses entretiens (à l'aide d'un logiciel de transcription), puis dans ce fichier texte, elle documente de manière très précise des informations complémentaires, si bien que cette ressource documentaire s'enrichit depuis les phases amont de l'entretien (préparation), jusqu'aux phases de formalisation de sa réflexion :

« Je me suis fait un template de retranscription : chaque entretien débute par un en-tête, où je remplis un certain nombre... ben de métadonnées quoi, et qui est un peu standard sur tous les différents entretiens. J'ai le nom du projet, mon nom, la date, le lieu de l'entretien, qui était présent, s'il y a un fichier audio, le nom du fichier audio sur l'ordinateur pour pouvoir le retrouver, et la date de retranscription aussi. Et ensuite, le fichier d'entretien lui-même il a une forme particulière, il commence par une synthèse d'une page ou une demie page où je note et je garde les idées que je garde de cet entretien. Après il y a le contexte de l'entretien, donc je décris en fait dans quelles conditions il s'est passé, ensuite ... [elle rit] je me rends compte que ça fait vachement sérieux! Après j'ai deux trucs: la mise en contact, c'est-à-dire comment j'ai réussi à entrer en contact avec eux, et aussi la préparation, c'est-à-dire les questions que je voulais leur poser, ou les renseignements

que j'ai cherché avant l'entretien. En fait c'est un fichier qui me sert en amont [de l'entretien] et après aussi ». (Marie, chercheuse non titulaire, octobre 2020)

Jacques, sociologue, distingue ainsi deux niveaux consubstantiels aux entretiens produits lors de l'enquête: il y a d'une part « le matériau que je constitue quand je suis en train d'enquêter », évoqué comme son terrain: "il y a mon terrain", "et un matériau intermédiaire : les comptes-rendus ». Il explique ainsi que tous les entretiens et échanges informels qu'il documente lorsqu'il est en situation d'enquête, « il faut que tout ça soit transformé en compte-rendu ». Pour lui le compte-rendu est très "précieux", car il remet tout en ordre, c'est une forme de retraitement élémentaire (ou de pré-traitement) des ressources qu'il constitue en situation d'enquête, nécessaire pour organiser des informations recueillies selon une méthode qui implique "beaucoup de compilation".

Ce caractère évolutif embarqué dans la fabrique des matériaux ethnographiques se traduit donc de manière privilégiée dans une diversité de niveaux d'élaborations de contenu, qui coexistent ou se succèdent, entre le moment où l'information est capturée sur le terrain et documentée, et la phase finale de son intégration dans des résultats formalisés de recherche. Entre ces deux pôles de la fabrique de la recherche, les matériaux sont rarement considérés comme des ressources stabilisées : au contraire, ils vivent et évoluent. En témoigne par exemple la rareté de l'utilisation de logiciels de CAQDAS (Computer-Assisted Qualitative Data Analysis Software) par les personnes que nous avons interrogées. S'ils permettent d'intégrer et d'indexer différents formats de contenus (texte, image, vidéo, son), ces logiciels d'aide informatisée à l'analyse des données qualitatives impliquent de s'appuyer sur des documents stabilisés. Nous ne pouvons préjuger des différents motifs en faveur de leur utilisation ou non-utilisation; néanmoins, sur les 13 chercheurs.euses interrogé.e.s, un seul avait déjà utilisé ce type de logiciel.

Un cas emblématique de la coexistence de niveaux hétérogènes d'élaboration de contenus ethnographiques est sans doute celui du "carnet", "cahier" ou "journal" de terrain. Objet de l'ethnographe par excellence, on recense à son sujet autant de manières de faire et d'y consigner des informations que d'interlocuteurs.trices. Pour ne retenir qu'un exemple, citons ce professeur d'anthropologie dont le récit d'expérience est très illustratif de l'imbrication de niveaux d'élaborations variés de contenus ([ENCADRE 3](#)).

Encadré 3. Les cahiers de Thomas, professeur d'anthropologie.

(récit recueilli en octobre 2020)

« Je travaille avec du papier. [...] **J'écris sur des cahiers, plusieurs cahiers.** Des cahiers de brouillon où j'écris à la volée ce qu'il se passe sur le terrain, et puis des retranscriptions et des mises au propre soit tous les jours, soit toutes les semaines, de ce qu'il s'est passé, pour récapituler et remettre au propre, voir ce que j'ai entendu, ce dont je me souviens... J'ai toujours des carnets sur moi ».

[Donc vous avez deux types de carnets? Un où vous écrivez à la volée, et un où vous mettez au propre?]

« Oui, **j'ai même un troisième type de carnet, j'ai un toujours un petit carnet dans la poche, où j'écris des choses très précises,** par exemple du vocabulaire. Et donc, **le carnet de brouillon où j'écris des choses plus longues, de l'observation, ou quand je pose des questions aux gens et qu'ils me répondent,** donc j'écris les réponses, et puis **le troisième, c'est à la fois... C'est d'abord un journal, dans lequel il y a régulièrement des mises au propre. C'est un journal global du terrain.** Donc le mieux c'est de le faire tous les jours, mais en fait on n'y arrive pas tous les jours.

Donc après je relis tout ça, et puis j'écris les éléments que j'ai relevés, et puis j'essaie d'en faire quelque chose de correct. Le souvenir le plus important que j'ai, c'est quand j'ai écrit ma thèse. Quand j'ai repris les carnets et mes notes les plus anciennes, j'ai retrouvé des mots, qui, à l'époque je les avais notés correctement, et puis ça avait donné un sens, mais ils ont pris leur vrai sens des années après dans un autre contexte. C'est-à-dire que quand je les avais notés, je n'avais pas conscience de ce à quoi ça référait réellement ».

Le même type de processus d'élaboration existe de manière plus ou moins étendue pour les différentes catégories de matériaux que nous avons énoncées plus tôt. On recense ainsi des témoignages très similaires pour la vidéo par exemple, où les praticien.ne.s décrivent comment un corpus vidéo peut se composer d'une série de rushs, compilées dans des formats intermédiaires d'élaboration en montages, qui impliquent des coupes et associations de plusieurs rushs, ou formats plus aboutis qui suivent une trame narrative et produisent une ressource documentaire nouvelle dans laquelle l'intervention (à la fois dans la forme et sur le fond) du/de la chercheur.euse, éventuellement aidé de technicien.ne.s vidéaste.s s'entremêle aux contenus ainsi agencés. Les photographies font aussi l'objet d'un processus d'élaboration qui est davantage relié à leur classement, leur regroupement en séries ou planches, leur archivage, leur nommage ou l'association d'éléments descriptifs. A l'inverse, la littérature grise ou les archives par exemple apparaissent, dans les discours recueillis, moins retravaillées. Mais elles peuvent aussi exister sous plusieurs formats : le recueil direct des documents, ou leur consignation sous forme de scans ou photographie, qui impliquent alors le type d'opérations que nous avons évoquées pour ces dernières.

Cette question du degré d'élaboration de la "donnée" n'apparaît pas comme un élément spécifique aux matériaux ethnographiques : c'est une problématique récurrente de la mise en base de données, quel que soit le champ disciplinaire et le type de "données" (Jaton et Vinck 2016). Bancariser les "données" implique à la fois de stabiliser les informations qui doivent être saisies, et de fixer un niveau d'élaboration et d'harmonisation (un standard) de ces informations retenues pour intégrer une base ou un entrepôt. Ainsi tout le débat porte sur le caractère "brut", "aggloméré", ou "élaboré" des données, et sur les différents niveaux possibles de stabilisation, qui sont, de fait, le résultat de choix et d'arbitrages opérés par les chercheurs.euses produisant ou alimentant ces bases. Finalement pour reprendre les termes de Lisa Gitelman (2013), la "donnée brute est un oxymore", quel que soit le champ disciplinaire où elle est produite, et sa bancarisation implique toujours de déterminer un niveau d'élaboration de référence. Cela apparaît comme d'autant plus problématique dans le cas des matériaux ethnographiques, où le caractère évolutif et non stabilisé des matériaux est particulièrement fort. Il est assumé par les chercheurs.euses comme constitutif du processus d'élaboration d'une pensée fondée sur une démarche inductive ou abductive : **matériau empirique par définition, il est travaillé et retravaillé de manière itérative, hybridé avec la subjectivité du chercheur, et finalement susceptible d'évoluer au cours de ce processus de fabrication de la recherche.** Ainsi, la coexistence dans les corpus ethnographiques de divers niveaux d'élaboration des

matériaux de recherche reflète l'enchevêtrement progressif entre les éléments recueillis et documentés par le.praticien.ne au cours de son enquête, et la pensée qu'il.elle produit sur la base de la compilation de ces informations. Un témoignage illustre littéralement cet entremêlement en évoquant le *continuum* que forment ses supports matériels et immatériels de données de recherche, qui comprennent sa "propre mémoire", citée comme du matériau produit par l'enquête au même titre que les autres.

Finalement, en se cumulant à d'autres spécificités propres à la fabrique de la recherche qualitative, ce caractère non stabilisé des matériaux ethnographiques les éloignent particulièrement d'une démarche de banalisation sur des entrepôts de données.

A RETENIR :

- La fabrique des matériaux ethnographiques embarque des opérations de recueil d'informations, de tri, d'interprétation, d'analyse et de compilation d'informations qui s'agrègent. **L'information recueillie se mélange donc progressivement avec d'autres et fait coexister ou se succéder des niveaux d'élaborations divers de contenus, qui s'inscrivent dans une démarche éloignée d'une recherche de stabilisation impliquée par la banalisation de données.**
- Ces gradients d'élaboration des contenus impliquent des **degrés de formalisation hétérogènes** de matériaux ethnographiques.
- En fonction de ces degrés de formalisation, **les contenus sont plus ou moins compréhensibles et appropriables par d'autres personnes que le.praticien.ne du matériau.** Il existe donc des niveaux de formalisation qui vont rendre des matériaux facilement détachables de leurs producteurs, et d'autres au contraire qui ne peuvent exister en tant que tel en dehors de leur producteur.trice.

Le rapport matériel aux données : organisations individuelles et formes d'attachements

L'enquête met en évidence des modes d'organisation individuelle des matériaux de recherche. Nous avons déjà vu apparaître à travers les différents témoignages une **hétérogénéité des supports documentaires** utilisés, qui s'ajoutent au caractère entremêlé de ces documents. Les supports peuvent être matériels (papier, supports analogiques par exemple), ou numériques. Les récits des praticien.ne.s interviewés reflètent une diversité de pratiques liées à l'organisation de ces différents documents.

Tout d'abord, nous l'avons vu, il ne s'agit pas de matériaux standardisés. On pourrait au contraire les qualifier de matériaux artisanaux : chaque corpus (ou ensemble documentaire constitué par un.e chercheur.euse) constitue un ensemble unique. Par ailleurs, la **manière d'organiser cet ensemble est propre à chaque praticien.ne** qui construit ses repères en fonction d'un schéma qui fait sens pour lui.elle. Plusieurs interlocuteurs.trices qui ont exercé leur métier avant l'arrivée massive du numérique évoquent de manière emblématique les **bouleversements que représentent l'évolution constante des supports et des formats documentaires**, qui met à l'épreuve les schémas d'archivage personnels, et leur rapport plus ou moins aisé à l'émergence de nouveaux supports ou formats. Il témoigne en particulier du fait que l'arrivée des ordinateurs et du numérique a assoupli leur rigueur d'archivage, dans la mesure où ces supports semblent moins contraignants de ce point de vue que des supports matériels. Adeline, ethnologue, décrit exemplairement cette érosion progressive de ses pratiques d'archivage. Elle souligne au passage sa préoccupation majeure quant aux enjeux de **conservation** de ses matériaux de recherche, sous-jacents à ceux de leur archivage : en plus du risque de s'y perdre dans ses propres archives, elle craint de perdre certaines archives du fait de l'obsolescence des supports sur lesquels elles sont conservées.

Encadré 4. Les entretiens d'Adeline, anthropologue et chercheuse non titulaire, à l'épreuve de leur archivage.

(Récit recueilli en novembre 2020)

« Ma première recherche, c'était il y a déjà plus de quinze ans. Donc **beaucoup de mes entretiens sont sur disquette**, je ne sais même pas s'ils sont encore lisibles. En fait, les premiers entretiens que j'ai réalisés sont sur disquette et je les ai tous retranscrits. Non, pardon, ils sont **d'abord sur cassette** : j'avais un petit enregistreur, c'était le dernier cri à l'époque, et c'était des petites cassettes audio. Donc là j'ai toutes les bandes, j'ai aussi l'enregistreur, donc là j'ai une sorte de musée de l'enregistreur chez moi ! Donc je pourrais les ré-écouter éventuellement, **elles sont conservées dans une armoire chez moi**, et elles ne sont pas ... Et je n'ai pas déposé de double quelque part, et **en fait elles me sont personnelles**. Et j'avais fait aussi des tirages sur disquette, et aussi **des tirages papier de tous ces entretiens**. [...] Ensuite j'ai utilisé un enregistreur numérique, donc **les entretiens sont sur mon ordinateur et une clé USB**. Et là j'ai un problème de conservation parce que toutes ces données-là je ne sais pas très bien où les conserver.

Et justement votre étude m'intéresse aussi parce que c'est un point intéressant la conservation de ces documents. Par exemple **tous les documents sur disquette, je n'ai plus l'ordinateur pour les lire, et je n'ai pas réécouté les bandes, mais j'espère qu'il y a encore quelque chose gravé dessus**. Mais on ne peut pas rêver à une conservation éternelle de tout ça. Je n'ai pas fait de manière systématique, **je n'ai pas été une bonne archiviste** : à part mes cassettes qui ont leur dates, et le nom de l'interlocuteur, après j'ai pas par exemple imprimé de manière systématique tous mes entretiens. Et là je pense que **je le fais de moins en moins, je pense que j'étais plus ordonnée avant**. Peut-être parce que justement avant c'était peut-être moins simple : moi je me sentais un peu, avec ces petites cassettes, le devoir de réaliser des vraies archives bien ordonnées. Et j'ai l'impression que maintenant **avec mes fichiers informatiques, puisque j'ai une manière de les rechercher qui est beaucoup plus simple**, parce qu'il suffit que je tape dans "rechercher" dans mon ordinateur, tel nom et tel terme dont je me souviens qu'il a été employé par quelqu'un, et donc je retrouve. Donc j'ai l'impression que les supports que j'ai utilisés au cours de ces années, toute l'évolution du support finalement, elle a été un peu à l'encontre de l'évolution de mes archives. Hélas ».

Des témoignages similaires décrivent en particulier les complications liées à l'évolution des technologies de photographie et de vidéo (de l'analogique au numérique), et soulignent plusieurs enjeux et préoccupations majeurs de ce phénomène : d'abord les **risques liés à l'obsolescence des supports analogiques** qui peuvent se dégrader physiquement, ce qui implique un risque à la fois de **perdre la possibilité de lire de tels contenus**, ou plus grave **une perte définitive de ces contenus**. En face de cela, le désir de sauvegarder ces contenus sous un format numérisé, se heurte au **coût colossal (moyens et temps) d'une numérisation massive** de ces corpus. Enfin, ces enjeux d'archivage et de conservation sont également mis à l'épreuve du maintien d'une **rigueur d'organisation (archivage, indexation) sans cesse remise en question** par le caractère multi-support des corpus et par l'émergence de nouveaux formats, concomitante de l'obsolescence d'autres formats. De plus, la série d'informations que l'on souhaite associer à chaque ressource pour la répertorier, la classer et la décrire peut évoluer dans le temps, s'enrichir, se réduire ou se modifier. On peut par exemple penser qu'une architecture d'archivage avec une entrée par date et type de ressource est la plus pertinente, puis migrer vers un classement privilégiant une entrée par terrain de recherche, ou encore par thème.

Cette problématique, majeure, n'apparaît **pas suivre un découpage générationnel** et se cantonne aux générations de praticien.ne.s qui ont traversé la révolution numérique. Elle touche également les plus jeunes générations de chercheurs.euses qui, ayant démarré leur carrière à l'heure des ordinateurs et d'Internet, témoignent tout autant de difficultés et d'incertitudes liées aux évolutions et projets de refonte de l'organisation de leurs pratiques et modèles d'archivage. Enfin, l'enjeu souverain est sans doute celui de la **conservation** des matériaux de recherche produits par les praticien.ne.s, en tant que **matériaux précieux et irremplaçables** du fait de leur caractère idiosyncrasique. Ainsi, quel que soit le niveau de rigueur embarqué dans les pratiques individuelles d'organisation et d'archivage, quel que soit le risque d'obsolescence des supports documentaires, **une attention particulière est accordée à la sauvegarde de**

ces matériaux “précieux”. A l’instar de la chercheuse citée ci-dessus qui conserve avec soin ses cassettes dans une armoire chez elle, sans être certaine de pouvoir un jour les relire, nous pourrions ici représenter nombre de bureaux foisonnant d’étagères bondées, où ouvrages de références côtoient cassettes audio et vidéo, disquette, CD ROM, carnets de terrain, boîtes de photographies, etc. Les auteures de ce rapport sont bien placées pour en témoigner : partageant le même bureau lors de notre doctorat, il s’agissait de nous faire une petite place dans une pièce étroite de l’arrière bureau de notre directrice de recherche. Chercheuse brillante ayant produit des centaines de publications scientifiques, son espace de travail était littéralement envahi par ses dossiers et matériaux de recherches. Plus de la moitié de notre bureau de 15m² était en quelque sorte sa salle d’archives personnelles, peuplée d’unités centrales périmées, de disquettes et CD ROM, d’artéfacts, de centaines de dossiers d’archives constitués de documents de terrain, de littérature grise, de versions de travail de productions scientifiques, de rapports, etc. qu’elle conservait précieusement et pour lesquels nous avons un respect immense : nous n’osions déplacer un objet ou empiéter sur leur espace. La perspective d’un déménagement du laboratoire en 2015 a généré une angoisse majeure, pour elle et nombre de nos collègues, confrontés au déplacement de ces fonds volumineux vers d’autres locaux, où les bureaux étaient moins spacieux. Cela impliquait à la fois de ne rien perdre de “précieux” dans ce déménagement opéré par un tiers (pour le déplacement des cartons), mais également de faire du tri (angoissant vu le volume à trier dans un moment où l’on n’a pas forcément envisagé de s’atteler à cette tâche) et de conserver la cohérence d’organisation personnelle de chaque ensemble, mise à l’épreuve par ce déplacement.

La transmission de ses archives en fin de carrière : le passage coûteux (et salutaire) d’un archivage personnel à un fonds appropriable par d’autres

Si la valeur des matériaux ethnographiques est très grande aux yeux de leurs producteurs.trices, elle s’ouvre à une préoccupation forte et pressante en fin de carrière : celle de transmettre ces matériaux uniques. L’émergence de cette préoccupation patrimoniale constitue une étape particulière du point de vue du rapport à l’organisation et l’archivage de ces matériaux : c’est un temps de mutation depuis un schéma personnel d’organisation sans cesse renouvelé comme nous l’avons vu, vers un schéma d’archivage plus indépendant du chercheur, c’est-à-dire à la fois être compréhensible et appropriable par d’autres, et qui donne accès à des matériaux qui puissent être eux-mêmes compréhensibles et appropriables par d’autres (ce qui suppose une sélection de ce qui peut être transmis, et ce qui ne peut pas). C’est donc en quelque sorte, le temps du grand tri et de la grande revisite de ses corpus et de sa rigueur organisationnelle à l’échelle d’une carrière entière. La photo ci-dessus illustre le travail minutieux et de longue haleine entrepris par l’ethnobotaniste Claudine Friedberg pendant sa retraite, aidée par une étudiante, pour mettre de l’ordre dans ses archives de terrain avant de les déposer au fonds d’archives du Muséum d’histoire naturelle. Elle est ici photographiée en 2009 avec son imposant herbier, qui ne constitue qu’une partie de ses fonds personnels (PHOTOGRAPHIE 1).



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

Dans notre enquête, deux témoignages sont particulièrement marqués par cette préoccupation de transmission, et sont sans surprise celui d'un chercheur honoraire (Jean) et celui d'un autre chercheur proche de la retraite (Thomas). Le premier étant engagé dans un tel processus de transmission, le second s'y projetant largement. Chacun explique les opérations de rangement, de tri et de documentation complexe des matériaux collectés au cours de leur vie d'ethnologue, en soulignant le coût majeur de ces opérations : elles nécessitent une grande disponibilité en temps, et éventuellement un appui externe, financier ou humain. Elles supposent aussi, en lien avec ce qui a été évoqué *supra*, des opérations coûteuses de numérisation, dans la perspective de faire exister et perdurer leur corpus après eux et les rendre appropriables par d'autres. Elles impliquent enfin une part de renoncements, et parfois de regrets vis-à-vis de choix qui ont été faits par le passé et dont la pertinence n'est plus avérée face à ce "nouvel" enjeu patrimonial.

Thomas raconte ainsi le coût de la numérisation de ses photographies, qui représente une opération massive et laborieuse, et bouleverse ses propres pratiques passées d'organisation et de documentation de ses corpus (ENCADRE 5).

Encadré 5. Face à la transmission. Thomas, Professeur d'anthropologie.

(Récit recueilli en octobre 2020)

« Mon premier terrain c'est en 1969. Ma période intense de terrain c'est de 1972 à 1986. Donc la majorité de mes photographies de terrain, elle est sur support... de négatifs. Donc j'ai pris très tôt l'habitude d'avoir deux appareils photos similaires, un chargé en diapositives et un chargé en noir et blanc, parce qu'à cette époque-là, les photos en couleurs étaient impossibles à publier. Donc avec le noir et blanc je faisais plus de photos d'objets, des photos techniques ou des photos plus documentaires. [...] Alors pour les négatifs, j'ai toujours fait des planches contact. Donc j'ai à la fois des négatifs dans des bandelettes et dans des classeurs et puis j'ai une planche contact avec toutes les photos d'une même pellicule, et **en général c'est sur la planche contact que j'ai marqué date et lieu, au moins date et lieu c'est quand même la base, et puis parfois le contenu.** Alors il y a eu une époque où j'étais très enthousiaste et je marquais à l'encre de chine sur la bande, mais je n'ai pas continué ça. [...] A l'inverse, pour faire des conférences, tu projetais des diapositives, tu ne pouvais pas faire des conférences avec des négatifs. **Donc la majorité des photos sont en diapositive, dont seulement une fraction a été scannée.** Il y en a quand même beaucoup qui ne le sont pas. Il y a quand même tout ce stock de photos qu'il va falloir que je gère. Alors les diapos je marquais dessus sur le cadre, au moins la date et le lieu, quelquefois le sujet. **Le bazar a commencé quand j'ai sélectionné des photos pour les faire scanner, et à ce moment-là elles ont commencé à être mélangées.** Et l'autre problème, c'est que **quand tu scannes des photos, tu perds des informations de base qui sont sur le carton,** parce qu'il faut rencoder tes photos scannées, tes images en JPEG et en général on ne le fait pas et là c'est vraiment le foutoir ».

[Et cette opération vous la faites vous-même ?]

« Oui c'est obligatoire, parce que d'abord on n'a pas de petites mains pour le faire et puis ensuite en gros **il n'y a que moi qui sais ce qu'il y a dans mes affaires, c'est un peu ça le problème quoi.** Et ça prend énormément de temps ».

A RETENIR :

- **L'organisation des corpus ethnographiques se conçoit à l'échelle individuelle** : les praticien.ne.s ne s'approprient pas de standards d'archivage mais construisent leur propre schéma.
- **Les ressources documentaires** peuvent exister simultanément sous différentes formes et avec différents degrés d'élaboration, qui constituent autant de traces de la pensée du chercheur mais qui **ne sont pas regroupables dans un format unique.**
- Il existe des transformations socio-techniques qui font que **tous les chercheurs.euses, y compris les plus jeunes, ont connu des changements d'outils, de supports et de formats qui rendent tout**

choix d'archivage coûteux et risqué. Ils.elles ne peuvent jamais faire un choix systématique et définitif qui leur permette de gérer et d'organiser l'ensemble de leur corpus, mais doivent au contraire faire cohabiter structurellement plusieurs formats.

- **L'enjeu de sauvegarde et d'archivage est fort, et particulièrement relié au caractère unique, original et non reproductible des matériaux ethnographiques, qui leur confère une valeur particulière et les rend précieux** aux yeux de leurs producteurs.trices.
- **La perspective de la fin de l'activité de recherche (fin de carrière) est susceptible de faire émerger des enjeux de patrimonialisation** liés à la transmission des corpus de recherche.

Typologie des rapports aux matériaux de recherche ethnographiques : profils "collectionneurs", "auteurs" et "coopératifs"

L'enquête met en lumière une diversité de liens et de caractéristiques qui rendent les matériaux de recherche produits dans le cadre d'une démarche ethnographique difficilement détachables de leurs producteurs.trices et de la relation qui lie ces dernier.e.s à leurs interlocuteurs.trices. Cette tension entre attachement et "détachement" (ou existence propre de ces matériaux de recherche en dehors de leurs producteurs.trices) s'énonce dans les entretiens menés auprès des chercheurs.euses selon quatre modalités principales :

- d'abord leur ancrage dans une relation de l'enquêteur.trice à l'autre, qui les inscrit dans **une éthique** de recherche particulièrement forte qui les institue en produits et traces de la relation de confiance qui lie le.la chercheur.euse à son terrain et ses interlocuteurs.trices,

- leur représentation comme **extension de la pensée du.de la chercheur.euse et de son acuité d'observation**, qui leur confère un caractère personnel du point de vue du.de la praticien.ne qui y incorpore son "filtre" (pour reprendre un terme cité dans plusieurs entretiens), son regard, son expérience, sa subjectivité, son interprétation, et donc une part de son **intimité**.

- leur **caractère entremêlé et évolutif**, qui peut revêtir diverses formes, et qui a pour principal effet d'éloigner ces matériaux de recherche de modèles d'unités élémentaires indépendantes, stabilisées et standardisées.

- leur **mode d'existence dans des organisations documentaires personnelles aux praticien.ne.s**, éloignés de standards d'archivage, et davantage centrés sur l'enjeu de se repérer, selon ses propres schémas organisationnels, dans la collecte et la mémoire des traces de son enquête, et dans l'élaboration d'analyses inscrites dans une démarche de recherche inductive.

Si ces différents registres d'attachements caractéristiques du rapport entre praticien.ne.s de la recherche et matériaux ethnographiques sont forts, il apparaît toutefois que le degré de "détachabilité" peut varier en fonction des types de matériaux : que ce caractère détachable ait été anticipé ou qu'il soit simplement possible "de fait" (c'est ce que nous aborderons plus en détail dans la partie suivante), certains matériaux paraissent plus susceptibles d'être publiés, partagés, détachés de leurs conditions de productions et d'usage originelles. C'est le cas en particulier des photographies, des artefacts, des archives et des vidéos, qui sont susceptibles de se prêter plus facilement à des modes d'exposition et de partage, lorsque leur contenu n'est pas jugé "sensible" (c'est à dire en capacité de nuire à l'enquêteur.trice ou l'enquêté) et ne fait pas l'objet d'une protection juridique particulière. Mais la mesure de ce degré de détachabilité ne peut être prise uniquement à l'aune des caractéristiques spécifiques de chacun de ces types de matériaux, de leurs contenus, et de leur statut juridique, que nous pourrions décliner à l'infini tant la diversité des configurations possibles est immense (c'est bien là tout le défi qui se présente pour la qualification juridique de ces matériaux qui ne parvient à pas à embrasser la diversité des situations) : elle embarque au contraire nécessairement la question du **rapport de chaque praticien à ses matériaux et à leur partage**.

A cet égard, notre enquête fait apparaître trois profils distincts parmi l'échantillon de personnes rencontrées: un profil "auteur", un profil "collectionneur" et un profil "coopératif". Les trois ne s'opposent pas : ils représentent davantage des *ethos* qui s'inscrivent dans un *continuum* de pratiques de fabrication de la recherche ethnographique, et peuvent parfois coexister chez un même praticien en fonction des situations et des projets de recherche.

Le profil **auteur** correspond aux praticien.ne.s qui considèrent que la "donnée", "leur matériau de recherche" est produit d'abord selon l'objectif de nourrir **l'élaboration d'une pensée et d'un récit sur la base du croisement des traces collectées dans l'enquête**. Ces matériaux, ces traces, sont ainsi restitués comme tant d'éléments empiriques qui intègrent des textes et autres œuvres produites par ces auteurs (des films ethnographiques par exemple) et qui embarquent à la fois des opérations analytiques et réflexives sur les matériaux produits et mobilisés. Dans cette conception, ces matériaux sont envisagés comme partageables uniquement sous ces formes élaborées et accompagnées par son récit. Leur partage à des stades antérieurs, parfois qualifiés faute de mieux de "bruts", ne fait pas sens dans leurs pratiques : le.la chercheur.euse auteur.trice produit et partage des réflexions et des textes, et pas des "données"⁴⁴. L'existence-même de ces matériaux (leur production) étant subordonnée au filtre ou à la subjectivité assumée du chercheur - par son regard personnel et informé sur une situation d'observation, par son expérience même de praticien.ne de la recherche ethnographique, par son talent d'observateur en quelque sorte - **leur partage n'est alors envisagé que dans le cadre de l'aboutissement du processus d'interprétation et de réflexivité qui se fonde dans et sur ces matériaux**. On pourrait comparer cette posture à celle d'un.e artiste qui dévoile son œuvre lorsqu'il la considère achevée. Ces matériaux de recherche n'apparaissent donc pas pouvoir exister en dehors de ces auteurs.autrices, seuls à en posséder les clés de lecture éthiquement acceptables et scientifiquement rigoureuses. Isabelle, enseignante-chercheuse incarnant particulièrement ce profil, souligne ainsi que de son point de vue, le partage de ses matériaux de terrain ne lui paraît ni possible, ni souhaitable. Pour elle, c'est le propre de la démarche ethnographique que de produire des matériaux où "tout passe par le filtre de [sa] sensibilité", et "tout se joue dans ce filtre" : il lui paraît ainsi difficile de partager "ce qu'on a recueilli, ou ce qu'on a compris avec notre subjectivité" (Isabelle, enseignante-chercheuse en anthropologie, octobre 2020). C'est pour cette raison qu'elle qualifierait volontiers "d'arts humains" plutôt que de sciences sociales sa démarche de recherche, qui selon elle, gagnerait à se rapprocher des disciplines littéraires.

Le profil du **collectionneur** se distingue de cette première posture, dans le sens où la démarche scientifique des chercheurs.euses s'inscrit davantage dans **une logique d'accumulation de matériaux qui se construit en embarquant une dimension de partage et de publicisation**. Cette démarche implique une détachabilité de certaines ressources documentaires collectées sur le terrain, dont le.la collecteur.trice se sent autant dépositaire qu'analyste, et qui existent alors en dehors (en complément) d'un processus d'élaboration et de restitution par le biais de l'autorité. C'est le cas en particulier pour les praticien.ne.s qui incorporent dans leur enquête ethnographique un objectif de partage à court terme de certains éléments collectés sur le terrain, tels que les chercheurs travaillent dans des configurations muséales et particulièrement sensibles à cet enjeu. La collecte d'objets, de photographies et de films prend pour eux.elles une place importante sur le terrain et sont voués à alimenter des collections pouvant faire l'objet d'expositions. Ils.Elles intègrent dès les phases de collecte une logique de partage et de dépôt dans des structures dédiées à l'archivage, la conservation et la publicisation de ces éléments qui sont (en partie) gérées par d'autres acteurs.trices : par exemple des collections scientifiques ou patrimoniales, des vidéothèques ou sonothèques numériques. Ainsi le collectionneur s'inscrit comme un dépositaire, un maillon, ou un passeur dans le chaînage entre une réalité - une société - observée et une instance de conservation et de valorisation de cette observation vers un public élargi. Le partage est donc incorporé à la démarche même de collecte, mais pas forcément dans les conditions et dans le cadre normatif générique de la publicisation des données de recherche. Les interlocuteurs.trices poches de ce profil « collectionneur » ont été particulièrement disert et détaillé dans

⁴⁴ Pour une théorisation de ce point et une analyse des effets de réalité construits par et dans les grands textes anthropologiques, voir: Geertz 1996.

leurs récits d'expérience à l'occasion de notre enquête, du fait de leur affinité plus prononcée envers le thème du partage que nous abordions en entretiens. Ainsi, les extraits de leurs entretiens, plus développés, alimentent plus largement ce rapport que ceux de deux autres profils.

Enfin, le troisième profil, **“coopératif”**, s'inscrit dans une logique de **mise à disposition et de croisement des matériaux de recherche à une échelle collective**. On retrouve principalement cette logique auprès des chercheurs.euses travaillant dans des configurations de recherche sur projets financés (type projets ANR ou ERC) qui créent des collectifs institués mettant en présence des équipes de recherche souvent interdisciplinaires et porteuses d'une intention de partage d'une partie des matériaux de recherche à des fins d'analyse croisée : le partage implique donc un processus de standardisation de la collecte et de la mutualisation. Dans ces configurations, des protocoles de collecte de données sont généralement produits dans les phases de lancement des projets, et définissent quels matériaux seront partagés et dans quel périmètre. Cela implique que l'objectif de partage est intégré par l'enquêteur.trice au moment où celui.celle-ci procède à ses enquêtes, et que ce partage est délimité. Dans les témoignages recueillis, il s'agit principalement de partages à l'échelle de l'équipe du programme, voire d'un groupe de travail restreint au sein de cette équipe. La démarche adoptée par l'équipe scientifique du programme PARCHEMINS présentée dans la partie précédente s'apparente à cette logique coopérative. Dans ce cas, on voit que l'intention de partage peut déborder les frontières de l'équipe interdisciplinaire, pour répondre à un enjeu de transdisciplinarité, c'est-à-dire un partage de certains matériaux de recherche avec la société, selon une approche valorisant la co-construction et la symétrisation des savoirs entre sciences et société.

A RETENIR

- **L'anthropologie est traversée par des clivages internes et historiques : pour les uns, elle est avant tout une interprétation écrite, alors que d'autres considèrent comme possible (voire recherchent) un détachement d'avec le matériau empirique collecté.** A la fois les postures scientifiques, mais aussi les rapports pratiques aux matériaux évoluent au cours du temps et ne sont pas figés.
- **Le “détachement” des matériaux de recherche se produit dans deux types de circonstance spécifiques :** lorsque le.la praticien.ne est animé.e par une logique d'accumulation qui est en partie construite publiquement (c'est le cas du profil “collectionneur”); lorsqu'il.elle évolue dans des configurations de recherche collectives qui engagent une mise en commun et un croisement des matériaux de recherche (profil “collaborateur”).
- **La perspective de détachement n'est que rarement envisagée par les profils “auteurs” :** ils.elles partagent des pensées et des récits, et pas des matériaux de recherche.

4.3.2. Rapports à “l’ouverture” des données

“L’ouvert” et le “fermé”: deux pôles hypothétiques d’un continuum abordé par le rapport au partage

Au même titre que la catégorie de “données” nous interroge sur le sens qui lui est ou non donné par les communautés praticiennes de la recherche, et la façon dont ces dernières définissent ce qu’elles produisent, le rapport à l’ouverture de données soulève le problème épineux de ce qui est entendu par “données ouvertes”, et selon qui?

Nous avons vu en partie 2 que l’ouverture des données telle que portée dans le mouvement de l’*Open Data* est conçue, pour résumer, selon deux principes majeurs : d’abord celui de l’ouverture comme règle et la fermeture comme exception selon le principe souverain “aussi ouvert que possible aussi fermé que nécessaire”; et deuxièmement le fait que cette ouverture, lorsqu’elle est possible, corresponde à une mise à disposition publique de données sur des entrepôts numériques dédiés respectant les principes FAIR, et certifiés.

Un premier constat qui émerge des témoignages recueillis dans notre enquête est que **la notion d’ouverture ne fait pas spontanément sens dans les discours et les pratiques des praticien.ne.s interrogé.e.s**, et que lorsqu’elle est utilisée, le sens qui lui est donné ne recouvre que peu ou pas celui qu’elle revêt dans les textes de droit et chez ses promoteurs. C’est pourquoi nous préférons ici partir de la notion de “partage”, privilégiée dans notre grille d’entretien, et qui évoque l’ouverture plutôt qu’elle n’en impose un sens. Cette approche permet de recueillir la diversité des expériences de nos interlocuteurs.trices, sans opposer de fait deux catégories hypothétiques et polarisées de pratiques (l’ouverture ou la fermeture des données), mais au contraire une multiplicité de possibilités, de niveaux et de formes de partages qui constituent un *continuum* entre une publicité sans limites des matériaux de recherche et un accès réservé uniquement à son producteur. Ainsi, en nous appuyant sur la caractérisation de la diversité des rapports des praticien.ne.s aux matériaux de recherche qu’ils.elles produisent (partie 4.3.1), l’objectif de cette partie est de comprendre dans quelles mesures et sous quelles formes ils.elles envisagent et pratiquent leur partage.

Pratiques et enjeux de partages exprimés par les praticien.ne.s

La partie précédente a mis en évidence des formes et des enjeux de partage incorporés dans les pratiques de recherche et qui apparaissent comme particulièrement signifiants pour les praticien.ne.s interrogé.e.s. La plus évidente est l’organisation pratique de la transmission, qui s’intensifie dans la perspective de la fin de carrière.

Transmission : les matériaux de recherche comme archives

La transmission des matériaux ethnographiques en tant qu’archives est la forme de partage la plus commune⁴⁵, et également la plus étudiée (Molinié et Mouton 2008; Mouton 2018; Zeitlyn 2019). C’est principalement par ce canal que les ethnographes abordent le sujet du partage. Sans revenir ici en détails sur les enjeux et pratiques de transmission que nous avons déjà évoqués dans la partie précédente, soulignons en particulier que leur spécificité est **l’incorporation *ex post* d’une dimension patrimoniale dans la conception que le.la praticien.ne a de son corpus de recherche**. Ce passage transforme l’ensemble des données transmises en “fonds”, pour reprendre la terminologie employée par les bibliothèques ou instances qui en sont dépositaires. Le.la praticien.ne engagé.e dans ce processus de transmission conçoit alors les matériaux ethnographiques qu’il.elle a produits comme des “archives” de sa recherche, portant sur des sujets, des terrains et des situations (la vie de communautés, leurs pratiques, rituels, chants, cérémonies,

⁴⁵ Notons ici que dans certains cas, cette transmission est mise en œuvre par les ayants-droits des praticien.ne.s après leur décès, sans que ces dernier.e.s n’aient pleinement participé à ce processus. Cette configuration ne semble pas être rare, néanmoins, nous nous concentrons ici sur le cas de chercheurs.euses qui s’engagent eux-mêmes dans ce processus dans la perspective de la fin d’activité.

systèmes de parentés, rapports au corps, rapports au milieu environnant, etc.) documentés sur une période donnée. Ces matériaux et ressources documentaires, mutés en fonds d'archives constitueront une trace unique (et donc précieuse) de ce que son activité de recherche ethnographique a permis de documenter. Pour Thomas, comme pour plusieurs des chercheurs.euses interrogé.e.s, transmettre ses archives apparaît comme une nécessité :

« De toutes façons, qu'elles soient archivées, c'est une nécessité. Pour ça, il faut que je les prépare correctement, parce que si je laisse derrière moi des boîtes entières, il y a toutes les chances que ça parte à la poubelle. Donc il faut organiser la survie de ces choses-là ».
(Thomas, professeur d'anthropologie, octobre 2020)

Jean exprime un point de vue très tranché à ce sujet : son expérience le convainc que ce qui reste de plus important de l'activité de recherche d'un ethnologue, avant même ses publications et ses écrits formels, **ce sont ses matériaux originaux.**

*« J'ai la conviction profonde, que ce qu'il reste d'une carrière de chercheur, ce ne sont pas les articles. [...] Ce qui m'a le plus ancré dans ça, c'est quand j'ai vu que [des doctorants] n'avaient pas lu la thèse de leur directeur de thèse. Et donc tu t'aperçois que les articles, personne ne les lit. Ou seulement si tu fais une recherche qui est très pointue et en rapport avec un sujet traité ailleurs. Je pense que certains ouvrages sont des références. **Mais dès que tu as ta collecte ethnographique, faite dans des conditions correctes, que ce soit par des photos, des vidéos, des machins, je pense que c'est ça qui reste.** Et par exemple à l'heure actuelle, qu'on parle du taux de citation de Lévi-Strauss depuis qu'il est mort et son taux de citation avant, tu t'aperçois que c'est très éphémère. Alors que les photos, les objets, les films, les documents bruts quels qu'ils soient... Et je le vois moi-même, dès que je vais faire des recherches, quand je tombe sur des matériaux ethnographiques, je lis. Par contre quand je tombe sur les analyses qui en ont été faites, eh bien je saute souvent. **Pour moi la partie concrète qui reste, ce sont les matériaux ethnographiques** ». (Jean, chercheur honoraire, octobre 2020)*

Les témoignages recueillis et notre expérience de membre de cette communauté de recherche ayant côtoyé ou côtoyant des chercheurs.euses en fin de carrière convergent avec le constat général que l'enjeu de transmission est une forme privilégiée de partage pour cette communauté de praticien.ne.s. L'équipe du programme BeQuali le considère comme l'un des principaux motifs de dépôts de corpus dans cette banque d'enquête (3.2.2.), et outre les fonds d'archives physiques, les banques de données ethnographiques numériques ou numérisées les plus connues à ce jour sont largement alimentées par ce type de dépôts (par exemple la phonothèque de la Maison méditerranéenne des Sciences de l'Homme, le catalogue Telemeta du CREM, la plateforme ODSAS du Centre de recherche et de documentation sur l'Océanie). L'ampleur de cette préoccupation s'illustre également à travers l'institutionnalisation, il y a une dizaine d'années, du Consortium des "Archives des ethnologues" (AdE), comme l'un des premiers consortiums de la TGIR Huma-Num⁴⁶. Au sein d'Huma-Num, les consortiums constituent des "dispositifs de concertation collective"⁴⁷ à l'échelle de différentes disciplines, qui portent une réflexion fondée sur les préoccupations et spécificités internes à chacun de ces domaines scientifiques des SHS. Les communautés d'ethnologues ont ainsi investi cette arène de réflexion en y entrant par leurs "archives", un enjeu porté en particulier par plusieurs bibliothèques et centres des ressources adossés à des laboratoires d'anthropologie⁴⁸, et se retrouvant

⁴⁶ Aujourd'hui fer de lance des politiques d'ouverture et de FAIRisation des données de la recherche en SHS, la TGIR Huma-Num se fonde avant tout sur l'objectif de faciliter le tournant numérique dans la recherche en sciences humaines et sociales en contribuant à équiper ces communautés de recherche.

⁴⁷ site de la TGIR : <https://www.huma-num.fr/>, consulté le 15/04/2021.

⁴⁸ En 2021, les différents partenaires du Consortium comptent : la Bibliothèque Éric-de-Dampierre du LESC, la Bibliothèque Claude-Lévi-Strauss du LAS, le Phonothèque de la Maison méditerranéenne des Sciences de l'Homme, Bibliothèque de recherches africaines

dépositaires de fonds d'archives et matériaux d'enquêtes collectés par des ethnologues et déposés en fin de carrière, surtout depuis les années 1990 (Molinié et Mouton 2008) :

« Contrairement à une époque antérieure, où seules quelques personnalités connues étaient sollicitées, c'est l'ensemble de la communauté des ethnologues qui s'est sentie concernée par la sauvegarde de ses archives, d'autant plus facilement peut-être que cette incitation rejoignait leurs propres préoccupations. Beaucoup d'ethnologues, en effet, se soucient d'assurer la sauvegarde de leurs matériaux de terrain ». (Op.cit. : 4)

Cet exemple témoigne par ailleurs du fait que la problématisation qui entoure cet enjeu de transmission des "archives des ethnologues" semble largement portée par les communautés d'appui scientifique et technique (documentalistes et personnels ingénieurs) alors que les chercheurs.euses l'abordent davantage à l'échelle individuelle de leur propre dépôt, et sollicitent, dans ce cadre, l'appui de personnels spécialisés dans l'information scientifique et technique. Ainsi, la constitution et l'activité du Consortium AdE autour de ces enjeux a plus fortement mobilisé des personnels ingénieurs et documentalistes, qui se trouvent, de fait, en première ligne pour assurer la gestion pérenne de ce type de fonds. Soulignons que de manière paradoxale, l'usage du terme "archive" est souvent évité par ces mêmes personnels, qui incitent, conseillent et accompagnent les chercheurs.euses dans ce processus (entretiens informels avec deux ingénieurs, février 2021) et gèrent ce type de fonds. Ce, du fait entre autre de la charge juridique associée à la gestion d'archives de leur point de vue, qui implique une caractérisation juridique et des obligations particulières dans le contexte du travail de documentation. Tout le problème résidant dans la distinction épineuse entre le statut public ou privé de ces archives : c'est-à-dire en tant que productions financées par la recherche publique, ou en tant que productions originales et privées d'un chercheur. Cette question n'est pas tranchée à ce jour et reflète sans doute toute l'ambiguïté des matériaux ethnographiques à cet égard. Ainsi, certains gestionnaires préfèrent parler de "ressources documentaires", et contourner cette problématique.

L'enjeu de transmission **ne porte pas en priorité sur des matériaux de recherche en cours d'élaboration**, c'est-à-dire des matériaux issus d'une recherche active (parfois qualifiés de "chauds" comme dans la terminologie utilisée par le TGIR Huma-Num par exemple), tel que cela est préconisé dans le mouvement de l'*Open Research Data* qui incite à publier les "données" au fil de l'eau, dès que cela est possible et avant ou de manière concomitante à la publication des résultats de recherche qui en découlent. La posture du.de la chercheur.euse transmetteur.euse est inverse : il.elle envisage cette forme de partage une fois que ses matériaux de recherche accumulés à l'échelle d'une carrière et d'une succession de projets de recherche sont "refroidis", et qu'il.elle en a déjà tiré une multiplicité de résultats formalisés. Plusieurs motifs sont apparus en filigrane de notre enquête pour justifier le fait que cet enjeu de transmission ne se manifeste pas précocement dans la carrière des praticien.ne.s, et n'est pas relié aux phases actives de recherche. Le principal est la **question du coût**, en particulier en temps, de cette "autre" forme d'attention à la donnée que nécessitent le tri et le traitement pour son archivage en tant que fonds, alors que les préoccupations des praticien.ne.s en activité sont centrées sur la collecte, l'analyse, la publication et la communication de résultats de recherche. Ils.elles sont par ailleurs largement occupé.e.s par d'autres tâches et responsabilités liées à leur activité, comme par exemple la recherche de financements, l'enseignement, l'administration, parfois des responsabilités de gestion de laboratoire, d'évaluation, etc. qui génèrent un ensemble de charges parmi lesquelles, la transmission, ou plus largement la mise à disposition des matériaux n'est pas considérée comme une priorité. Cela, d'autant moins au regard de la représentation du rapport coût-bénéfice dont elle fait l'objet dans ce contexte. Ainsi la fin de carrière, voire la retraite, offrent des conditions plus favorables pour s'adonner à ce type d'opérations. Elle correspond par ailleurs à un moment de la vie du.de la chercheur.euse où la préoccupation de "faire le tri" et de transmettre ce qu'il.elle a à léguer est plus pressante, et où les enjeux de concurrence et d'évaluation sont dépassés. Par ailleurs, paradoxalement,

de l'IMAF, Institut de recherches Asiatiques (IrAsia), Centre de Recherche et de Documentation sur l'Océanie (CREDO), Bibliothèque Yves Le Gallo du CRBC, Centre d'Études Himalayennes (CEH).

l'œuvre du temps ne semble pas atténuer les enjeux et éléments préjudiciables contenus dans les matériaux de recherche, qui pour certains ont été collectés depuis plusieurs décennies, et dont le partage aurait pu faire l'objet de réticences fortes au moment de leur collecte. A partir de l'exemple de corpus photographiques, Marie-Dominique Mouton (2018) souligne que les enjeux qui touchent à la transmission des archives des ethnologues - qui s'organise souvent soit en fin de carrière de l'ethnologue soit après leur mort - demeurent, y compris lorsque les matériaux ont été collectés depuis longtemps. Leur partage en tant qu'archive nécessite des arbitrages fins au cas par cas et une connaissance approfondie des enjeux éthiques qui les relie à leur contexte de production :

« Sans que l'on puisse affirmer qu'il s'agit d'une règle absolue, la divulgation de ces données et de ces portraits, peut, dans certains cas, aller à l'encontre des règles locales, briser des interdits. On citera tout d'abord le tabou de la nudité qui dans de nombreuses sociétés africaines ne doit pas permettre à un descendant de voir la nudité de ses ascendants, les interdits qui réservent aux hommes ou aux femmes la connaissance de certaines cérémonies ou rituels. Mais d'autres informations sur la répartition des terres, la façon dont l'ethnologue fait état d'inimitiés entre groupes, sur la foi de ses informateurs, peuvent également, des décennies plus tard, réanimer des querelles entre des familles ou des villages, envenimer des rivalités entre des groupes, menacer l'équilibre dans la société et entre sociétés. Des détails sur un cliché, difficilement repérables à première vue, peuvent porter préjudice à une famille, en rappelant, par un élément vestimentaire, son appartenance antérieure à un groupe social inférieur, alors que la révélation des pratiques, aujourd'hui décriées, comme celles de sacrifices d'animaux domestiques, pourrait jeter l'opprobre sur un groupe. » (Mouton, 2018 : 79)

A partir de sa propre expérience de transmission d'une vidéothèque constituée au cours de sa carrière et mise en ligne à partir des années 2000-2005 (au tout début des possibilités offertes par Internet), un chercheur honoraire témoigne du fait que **les conditions éthiques et juridiques d'ouverture de matériaux ethnographiques** (ici des vidéos) **ne peuvent jamais être tenues pour acquises**, et qu'elles peuvent évoluer non seulement en fonction du caractère sensible contenu et identifié dans les données au moment de leur collecte, mais également selon l'évolution des contextes et des périodes qu'elles seront amenées à traverser, et qui changent les critères d'appréciation de ce qui est légalement et éthiquement publicisable ou non. Il évoque ainsi, comment toute une catégorie de films ethnographiques qu'il a travaillé à produire lui-même ou à rassembler auprès de collègues pour les inclure dans cette vidéothèque, portant sur l'anthropologie de la santé et l'anthropologie du corps, sont devenus un "enfer", dans un double sens. Enfer, par référence à l'isolement des œuvres iconographiques et textuelles jugées non conformes à la morale dans plusieurs départements de la Bibliothèque nationale de France ainsi intitulés. Enfer, ensuite, par les épreuves incessantes auxquelles la publication de données audiovisuelles sensibles l'oblige à se confronter. Se pose en effet à lui la problématique de transmettre ses propres matériaux, selon ses propres choix, ainsi que ceux qu'il a rassemblés auprès d'autres chercheurs, et qui le positionnent comme acteur intermédiaire dans ce chaînage de transmission ([ENCADRE 6](#)).

Encadré 6. La vidéothèque infernale. Jean, Ethnologue, chercheur honoraire.

(Récit recueilli en en octobre 2020)

« Donc cette vidéothèque de 600 films, **une des raisons qui justifie que je vienne toujours travailler alors que je suis retraité, c'est que j'essaie de transmettre cette vidéothèque.** Mais c'est d'une complexité énorme parce que pour toute une série de gens, la remise d'une vidéo, c'est issu d'une relation personnalisée, au mieux il y a un mail que éventuellement j'ai perdu où il est écrit «OK pour la mise en ligne». Donc quand le centre de ressources a récupéré la vidéothèque, ils ont fait un travail de recherche de tous les auteurs, ils ont éliminé les auteurs qu'ils ne pouvaient pas joindre, et ils leur ont fait signer des documents *ad hoc*. Donc pour transmettre toute la vidéothèque, mon château en Espagne, ce que j'espère qu'il va se passer, y compris avec le centre de ressources qui va être hébergé à la bibliothèque, c'est que ça va faire des postes de consultation informatique, et donc on pourrait mettre accessibles des films de la vidéothèque, où tu as donc : l'« enfer ». C'est-à-dire, pour nous, c'est des films par exemple où on voit des jeunes qui se lavent. C'est ça l'« enfer »: c'est des sortes de visions du corps, des organes génitaux, de seins, chez des jeunes. [...] Je prends par exemple une image qui est un de mes films classiques. On voit un bébé pendant sa toilette. Même si on ne voit pas sa vulve car la maman a sa main devant, **dans l'idéologie pédophile qui est montée en puissance de façon extraordinaire, ça pose problème... Donc se combinent à des logiques de restitutions idéologiques, des logiques qui relèvent de la pudeur... [...]** C'est surtout ça qui nous pose problème. **Au-delà de l'incertitude des consentements** [il a évoqué plus tôt dans l'entretien la difficulté d'obtenir les consentements écrits dans son contexte de travail]. Donc **l'idée c'est de faire un système où, si tu as des légitimités de recherche, thèse, etc., tu peux consulter les films.** Les films de l'« enfer » ne sont accessibles que parce que tu as un sujet de recherche adéquat. Ensuite, il y a le niveau de consultation où les gens [les auteurs] ont dit : "non, je ne veux pas que ce soit en ligne et je veux être sollicité." Et puis, il y a les films pour lesquels les gens sont d'accord. Et là-dessus, va se placer aussi quels sont les films qui sont téléchargeables ou pas. Donc on est en train de refaire une nomenclature par rapport à tout ce qui avait déjà été mis dans la vidéothèque. Et la même chose pour les photos ».

Le récit de ce chercheur pose une question majeure vis-à-vis de cet enjeu de partage, qui dépasse le cadre seul de la transmission : **avec qui envisage-t-on de partager ses matériaux de recherche?** Ce témoignage souligne en effet combien l'enjeu éthique réduit le champ du partage, et suscite un problème de maîtrise de qui a accès aux données, et selon quels motifs. **Le partage ne se définit ici donc pas selon une polarité "ouvert" versus "fermé", mais davantage selon un gradient** (ici à trois niveaux : consultation très contrôlée, consultation sur autorisation, et consultation ouverte en ligne avec possibilité de télécharger la ressource) qui est ici défini par le chercheur producteur en fonction de critères qu'il va participer à déterminer.

Thomas exprime également **la nécessité de gradients dans sa conception du partage** de certains de ses matériaux de recherche, en particulier ses observations (carnets de terrain) et notes d'entretiens. Il ne voit pas de sens à les rendre librement publics sur Internet, alors qu'il conçoit tout à fait leur dépôt comme archive, soulignant en filigrane qu'il considère cette forme de partage comme *en capacité de maintenir un contrôle précis sur les droits d'accès et les motifs de consultation* de ses archives :

« Les documents bruts, ils ne sont pas vraiment utilisables, je trouve. Parce que ce n'est pas ordonné, c'est des choses au jour le jour, mon écriture n'est pas toujours bien lisible. En plus, personne ne me les a jamais demandés. Après pour les mettre à disposition du public, c'est plus compliqué. Parce que là, c'est compliqué. Il y a quand même le fait que c'est manuscrit, après comment on fait pour rendre ça public, on les scanne et on les met sur le net ? Moi je ne vois pas que ça ait du sens. Que ça aille aux archives, ça même de mon vivant, ça oui, ça me paraît évident. De le diffuser sur le net, c'est compliqué. Je ne le sens pas franchement ». (Thomas, professeur d'anthropologie, octobre 2020)

Deux types de publics apparaissent être visés en priorité par les ethnologues qui partagent leurs archives. D'abord un public académique (chercheurs.euses et étudiant.e.s) susceptible de mobiliser ces fonds pour des recherches futures et selon une approche relativement experte (revisite des terrains, apprentissage de la méthode ethnographique, histoire de la discipline par exemple). D'autre part, les communautés ethnographiées, dites "sources", auprès desquelles et avec lesquelles ces matériaux ont été produits. Il n'est pas rare que les archives des anthropologues soient consultées par des ayants-droits ou des membres des communautés étudiées qui recherchent des informations sur leur passé et leur patrimoine⁴⁹. On touche là à une seconde grande dimension du partage des matériaux ethnographiques : le retour vers le terrain, ou la « restitution ».

Retour vers le terrain

S'il est un autre motif de partage évident pour un ethnographe, c'est celui qui s'adresse à ceux.celles qui sont *a minima* considérés comme "informateurs.trices" ou "interlocuteurs.trices" des enquêtes, et sont de fait co-producteurs.trices des données ethnographiques: c'est-à-dire les membres des groupes ethnographiés sur le terrain. Le "retour" vers le terrain, ou la "restitution" à ces communautés sources constitue ainsi une forme de partage embarquée dans la production des données ethnographiques. Les évolutions liées aux technologies numériques ont grandement facilité ce type de partage et sa logistique : plutôt que d'envoyer ou de transporter des valises de photos, bandes sonores ou autres supports de partage sur le terrain, les documents numérisés peuvent, depuis une vingtaine d'années environ, être rendus d'autant plus facilement accessibles aux communautés dites "sources" que leur circulation numérique s'est accrue - sous réserve que ces communautés disposent d'accès à des outils informatiques et à Internet. Quoi qu'il en soit, leur numérisation et leur stockage informatique facilitent la logistique du "retour" au terrain. Ceci est apparu comme une formidable opportunité pour Jean, qui a produit des films ethnographiques tout au long de sa carrière, avec l'intention de les partager sur le terrain :

*« A partir des années 2000-2003, quand est apparue la possibilité de mettre en ligne des films, moi j'ai trouvé que c'était quelque chose d'absolument épatant. **Une vraie restitution, une vraie rentrée dans le patrimoine immatériel de l'humanité**, et en même temps je ne peux pas vraiment te montrer de document avéré, prouvant par A+B que l'ensemble des gens que j'ai filmés ont exprimé leur consentement. Dans tous mes films Ebola, j'ai monté les films, j'ai montré sur place et diffusé, parce que l'idée c'était de passer dans les ethnies, parce que l'idée c'était de monter ce qu'il se passait dans les pavillons d'isolement où personne n'allait et on imaginait des choses épouvantables. Mais c'était évident que je n'ai pas pu, pendant l'épidémie, avoir les contacts qu'il fallait avec les familles pour leur [demander leur accord]... Enfin dans certains cas si. La restitution, tu peux la faire avec des assistants informateurs ou la famille dans laquelle tu restes. Plus, c'est plus compliqué ». (Jean, chercheur honoraire, octobre 2020)*

La vidéo, si elle suppose un processus éditorial, s'avère toutefois un format privilégié par plusieurs chercheurs.euses pour partager des matériaux ethnographiques au-delà des sphères académiques, et en particulier auprès des populations étudiées:

« Les vidéos je les fais déjà un peu pour moi, pour les gens sur place pour après retourner, montrer les vidéos aux gens qui sont dans les vidéos, et pour partager aux amis, à la famille, au réseau. Mais c'est le seul matériel brut en fait que je partage, qui sont en fait des petit formats vidéo de terrain sur des pratiques de chasse, de cueillette, la vie au village. Je les mets sur canal U. Et donc à chaque vidéo il y a un grand texte, avec des mots clés, pour que ce soit référencé, etc. [...] Je pense que c'est l'envie de partager quelque

⁴⁹ Entretien informel avec un gestionnaire d'archives, février 2021

chose que j'ai vécu sur le terrain ou que j'ai vu sur le terrain, avec une communauté plus large que le petit monde académique ». (Pierre, chercheur non titulaire, octobre 2020)

Le partage en tant que retour vers celles et ceux qui ont participé à l'enquête ethnographique apparaît comme une évidence pour les chercheurs.euses. La question qui se pose à eux.elles dans cette intention est **celle de trouver la forme et le niveau juste pour cette restitution** : c'est-à-dire celle d'un matériau qui soit appropriable et fasse sens pour les communautés étudiées. La publication académique apparaît comme un format peu approprié dans ce contexte, alors que les éléments visuels et sonores (photographies, vidéos, enregistrements sonores de chants, etc.) se prêtent davantage à cet usage. Dans certains cas, des formats plus élaborés et emprunts d'une forme de créativité peuvent aussi faire l'objet d'une implication particulière de la part des ethnographes : des émissions de radio par exemple (citée par un interlocuteur) ou des expositions, notamment photographiques. Les enjeux éthiques sont à nouveau une dimension majeure du retour : au-delà de l'engagement moral qui lie l'ethnologue à son terrain dans une relation de don/contre-don et de co-production, il s'agit aussi de choisir parmi les matériaux restitués **ceux dont les contenus ne risquent pas de porter préjudice ou de créer des tensions auprès des groupes étudiés**. Un exemple très parlant est celui d'un chercheur qui documente des pratiques de braconnage, et ne peut évidemment pas faire porter l'objet de ses retours vers ce type de contenus. Cette dimension touche également à la **question de la relation de l'ethnologue à son terrain** : à la fois cette relation est, dans la plupart des cas, conçue comme faite d'allers-retours (et non pas d'une visite unique), qui engagent l'ethnologue envers ses interlocuteurs.trices. Et d'autre part, la qualité de ces "retours", à la fois en tant que revisite et en tant que ce qui est rapporté, ou restitué d'une mission à l'autre, peut influencer la qualité de l'accueil du.de la chercheur.euse sur le terrain, et donc la possibilité de pouvoir y poursuivre ses recherches : outre l'impératif éthique, partager des contenus préjudiciables constitue donc un risque de fermeture du terrain à un.e chercheur.euse.

Autres formes de partages et de création : formes hétérogènes et publics élargis

Au même titre que les matériaux ethnographiques n'apparaissent pas comme des contenus standardisés, les formes et niveaux de partage dont ils font l'objet revêtent un caractère artisanal et singulier à chaque situation rencontrée. On l'a vu apparaître dans les deux principaux motifs de partage que représentent la transmission et la restitution, qui se construisent et s'organisent autour d'une forme de singularité de chaque situation. On le voit également apparaître dans d'autres formes de partages, qui se manifestent selon les configurations de recherche dans lesquelles les ethnographes travaillent, et selon leurs affinités individuelles et leur appétence pour des formes de diffusion et de création autres que la publication académique classique. Sans pouvoir évoquer ici de manière exhaustive leur étendue et leur diversité, cette partie en souligne quelques exemples, qui se construisent largement autour du projet **d'élargir (et de diversifier) à la fois les formes de partage et les publics visés, par contraste avec une démarche et une audience principalement académiques**.

Plusieurs chercheurs.euses interrogés travaillent dans une configuration où la **gestion de collections muséales** est une composante de leur activité ou de leur contexte de travail. Ils sont ainsi particulièrement sensibilisés à l'enjeu de présenter des matériaux ethnographiques dans le contexte d'**expositions** destinées au grand public. Cette forme de partage s'appuie, dans les cas décrits, principalement sur des artefacts, des vidéos, des photographies ou des bandes sonores, souvent reliés entre eux pour donner un sens à cette publicisation de contenus, comme le souligne Thomas, très impliqué dans la collecte et l'exposition d'objets, et qui explique ses motivations à la fois à collecter et à rendre publiques des collections :

« J'ai ramené des objets depuis mon premier voyage auprès [d'une communauté] en 1972. C'est quelque chose auquel j'ai toujours tenu. Bon j'aime ça. Et effectivement je collecte, c'est même maniaque. Mais actuellement ça me pose des problèmes, parce que dans l'état d'abandon, pas d'abandon mais dans l'état matériel où se trouve la réserve, on ne peut plus y rentrer, tous les trucs sont entassés, c'est scandaleux, et c'est pareil pour les autres collections même patrimoniales : si on leur propose deux paniers, ils ne vont en prendre qu'un, et s'il est un peu gros, on ne sait pas où on va le mettre. Donc je me pose la question,

*quand je vais en mission, à quoi ça sert que je collecte? Parce que je ne sais pas ce que cette collection va devenir. Donc j'en suis au point de me demander : qu'est-ce que je vais en faire? Je ne vais pas faire mon musée à moi! Ça devient critique. Mais moi j'aime ça, ça me plaît. Il y a un côté attrait personnel et il y a un côté plus réfléchi, parce que je ne collecte pas tout, donc il y a un lien avec l'artisanat, l'alimentation, ou le travail avec la nature. C'est l'angle que je prends. **Et la motivation c'est toujours de dire : ces objets c'est un témoin d'aujourd'hui, d'un état. [...]** Donc la motivation, **c'est l'objet témoin, témoin d'une particularité locale et d'aujourd'hui quoi. C'est ethnographique.** Mais il y a peu de gens qui collectent, c'est vraiment un attrait. Donc l'objet n'est pas vraiment un support de recherche, c'est en plus. [...] **Et la motivation à les rendre public, c'est pour les faire exister, sinon ça n'a pas de sens.** En plus c'est pas de choses qui sont jolies, quelques-unes sont jolies mais la plupart, c'est anodin. **Donc elles n'ont du sens qu'ensemble. C'est-à-dire qu'on crée du sens en les associant. Donc c'est un peu ça, c'est relié à la technique, à l'action sur la matière, et ça marche par ensemble, donc de les associer pour montrer la diversité et la variabilité** ». (Thomas, professeur d'anthropologie, octobre 2020)*

Si ce chercheur relie largement sa démarche à une appétence personnelle, nous constatons de manière assez nette au sein de notre échantillon d'entretiens l'influence des configurations de travail sur la sensibilité du/de la praticien.ne à collecter des éléments susceptibles d'alimenter des projets de partage, tournés vers l'exposition et le "grand public", quelle que soit la génération ou le statut professionnel du/de la praticien.ne. Quatre des cinq interlocuteurs.trices relevant de la catégorie "Collectifs de recherche en anthropologie avec une forte culture de l'interdisciplinarité avec des sciences biophysiques et un contexte de gestion de collections muséale" ont ainsi abondé des collections de recherche, et continuent à le faire. Ils se décrivent à la fois comme sensibilisés par ce contexte professionnel dans lequel ils évoluent, et considèrent dans le même temps ces contributions comme partie intégrante de leur mission scientifique de diffusion des connaissances. Ils incluent donc pleinement cet enjeu dans leurs pratiques d'enquête. Thomas, dont le témoignage est cité ci-dessus, est en charge de gérer ce type de collections à l'échelle de son établissement. Il explique la difficulté qu'il rencontre à **valoriser et rendre visibles ce type de collections**, qui nécessitent un effort soutenu de la part des praticien.ne.s porteurs.euses de ce type de projets, et également la mise en place de collaborations avec d'autres corps de métiers. Il raconte ainsi l'un de ses projets les plus récents de valorisation virtuelle, dans le cadre d'une collaboration avec une école de *design*, sur fonds de recherche de nouveaux formats de publicisation de ces collections qui n'ont de sens, selon lui, que si elles sont rendues visibles :

« A propos de l'utilisation des collections, j'ai eu quelque chose d'absolument excitant les deux dernières semaines : j'ai été amené à travailler avec des étudiants d'une école de design. Donc avec leur prof, on a fait l'atelier ici. Je leur ai montré la réserve, et je leur ai sorti des objets qu'ils ont pu regarder. Et puis ils ont imaginé des projets à partir de la réserve, et puis ils ont imaginé tout ce qu'ils voulaient, ils pouvaient prendre des photos, etc. Puis pendant la deuxième semaine, ils ont fait leurs prototypes, qu'ils nous ont exposés. Donc le titre de l'atelier, ça s'appelait "donner à voir". C'était sous-entendu, comment on fait pour rendre visibles des objets qui ne servent à rien dans la réserve. Et il y a même une personne qui a appelé son projet « la réserve invisible ». Ça montre qu'ils ont compris l'état d'esprit. Donc ils ont fait plusieurs propositions. [...] L'un, c'était de faire des entretiens avec plusieurs collecteurs, autour des objets qu'ils ont collectés, qu'ils en parlent pour faire des podcasts. Bon après il y a des trucs plus classiques d'exposition, mais plusieurs présentations numériques des objets, en les mettant en relation les uns avec les autres. Donc tout ça je ne sais pas ce que je pourrais utiliser, mais le concept il est excellent, et ça répond au souci que j'ai de valoriser ces collections ». (Thomas, octobre 2020)

L'exposition, physique ou virtuelle, génère ainsi un **processus de création**, qui, dans les expériences décrites, est accompagné de nouvelles collaborations : dans les cas plus classiques, il s'agit de travailler avec une équipe muséographique et scénographique. Cela suscite une série de négociations et d'arbitrages sur ce qui

peut être rendu public dans ce cadre, sous quelle forme, pour délivrer quels messages et informations. Dans cette configuration, des tiers incorporent la relation du/de la chercheur.euse aux matériaux de terrain rendus public et au terrain lui-même. Dans d'autres cas, il s'agit de **collaborations plus artistiques**, liées à des démarches de recherche-création par exemple. Dans ce contexte, Emmanuel raconte qu'il a participé à créer des "pastilles sonores" en collaboration avec une radio :

"J'ai des formats un peu étranges, c'est des pastilles sonores, avec une réalisatrice radio de RFI. On avait fait des montages sonores, il y avait juste une phrase d'accroche, et après c'est des ambiances sonores et des voix [enregistrées]. Et ça rendait compte d'une partie de mon travail en fait. Donc là c'est une forme, entre artistique et anthropologique, et sans... que sonore". (Emmanuel, chercheur titulaire, octobre 2020)

Cette dimension de création artistique constitue pour un autre chercheur, Pierre, une voie (et une voix) particulièrement favorable à l'exploration d'autres formes de mises en partage de l'ethnographie, qui peut se construire au gré de la recherche et l'accompagner. Il décrit ainsi comment la vidéo lui offre des possibilités pour assumer une dimension plus créative et personnelle dans son travail. Ce format lui permet à la fois de rendre compte de manière très fidèle des observations ethnographiques qu'il réalise, tout en assumant parfaitement son rôle sur le terrain et son regard particulier sur la situation observée. C'est surtout pour lui un moyen de partager sa recherche plus largement que par les canaux académiques, et en particulier auprès de publics qui font sens pour lui :

*« Ce que j'aime dans la vidéo, c'est qu'il y a aussi une part plus personnelle et créative, ou quand tu filmes, tu assumes un point de vue, et que ce soit au cadrage ou au montage, tu assumes un point de vue, une interprétation. Et c'est ça qui me plaît en fait : **le fait d'être dans un espace un peu intermédiaire entre la création et en même temps, c'est de l'ethnographie**, qui suit par exemple une chaîne opératoire pour fabriquer une arme de chasse, ou qui suit un groupe en forêt. **Je suis une activité du début à la fin, donc c'est de l'ethnographie filmée, et en même temps j'y mets ma patte plus facilement que dans un article scientifique pour une revue à comité de relecture, où la possibilité de personnaliser ton rapport au terrain est plus complexe.** Donc moi j'ai besoin de ça. De dire que je ne suis pas juste un collecteur de données qui va pêcher les données, les mettre en forme et les publier. **Il y a aussi un rapport de moi au terrain que j'aimerais mettre en avant.** C'est ça que j'aime bien dans la vidéo ethnographique, que ce soit visible que tu interagis avec les gens. Et après c'est **une volonté de partager avec mon réseau d'amis, de contacts, ou de chercheurs, de connaissances dans le monde académique qui ne vont pas forcément travailler sur les mêmes sujets et lire les articles que j'écris, de partager en fait ce qui se passe sur mon terrain**". (Pierre, chercheur non titulaire, octobre 2020)*

Entrepôt de données

Parmi nos interlocuteurs.trices, seul Emmanuel témoigne avoir déjà eu une expérience de dépôt de données ethnographiques sur un entrepôt dédié. Son récit met en lumière son adéquation assez forte avec les valeurs qui soutiennent l'Open Research Data, alors que ce chercheur se dit par ailleurs un "fêru" de l'Open Access: il dépose depuis plusieurs années toutes ses publications sur HAL, et a fait oeuvre de sensibilisation auprès des collègues de son laboratoire pour en faire de même avant que cette pratique ne devienne une règle. Ce qui est intéressant dans son témoignage est la nuance qu'il apporte du point de vue de son rapport à cette forme de partage ouvert, en fonction des configurations : à la fois sur le type de données concernées, le type de situations et le sujet de l'enquête. Il explique ainsi en début d'entretien qu'il lui paraît impossible de partager des matériaux d'enquêtes tels que ses carnets de terrains ou ses observations, vis-à-vis du risque que cela ferait encourir à des interlocuteurs.trices, du fait que ces matériaux sont trop empreints de son regard et donc interprétables par lui uniquement, et qu'ils ne font de sens que combinés entre eux. Plus tard, il raconte comment est récemment née, dans le cadre d'une collaboration avec un autre ethnologue, sa première expérience de dépôt de données sur entrepôts. Ils ont à la fois **considéré que ces données**

n'étaient pas sensibles et en ont retiré les éléments identifiants, et surtout qu'ils avaient **achevé leurs propres analyses**, mais que d'autres analyses pouvaient encore être faites (ENCADRE 7).

Encadré 7. Le dépôt sur un entrepôt de données par Emmanuel, chercheur titulaire.

(Récit recueilli en octobre 2020)

« Il n'y a pas longtemps j'ai fait un article sur un projet qu'on avait en commun avec un collègue, sur la perception des ambiances sonores. Donc comme c'était un point final, voilà on n'avait pas envie de continuer dessus. Mais on avait conscience qu'une partie de nos données on ne pouvait plus, on était tous les deux partis sur autre chose, et il faut savoir s'arrêter. Du coup on a décidé de mettre, on a mis dans l'article qui est publié, on a marqué un petit mot en disant que toutes nos données étaient intégralement en ligne. **On les a déposées sur Zenodo, c'est un open access européen, et là on a mis toutes nos données, y compris audio.** On avait demandé à des habitants d'enregistrer un trajet, par exemple entre chez eux et le café, et après il y a des entretiens sur l'ambiance sonore du trajet. **Et on a aussi fait une transcription et une traduction de ces entretiens. Et on a mis tout ça en ligne. Données audio, textuelles, et aussi lexiques, etc.** Donc là pour le coup on a tout mis en ligne. On voulait les donner à la revue mais ils ne savaient pas quoi en faire, ça ne les intéressait pas en fait.

[Pourquoi dans ce cas spécifique vous avez voulu les partager ?]

Parce qu'on avait le sentiment qu'on les avait sous-exploitées. D'une part parce que c'était quand même pas mal, même si on avait continué de travailler là-dessus, **c'était pas mal que les données accompagnent l'article. C'était aiguillonné par le fait qu'on allait arrêter de travailler sur ces données.** Mais que les données accompagnent l'article c'était pas mal aussi : en termes de vérification, si les gens voulaient retourner aux données pour vérifier ce qu'on disait, faire des analyses concurrentes, ou d'autres analyses, il y a plein de choses. **On peut imaginer très facilement les linguistes s'intéresser à ce corpus.** Il est très intéressant, parce qu'on entend des gens qui parlent dans leur langue naturelle, des ambiances sonores qui accompagnent leur vie quotidienne. Des gens de tous milieux, de tous genres et toutes classes sociales, et de toutes localisations dans [la ville concernée]. Et par ailleurs, en plus, ce qui nous a donné l'idée, peut-être pas ce qui nous a décidé, **mais ce qui a donné l'idée de mettre en ligne c'est que nous on n'allait plus s'en servir.** En tous cas c'est en ligne et ça nous convient très bien, on est contents de ça. D'autant que ça existe pas beaucoup en SHS ce genre de démarche, à notre avis. [...] **Ces données-là, on les considère pas sensibles.** Par ailleurs on a anonymisé les gens, parce que certains interlocuteurs préféraient ne pas voir leurs noms apparaître, du coup on a anonymisé tout le monde. On a mis des pseudonymes, et parfois dans les entretiens il y avait le nom qui apparaissait, donc on les a masqué avec un bip sur les fichiers sonores. On est allés jusqu'au bout de la logique ».

Ce chercheur souligne par ailleurs qu'il a déjà eu ce type d'expériences, mais concernant des données issues des sciences biophysiques, c'est-à-dire dans le cadre de collaborations interdisciplinaires avec des domaines scientifiques pour lesquels la pratique est plus courante et les données sont nativement plus standardisées:

« Dans des articles où j'ai collaboré avec une collègue généticienne, là souvent les revues de sciences dures demandent les données, et puis nous on le fait volontiers de toutes façons. On dépose les données génétiques dans une banque. Donc toutes ces données-là sont récupérables et vérifiables par d'autres chercheurs ». (Emmanuel, chercheur titulaire, octobre 2020)

Isabelle, enseignante-chercheuse, souligne cette distinction entre données standardisées et considérées comme "objectivées", et matériaux qualitatifs qui incorporent une perspective subjectiviste. S'étant intéressée aux types d'initiatives existantes de bancarisation de matériaux de recherche dans des bases de données, elle souligne que seules les expériences centrées sur des types de matériaux très particuliers et **objectivables** lui paraissent avoir du sens, comme des taxonomies vernaculaires par exemple. Outre le fait que les matériaux qualitatifs se prêtent mal à ce type d'exercice, selon elle ce type de collectivisation des matériaux de recherche est très éloigné des pratiques qu'elle maîtrise : elle ne saurait comment s'y prendre

pour contribuer à un projet de ce genre et ne conçoit pas que le travail sur ce type de données puisse être réalisé en dehors de celui.celle qui les a produites.

*« A chaque fois que je vois un peu ce qui se fait comme bases de données, je trouve que celle qui était vraiment bien faite, c'était celle du projet LINKS, parce que justement on avait la possibilité d'injecter des données, mais sans refermer leur sens. Alors que là, il n'y a pas de constitution faite dans ce sens-là. J'avais par exemple construit un site, avec des données taxonomiques, sur comment les coraux étaient nommés par exemple. Mais j'en étais un peu restée là parce que justement je n'ai pas trop d'expérience de ce genre de choses. **Autant dans un article tu sais comment citer des données, etc. Autant dans une base de données.... Je ne savais pas trop comment les valoriser autrement que dans des articles.** A part sur des éléments sur la classification des espèces par exemple, mais après sur des enjeux politiques c'est très difficile de faire des bases de données, tu ne peux pas. Non moi je ne serais pas du tout favorable. Aussi parce que, je reste persuadée qu'on mène des enquêtes, ce n'est pas des données brutes, c'est des données qu'on a recueillies, qu'on a comprises avec notre singularité, notre complexité. **A part certaines données très objectives comme par exemple le nom de certains poissons, mais sinon hors de ce cadre, je ne vois pas trop, et je serais très réservée. Je sais pourtant que c'est une pratique qui s'est faite dès le début de l'anthropologie et qui continue à se faire dans certains pays comme au Québec par exemple, de déléguer par exemple le terrain à des étudiants et après de reprendre ses données et d'en faire..., mais moi ça me pose beaucoup de problèmes ce genre de démarches. Je pense qu'on n'est pas attentif aux mêmes choses et que les questionnements d'un tiers ne sont pas.... Ça me paraît difficile de penser les données comme des matières objectives. **Dès qu'on va sur des terrains qui sont plus subjectifs, ou plus qualitatifs, c'est vrai que pour moi les bases de données n'ont pas grand sens** ».***
(Isabelle, enseignante-chercheuse titulaire, octobre 2020)

Partages entre pairs

Si la mise à disposition des matériaux de recherche de manière libre et détachée de son.sa producteur.trice apparaît minoritaire et ne fait sens que dans des cas très spécifiques, le cadre de la collaboration professionnelle avec d'autres chercheurs.euses est un motif de partage légitime pour les personnes interrogées. Ce partage s'opère dans le cadre d'**un accès très contrôlé, restreint à des collègues proches, et où le.la producteur.trice de la donnée reste toujours impliqué.** Ces partages sont décrits comme accompagnés d'échanges, notamment parce qu'ils se produisent dans et au service des phases d'élaboration des interprétations et des analyses fondées sur les matériaux d'enquête. Nous avons recensé plusieurs types de pratiques. Par exemple : une chercheuse met à disposition de chercheurs.euses avec qui elle collabore tous ses entretiens retranscrits, dans le but spécifique de discuter et d'échanger sur leurs contenus, en croisant les diverses interprétations fondées sur ce même contenu. Elle explique que les entretiens ne sont pas anonymisés, puisque l'objet de cet échange est justement de saisir qui parle et quelles sont les particularités de son discours. Un autre chercheur ne met en partage que ses comptes-rendus d'entretiens, c'est-à-dire un document de travail qui incorpore déjà un niveau avancé d'interprétation et de cadrage. Une troisième chercheuse explique de son côté qu'elle ne met pas en partage les documents eux-mêmes (entretiens et observations), mais elle partage sur le fond, à l'oral, avec des collègues les informations qu'elle a recueillies. Un autre type de partage évoqué à plusieurs reprises, en particulier par les enseignant.e.s-chercheurs.euses est celui de la mise à disposition, par des étudiants qu'ils.elles encadrent, des transcriptions d'entretiens que ces dernier.e.s ont réalisés dans le cadre de stages ou de programmes de recherche pour lesquels ils.elles ont été recruté.e.s. Outre la dimension pédagogique, ce partage à une vocation de collectivisation des données pour élaborer une analyse avec les chercheurs.euses référent.e.s.

Isabelle témoigne ainsi de ce type de mise en partage organisée au sein d'un projet collectif et comparatif, qui implique également des chercheurs.euses post-doctorants et qui vise à documenter et comparer une

enquête multi-située. Elle décrit comment des niveaux divers d'élaborations de ces matériaux ethnographiques sont mis en commun pour alimenter une analyse transversale :

« J'avais demandé à chaque personne qui s'était rendue sur un terrain de me rendre un fascicule qu'on avait construit ensemble et qu'on puisse interagir ensemble et poser des questions sur des éléments plus précis. Et effectivement j'avais regroupé toutes les données que j'avais pu recueillir, et j'avais essayé de les synthétiser, mais ça c'est encore un travail qui reste à faire. On travaillait tous sur des terrains relativement différents, parce que pour avoir une idée générale sur plusieurs [territoires], il fallait nécessairement que j'envoie des gens où moi je ne pouvais pas aller. Donc j'étais partie en mission à [un endroit], et d'autres étudiants et post-docs sont partis à [d'autres endroits], et on comblait comme ça tous les espaces importants. Et puis on reparlait par des réunions régulières, et puis par les données, par ces fascicules qui pouvaient être très substantiels, même s'ils ne l'étaient pas toujours. [...] Les étudiants mettaient en partage les annexes de leurs mémoires, qui étaient les entretiens, et en même temps, chacun devait constituer un stock de données brutes, même si les données brutes ça n'existe pas, mais que je pouvais consulter et sur la base desquelles je pouvais faire moi cette lecture comparative. [...] Et les post-docs, ils travaillaient sur des questions vraiment très particulières, et là ils essayaient de rédiger quelque chose, sur des points beaucoup plus précis autour de ce questionnement global ». (Isabelle, enseignante-chercheuse titulaire, octobre 2020)

Cette collectivisation du travail sur les données, qui impulse leur mise en partage, est plus particulièrement **anticipée, organisée et développée** dans le cadre de **programmes de recherche financés**, qui participent à fédérer des équipes autour d'un objet de recherche commun avec des moyens dédiés. Les témoignages recueillis dans l'enquête montrent toutefois que cette collectivisation tend à se restreindre aux chercheurs travaillant sur des **types de matériaux similaires ou relevant de domaines disciplinaires proches**, quand bien même ils.elles participent à des collectifs élargis et pluridisciplinaires. Le récit de Pierre, jeune chercheur recruté dans un programme interdisciplinaire impliquant des chercheurs.euses en sciences de la santé et en sciences biophysiques, décrit parfaitement cette organisation à deux niveaux : d'une part un partage et un travail de croisement des interprétations qui se jouent à l'échelle de la donnée, et d'autre part, un travail d'analyse interdisciplinaire qui se joue à l'échelle générique du partage d'analyses intermédiaires ou finalisées. Il revient par ailleurs sur la nécessité d'organiser collectivement le partage des données et les préparer de sorte qu'elles puissent être consultées et mobilisables par celles.ceux qui ne les ont pas produites.

Encadré 8. La collectivisation des données dans le programme de recherche auquel participe Pierre, ethnologue, chercheur non titulaire (récit recueilli en octobre 2020).

« C'est encore à construire. Pour l'instant il y a deux pôles dans le projet : il y a ceux qui vont faire de la métagénomie, de l'analyse du microbiote intestinal, du séquençage ADN, etc., qui n'ont pas l'habitude de lire de l'anthropologie et du qualitatif. Donc là vraiment il y aura un échange à faire, mais là nos interprétations ne vont pas influencer la manière dont ils vont traiter les données. Et un peu *vice versa*, nous on collecte des informations sans avoir les résultats des séquençages ADN par exemple. C'est pas qu'on s'en fout, mais il y a plus d'échanges entre ma directrice qui est historienne anthropologue, et un autre chercheur qui est aussi anthropologue. **Tout notre groupe où on a l'habitude de faire du terrain, de collecter, d'échanger avec les gens. Donc il y a plus d'échanges de données, d'interprétations de réflexions, parmi les scientifiques qui travaillent en SHS.** Et le partage d'informations et de réflexion est peut être plus complexe avec ceux qui sont vraiment sur de la virologie, de la génétique.

[...] Tout est en commun dans un dossier partagé : toutes les retranscriptions d'entretiens, toutes les grilles d'analyses, les codages NVivo, tout ça c'est en commun. Et après on en discute souvent : notre petit groupe, on est 3 anthropologues, on fait des réunions une fois par semaine. On échange en fait les questions qu'on se pose, des débuts de résultats. A la fois les données sont en accès libre entre nous, et en même temps on en discute assez régulièrement. Donc il y a des échanges de données et de l'échange de la formation aux outils d'analyse qui sont différents, entre R et NVivo par exemple.

[...] Dans les dossiers informatiques qu'on a en commun, il y a les enregistrements bruts, qui sont codifiés : le nom du fichier il est codé. On retrouve le code dans la retranscription d'entretiens entière avec un résumé de l'entretien. On essaie de faire au maximum un résumé de l'entretien pour avoir accès aux grandes lignes. Les documents NVivo de codification et les fichiers Excel d'analyses plus quantitatives. En fait, tout est mis en commun. Et les observations, moi j'ai des documents word, que j'essaie de mettre au propre au fur et à mesure, qui s'appellent notes de terrain, avec date, lieu, et certaines observations pertinentes qui sont retranscrites, parfois un peu augmentées avec des questions, des interprétations, etc., et ça je les mets en commun au fur et à mesure. Après les photos je ne les mobilise pas trop dans l'analyse, elles sont là pour servir d'illustration à certains moments, et je ne les mets pas trop dans le dossier car ça prend de la place. Et les vidéos, c'est l'idée oui, que je les mette en commun. Mais là il y a déjà un travail de montage, et le montage c'est une forme d'analyse parce qu'on fait des choix dans le montage, de ce qu'on veut montrer.

Et là par exemple je suis en train de travailler sur les chiens, et sur les relations sur les humains et les chiens dans les villages, et prendre le chien comme un sujet intermédiaire entre le village et la forêt et comme quoi c'est intéressant d'analyser le chien comme transmetteur de pathogènes. Et moi j'ai beaucoup de vidéos là-dessus, donc là je suis en train de travailler sur un montage vidéo qui pourrait être utilisé comme un support, comme une analyse illustrative. En fait j'appelle ça de l'observation participante filmée : en fait je suis sur le terrain, je suis une activité de chasse, et plutôt que de prendre des notes dans mon carnet pendant la journée où je suis les chasseurs et leurs chiens en forêt, je filme. Et ces rushes-là sont en fait mon carnet de terrain, et sont encore plus riches. Parce que là pour le coup c'est mon regard qui enregistre tout, et après coup, je peux analyser les matériaux vidéos quoi. **Et là pour l'échange avec les autres chercheurs, je trouve ça hyper riche parce que c'est beaucoup plus complet qu'une prise de notes. Et c'est surtout que même s'il y a un début d'analyse en soi, un virologue, ou un primatologue qui voit une vidéo où je filme de relations humains animaux, va y aller de sa propre interprétation, et va peut-être voir des choses que moi je n'ai pas vu dans cette vidéo.** Donc moi je trouve que c'est un matériel brut qui est pas mal aussi pour échanger, avec d'autres disciplines ou d'autres chercheurs.

[...] **Ça me force à mieux organiser les fichiers, les documents, pour être sûr que ce soit bien accessible, compris par d'autres.** Mais c'est vrai que je porte une attention plus grande au fait que d'autres personnes vont aller regarder ces données. Donc j'essaie que les résumés soient clairs. C'est devenu un automatisme ».

Ces différentes pratiques de partages de matériaux de recherche entre pairs mettent finalement en évidence deux dimensions majeures :

- d'abord, elles témoignent d'une **collectivisation accrue du travail interprétatif et de l'élaboration d'analyses fondées sur les matériaux empiriques** dans le cadre de la recherche financée sur projets, qui favorise ce type de coopération.
- Ensuite, elle souligne **l'importance, pour les producteurs.trices de données, d'accompagner cette mise en partage marquée par un double impératif de confiance envers les collègues avec qui les matériaux sont partagés, et de contrôle** sur le périmètre et les modalités du partage.

Pas de partage (public) possible

Enfin, une diversité de configurations et de raisons ont été évoquées par les personnes interrogées pour décrire des situations où **le partage de matériaux de recherche ne leur apparaissait pas possible, en particulier dans une configuration ouverte et libre de réutilisations**. Nous avons évoqué dans la partie 4.3.2 la diversité des liens (visibles et invisibles) qui unissent les chercheurs.euses à leurs matériaux de recherche (relation de confiance, subjectivité du regard du chercheur, intimité, caractère entremêlé et évolutif des matériaux) et qui les rendent difficilement détachables, en particulier en tant qu'unités élémentaires indépendantes.

Il apparaît dans certains cas que, quand bien même le praticien considère qu'il existe un intérêt fort à partager des matériaux de recherche, d'autres enjeux peuvent rendre ce partage impossible. Le récit de Thomas (ENCADRE 9) illustre ainsi comment le **problème du recueil des consentements** l'empêche aujourd'hui de partager des corpus photographiques collectés à une période où cet enjeu juridique n'apparaissait pas comme déterminant. Il explique pourtant que de fait, prendre des photos de personnes supposait nécessairement qu'il ait leur accord, mais sans pouvoir aujourd'hui en fournir la preuve. Par ailleurs, la mise à disposition libre de ses photos lui apparaît risquée du point de vue de réutilisations non maîtrisées.

Encadré 9. Une éthique de la publication d'images. Thomas, Professeur d'Anthropologie (récit recueilli en octobre 2020)

« Les photos je pense qu'elles sont faites pour servir. Il faut trier. Mais il y a un autre problème qui est plus actuel, c'est le problème de la propriété de l'image, par les gens photographiés. Par exemple l'IRD ils font parfois des appels pour publier des photos, où ils demandent l'autorisation des personnes prises en photo. Je n'ai aucune autorisation d'aucune personne que j'ai prise en photo. Pour plein de raisons. La première c'est qu'à **une époque on n'avait même pas idée de le faire**. La deuxième raison c'est qu'en Afrique les gens sont très sensibles aux photographies, et si on leur demande un papier signé ils vont me dire, "Ah mais tu me payes, parce que toi tu vas être payé pour ça". **Donc ça met tout de suite de la suspicion**. Donc ma seule réponse, c'est que quand je prends une photo, je prends peu de photos de personnes, de gens... En fait j'ai été éduqué à ça justement par rapport à ce que je viens de dire : en République Centrafricaine, tu prends une photo dans la rue, t'as immédiatement la police qui te tombe dessus en te disant que tu es un espion, et la personne qui est au fond de la rue que tu n'avais pas vue vient vers toi en te disant "tu m'as pris en photo, paye-moi". **Donc j'ai été éduqué à ça. Donc je ne prends des photos de personnes que quand je suis chez eux et que je les connais**. J'ai très peu de photos avec des gens sur ma photo. C'est pour ça que j'ai plein de photos d'étals de marchés et très peu de photos des personnes qui vendent. Donc c'est un réflexe, ce qui est absurde, parce que quand tu vas en Asie ou en Amérique centrale, les gens te demandent de les prendre en photo, c'est l'inverse. Ça veut dire que les photos avec des gens que j'ai, c'est quand j'ai habité chez les gens. Donc là je considère que l'autorisation de les prendre en photo je l'ai, sinon ils m'auraient mis dehors. Après il y a aussi quand les gens te regardent, mais comme j'ai beaucoup de photos techniques, ils sont en train de faire quelque chose, donc ils ne me regardent pas. **Donc je considère que l'autorisation de les prendre en photo je l'ai, par contre officiellement je ne l'ai pas, et donc je ne peux pas faire état de ça à la publication. Alors les mettre en ligne, ça pose exactement cette questions-là : est-ce que j'ai le droit de mettre le visage reconnaissable de quelqu'un à disposition de l'humanité entière**. Sachant en plus que les photos elles sont volées pour faire n'importe quoi d'autre, je ne parle même pas d'argent. Tu peux la

retrouver sur une affiche pour je ne sais quoi. Les banques d'images, c'est catastrophique en fait. T'as aucune maîtrise dessus ».

Son témoignage se poursuit par l'évocation d'une autre dimension problématique du partage de ces photos devenues "historiques", vis-à-vis du regard actuel des membres de la communauté photographiée. Thomas raconte comment, il y a quelques mois, il a reçu à Paris avec d'autres collègues des membres du groupe auprès duquel il avait travaillé dans les années 1980. Entre autres visites et moments de partages ayant eu lieu à cette occasion, une anecdote malheureuse l'a profondément marqué : alors qu'un chercheur du groupe montrait ses photos de terrain de l'époque, il explique la réaction d'une femme:

« Elle a hurlé en nous disant: c'est quoi ces horreurs, cachez-nous ça. Je ne veux pas voir ça, c'est des sauvages, etc. Bon c'est ses grands-parents. Donc elle était dans le refus absolu. C'est-à-dire qu'elle était contente de nous voir, contente de savoir qu'on avait travaillé avec sa communauté, et tout ça, mais quand ils ont vu les photos... Mes photos, les gens ils sont en pagne, ils sont poitrine nue, c'était comme ça qu'on vivait. Ils ont refusé ça, parce que maintenant ils sont habillés comme nous, surtout quand ils viennent en France. Donc ils n'étaient pas du tout dans l'admiration de l'image de leurs ancêtres, mais dans le déni de leur passé. Donc ça veut dire que si moi je mets une photo, et maintenant quand je fais une conférence, je projette et je dis bien que ces photos elles ont été prises en 1975 ou en 1980. Parce que c'est certain que si tu vas là-bas aujourd'hui, ils n'ont plus du tout le même accoutrement : dans le même village, ils ont les mêmes gestes, ils font les mêmes activités, ils ne sont absolument pas habillés pareil. Donc c'est important, il y a la question de la conscience de leur histoire. Et en même temps dans le monde actuel, une femme aux seins nus ou un petit enfant, tu peux avoir des vraies emmerdes avec ça. On est quand même dans une ère pudibonde assez marquée. Donc ça veut dire que la mise à disposition des photos, elle est difficile à imaginer. Et pour moi c'est embêtant parce que les photos elles ont du sens que si elles sont utilisées comme documentation, enfin j'entends consultées ou regardées. Donc c'est un vrai problème à mon avis ». (Thomas, octobre 2020)

Pour d'autres chercheurs.euses, c'est la question de **la confiance qui leur est accordée dans la relation d'enquête qui est remise en question par le projet de partage**. C'est un argument récurrent dans les entretiens. Marie se dit très sensible aux enjeux de partage des données de la recherche dans une logique d'accumulation, mais elle explique en quoi ses entretiens, qui constituent une grande partie de ses matériaux ethnographiques, ne sont pas partageables. Au-delà de la question légale des données personnelles identifiantes, ce qui lui pose problème c'est de trahir la confiance qui lui est accordée :

« Et là où je ne partage pas, c'est surtout les entretiens, où je considère que c'est intime. Mais sinon ce qui ne concerne que moi, que j'ai produit et qui n'engage personne, j'ai toujours eu tendance à essayer de le partager, avec les collègues, etc.

[Et quand vous parlez d'intimité, c'est l'intimité de la personne ou la vôtre, ou les deux ?]

C'est les deux, je trouve que c'est les deux en fait. C'est privé en fait, plus qu'intime. En fait c'est confidentiel le terme. Je ne les partage pas, parce que c'est le contrat implicite. Les partager ce serait une trahison parce que c'est quelque chose qu'on me donne. Je les cite de manière anonyme ». (Marie, chercheuse non titulaire, octobre 2020)

Cette notion "d'intimité", difficile à définir comme le souligne cette citation, apparaît à diverses reprises dans les entretiens que nous avons collectés. Une autre chercheuse affirme ainsi que les "carnets de terrain, ça relève tout autant de l'intime que de l'observation scientifique", et à ce titre, cela ne lui paraît pas partageable. Il existe ainsi une argumentation très forte dans le sens de protéger ce qui a été "donné", ou "confié" à un chercheur, dans une relation de confiance, et selon un engagement réciproque, qui implique en règle générale la protection de l'anonymat de la personne. On s'inscrit pleinement ainsi ici dans la partie "fermé autant que nécessaire" énoncée par le mouvement de l'ORD. La spécificité est que la caractérisation

juridique du matériau ne constitue pas le seul critère en jeu pour rendre cet arbitrage, alors que le critère éthique apparaît déterminant (Mouton, 2018). La difficulté en l'occurrence, est qu'**une grande partie des matériaux produits par les ethnographes apparaissent intégrer ce caractère d'exception "fermé autant que nécessaire", ou se situer à sa marge.** C'est pourquoi on ne peut pas conclure des témoignages que nous avons recueillis que les ethnographes sont contre "l'ouverture autant que possible" de leurs données de terrain : les formes et motifs de refus catégoriques de partager restent circonscrits à des configurations particulières (mais fréquentes) qui posent des problèmes éthiques ou déontologiques majeurs, et peuvent nuire soit aux communautés étudiées, soit au.à la chercheur.euse, comme nous en avons vu ici un échantillon d'exemples assez restreint⁵⁰. En revanche, **il existe une diversité de motivations, de projets et de formes de partages plus artisanales, qui ne correspondent pas dans la majorité des cas au modèle de la mise à disposition des données ouvertes sur des entrepôts dédiés et interopérables.** Dans ces formes de partage, l'enjeu qui apparaît comme majeur au regard des témoignages recueillis est **la conservation d'une maîtrise sur l'accès aux données, et les formes de réutilisations qui peuvent en être faites.** Cela explique par exemple pourquoi le cadre de la transmission des archives via des structures dépositaires spécialisées apparaît particulièrement rassurant, puisqu'il garantit un suivi et une conservation des archives déposées par une structure et un personnel compétent dans la gestion et la valorisation de ce type de fonds ; ou encore, le partage entre pairs, qui permet le suivi et l'accompagnement du travail réalisé sur les données, dans le cadre d'un périmètre et de relations de confiance.

A RETENIR

- **Le partage des matériaux de recherche, à divers degrés et selon des formes souvent artisanales, est ancré dans la discipline anthropologique** et n'est pas rare chez les ethnographes.
- Les lourdes opérations de tri, d'indexation, et de traitements des données pour permettre leur appropriation par d'autres, supposent un **très fort investissement des chercheurs.euses. sur plusieurs années.**
- **Les formes de partage ouvertes offrant des possibilités de réutilisations libres sont très minoritaires** : les ethnographes enquêté.e.s publient très peu, au regard de la masse des matériaux qu'ils.elles collectent, sur des entrepôts ouverts, quelle qu'en soit la nature.
- Pour les ethnographes enquêté.e.s, **conserver une forme de contrôle sur le public destinataire de ces partages et sur les motifs de réutilisation** (en général circonscrits à la recherche, l'enseignement, la création et la réappropriation culturelle) **apparaît comme une préoccupation majeure.**

⁵⁰ Notre enquête met en lumière une plus grande diversité de situations et de motifs qui font que le partage des matériaux apparaît impossible. Nous avons fait le choix ici de nous limiter à quelques exemples concrets dans la mesure où le rapport présente déjà par ailleurs de nombreuses illustrations de cet état de fait.

4.3.3. Rapports à la protection des matériaux ethnographiques : entre éthique de recherche et cadre juridique

Les témoignages divers sur les impossibilités de partager des matériaux ethnographiques renvoient au cadre plus général de la protection des données, et en particulier, la protection des personnes incorporées dans ces données, c'est-à-dire le champ d'application du RGPD (Règlement général pour la protection des données personnelles). Les différents exemples que nous avons évoqués dans la section précédente montrent que les chercheurs tendent à aborder cet enjeu de protection de leurs interlocuteurs à travers leur éthique de recherche davantage que par le cadre juridique applicable, même si, dans certains cas, des manquements vis-à-vis de la mise en conformité juridique peuvent les empêcher de partager leur données (comme on l'a vu à travers l'exemple de photos pour lesquelles les consentements officiels n'ont pas été recueillis). Cette posture est très souvent reliée à une double spécificité de l'approche ethnographique, qui repose fondamentalement sur **une rencontre interculturelle**, et donc d'une part sur la nécessité de créer **une relation de confiance avec ses interlocuteurs.trices**, et d'autre part, celle de **s'extraire de son ethnocentrisme pour pouvoir aborder les points de vue émic des communautés ethnographiées**. C'est-à-dire qu'il s'agit de comprendre ce qui est d'usage et ce qui est acceptable ou non pour ces dernières, selon leur propre organisation sociale et leur manière de concevoir leur relation à l'autre, en particulier à l'ethnographe. L'enjeu pour le.la chercheur.euse est de documenter, au prisme de cette relation particulière, une situation ou une réalité observée, un "réel de référence" pour reprendre les termes d'Olivier de Sardan (1996), tout en respectant cette relation de confiance et d'attention qui lui a permis d'y accéder. Ainsi, on voit apparaître dans leurs témoignages que la question qui se pose à eux.elles à ce moment-là, et plus tard lorsqu'un éventuel partage est envisagé, n'est pas tant celle du droit applicable par un chercheur français qui collecte des données à caractère personnel, que **celle du respect du contrat de confiance qui les lie au terrain**, et qui peut, s'il est rompu, mettre en péril aussi bien les membres des communautés étudiées que le.la chercheur.euse lui.elle-même. Plusieurs témoignages évoquent ainsi comment cet enjeu de protection -majeur- de ce qui est défini comme des "données personnelles et sensibles" dans le cadre du RGPD, est incorporé dans l'éthique et les pratiques de recherche des ethnographes, et ne porte pas nécessairement sur les mêmes éléments que ceux définis par la loi. A cet égard, le rapport au recueil du consentement s'avère un thème épineux, et à géométrie variable en fonction des situations, des contextes culturels dans lesquels on travaille, et aussi des générations de praticien.ne.s qui n'ont pas tous reçu les mêmes formes de sensibilisations (ou formations) au sujet - sujet qui a également évolué dans la période d'activité des chercheurs.euses. Le récit précédent de Thomas (voir supra, [ENCADRE 9](#)) évoquant son impossibilité à produire aujourd'hui des autorisations qu'il a obtenues sous une autre forme il y a 40 ans est à ce titre très éclairant. Un autre chercheur, Emmanuel, qui travaille dans les contextes géopolitiques difficiles des pays du Moyen-Orient, raconte de son côté comment il s'est heurté à l'enjeu de formalisation des consentements, alors même que cela ne peut pas faire sens de son point de vue dans les situations sur lesquelles il enquête ([ENCADRE 10](#)). En amont de l'extrait suivant, Emmanuel avait souligné comment il écrit volontairement très mal dans ses carnets de terrain, en français (alors qu'il travaille en langue arabe), de sorte que personne ne puisse lire ou comprendre les informations qui y figurent et qui pourraient porter de graves préjudices à certains interlocuteurs, comme des peines de prison.

Encadré 10. Les accords tacites sur le terrain. Emmanuel, Ethnologue, chercheur titulaire
(récit recueilli en octobre 2020)

« Quand je suis dans une démarche ethnographique, les gens, **je ne leur cache pas ce que je fais, ils savent que je suis anthropologue**. Euh donc je ne suis pas masqué, et quand on passe 2 heures à discuter, et que je leur pose une question, et que je leur repose une autre question, **ils savent que je suis en train de travailler. Et donc je considère que ce qu'ils me disent dans ce cadre est déjà éclairé, dans le sens où ils savent sur quoi je travaille. Il n'y a pas une déclaration formelle de consentement éclairé, mais ils sont éclairés et ils sont d'accord puisqu'ils me parlent**. C'est comme les photos, quand les gens me regardent, ils savent que j'ai 25 photos d'une même personne en train de bêcher dans son jardin, ils savent, ils se prêtent à l'exercice. Donc s'ils ne veulent pas je ne les prends pas en photo. Je ne fais pas de photos volées. C'est quand les gens te connaissent, ils te regardent, il y a un regard. Voilà, ils sont... Ils t'ont oublié à force.

Une fois j'ai essayé d'obtenir des déclarations formelles, c'est parce que [une instance] m'y obligeait, pour une vidéo qui apparaissait dans [une exposition]. Et on m'a demandé, il faut qu'ils remplissent le formulaire, consentement éclairé. Sauf que j'ai dit, ben il est en français votre formulaire, donc ils me l'ont traduit en arabe, et quand j'ai tourné, il n'y en a qu'un qui a bien voulu signer, il m'a dit je te fais confiance, donc je signe. Parce que signer pour eux, c'est s'engager... On a pris l'habitude en France de signer partout, je ne passe pas une journée sans signer un papier sur un PDF. Chez eux c'est un truc, c'est important : c'est pour la police, c'est un mariage, du coup signer c'est... Mais pourquoi ? Non, je t'ai dit que j'étais d'accord. Bon alors je peux te filmer en train de me dire que tu es d'accord ? Bon d'accord, oui. Mais ils ne voulaient pas signer un bout de papier. Donc je n'ai pas eu les autorisations. [...]

Si on m'oblige à faire signer, oui ça peut me poser des problèmes pour ma recherche. C'est complètement irréaliste par ailleurs. Non faisable, parce que, encore une fois, je ne passe pas mon temps à faire des entretiens formels avec les gens. C'est une discussion flottante, au cours de soirées, de café, et thé pris ensemble, il y a une dizaine de personnes, il y en a un qui dit ça. Ils savent que je suis là, peut-être qu'ils n'y pensent pas en permanence quand ils disent des choses. Mais moi aussi **je me sens tenu par un devoir de confidentialité et de ne pas nuire aux personnes qui m'ont donné ou dont j'ai pris des informations. Donc c'est vrai que je suis un peu juge là-dedans**, mais oui voilà. Par exemple en Arabie Saoudite, s'il y a quelqu'un qui me dit ce qu'il pense du roi, qui est un petit peu divergent, je peux le noter dans mon carnet, personne ne pourra le comprendre, voilà, c'est important que je comprenne qu'il y a des points de vue divergents par rapport à l'unanimité, puisque c'est une monarchie absolue, je ne peux pas demander au type : vous êtes d'accord pour que j'enregistre sur mon carnet ? Bien sûr qu'il ne va pas être d'accord, pourtant c'est une donnée importante. Ça voudrait dire, puisque je ne l'ai pas noté, il faudrait que je note dans les conclusions d'un article que tout le monde est d'accord avec la monarchie absolue, donc ça n'a pas de sens. C'est juste impossible ».

Par contraste avec ce récit d'expérience, celui d'Amandine, jeune enseignante-chercheuse, témoigne d'un contexte de recherche tout à fait différent, au Brésil, où elle a fait sa thèse. Dans le contexte brésilien, les autorités locales imposent aux chercheurs.euses de passer par un processus très élaboré de demandes d'autorisations de recherche et de collecte de consentements éclairés formalisés. Ce processus est contrôlé par plusieurs comités d'éthique locaux, et conditionne l'activité d'enquête et le devenir des matériaux collectés. Elle évoque le caractère "lourd" et coûteux en temps de ces démarches, mais qui au final lui ouvrent des portes sur le terrain, et l'aident à instaurer un climat de confiance dans son enquête (ENCADRE 11).

Encadré 11. Face au formalisme institutionnel. Amandine, Enseignante-chercheuse titulaire en socio-anthropologie (récit recueilli en novembre 2020).

« Sur les formalités, pour ma thèse, on avait travaillé du début à la fin avec des autorisations : **consentements préalables libres et éclairés, signés par toutes les personnes qui ont fait des entretiens, à qui on avait présenté le projet, donc on leur laissait aussi un document écrit avec la présentation du projet.** Donc tous mes matériaux de thèse, en tous cas on a fonctionné comme ça, parce que c'était au Brésil, et que c'était hors de question pour ma directrice de thèse et moi de se plier aux formalités, et c'était important en termes éthiques.

[...] [Mon institution en France] ne m'a pas aidée, ils n'ont pas de service d'appui là-dessus. Donc c'est énormément moi qui ai pris en charge ces questions-là, et ma directrice m'a beaucoup aidée, elle était très impliquée. Parce que c'était un peu un truc étrange où on avait besoin des autorisations pour démarrer, et en même temps on était obligé de démarrer avant d'avoir les autorisations. Donc c'était un truc un peu itératif où on avançait pas à pas. Donc mon directeur de thèse au Brésil m'a beaucoup aidée aussi, il était un peu le chaperon sur ces questions-là. Tu as toute une série d'autorisations à obtenir, et ce que tu cherches à obtenir au final, c'est à la CONETH : Commission nationale d'éthique pour la recherche au Brésil. Donc avant d'avoir les autorisations de la CONETH, il faut les autorisations du comité d'éthique de ton université de rattachement au Brésil, et pour les avoir, il te fallait déjà avoir les signatures des organisations auprès desquelles tu allais enquêter, donc il fallait déjà les avoir rencontrées. Donc en fait tu es obligée d'avoir déjà commencé ton terrain pour les rencontrer, pour pouvoir avoir les autorisations. Et il fallait aussi les autorisations de la fondation de l'indien au Brésil, pour aller en terres indigènes. **Donc c'est très compliqué d'avoir les autorisations.** Par contre en termes d'entretiens individuels, il y avait quand même des modèles qui étaient fournis par les différents comités d'éthique, donc on suivait un peu les modèles pour le consentement préalable qu'on faisait signer. Et puis pour le modèle de présentation du projet de recherche, on avait été guidé et on avait des retours pour améliorer le document. Donc au final j'ai obtenu toutes les autorisations et fait signer tous les consentements.

[Donc c'est le cadre brésilien qui a amené à faire ces démarches, pas des obligations du côté français ?]

Oui. C'est important au Brésil de se soumettre à ce processus, si on veut publier derrière, et aussi si on ne veut pas se faire sortir quand on est sur le terrain. Mais ça fait réfléchir, parce que par exemple sur un autre projet en France, on ne l'a pas fait, de faire signer des consentements. **Mais c'est tellement lourd comme processus, que quand on n'est pas obligé de le faire, j'imagine qu'on ne le fait pas.** C'est vrai que du coup ici on ne s'est même pas posé la question, on ne l'a pas fait.

[Et ce processus, est-ce que ça modifie ta recherche, la manière dont tu travailles sur le terrain ?]

Au Brésil **c'était vraiment lourd, parfois ça m'a obligée à rentrer à Manaus alors que j'étais sur le terrain.** Mais je pense que ça modifie dans un sens positif : **c'est-à-dire que ça prend du temps, de l'énergie, mais ça pose vraiment les bases...je ne sais pas comment dire, les bases claires et de confiance avec les acteurs auprès desquels tu vas enquêter.** Et au Brésil, j'enquêtais avec des groupes de producteurs qui étaient concurrents au niveau des filières, des marchés, etc. Donc c'était important pour eux de savoir ce que j'allais faire des données, si je n'allais pas aller diffuser des contenus d'entretiens à des producteurs concurrents. Et du coup d'avoir ce processus avec des papiers signés, où eux s'engagent mais aussi moi surtout je m'engage, c'était un processus je pense hyper bénéfique. Après sur la qualité des données recueillies, sur la confiance entre l'enquêtrice et les enquêtés, je pense que c'était vraiment bien ».

Ces deux exemples opposés, d'une approche fondée uniquement sur l'éthique individuelle du chercheur d'un côté, et de formalisation et d'institutionnalisation très forte du recueil de consentement de l'autre, montrent combien les singularités culturelles, politiques et réglementaires du contexte étudié influencent les pratiques des chercheurs. euses au regard de la forme et du niveau de formalisation du recueil de consentement afférent à la protection des données collectées. **Pourtant dans les deux cas, l'intention de "protéger" les interlocuteurs existe et guide la pratique** : dans le premier cas, l'excès de formalisation du consentement risque de nuire à la personne et à la qualité de l'information collectée en laissant une trace

écrite de sa relation à l'enquêteur et des informations subversives qu'elle aurait pu lui fournir, plaçant ainsi le chercheur comme l'unique instance dépositaire et responsable de cette information qui lui a été confiée. Dans le second cas, le fort niveau de formalisation du consentement rassure les interlocuteurs et participe à créer une relation de confiance vis-à-vis de l'enquêtrice, via la garantie apportée par des instances tierces de contrôle. Une configuration qui se comprend à l'aune de l'histoire de la recherche scientifique et du colonialisme au Brésil et la crainte de pillage des savoirs, en particulier les savoirs traditionnels écologiques.

Dans l'échantillon de personnes interrogées, très peu avaient déjà utilisé des formats de consentements écrits, à l'exception de configurations précises :

- des chercheurs.euses travaillant dans des contextes locaux qui l'imposent dans la mesure où ils n'est pas possible de mettre en œuvre un programme de recherche sans autorisations spécifiques et remise des consentements signés
- des programmes de recherche interdisciplinaires avec les sciences de la santé en particulier, pour lesquelles il existe une plus grande sensibilisation à ces enjeux et un cadre très strict vis à vis des données de santé ; donc ce qui est appliqué en sciences de la santé est transposé aux enquêtes ethnographiques. D'autre part, la mise en œuvre de ces programmes fait l'objet d'évaluations par des comités éthiques qui imposent d'aborder la question du droit applicable par types de données collectées.
- des contextes de diffusion (vidéothèque dont la gestion a été transmise à un centre de ressources) ou d'exposition (musée) de certains matériaux de recherche (photos, vidéos) où en particulier la voix ou l'image de personnes sont exposées.

En dehors de ces configurations, les praticien.ne.s expliquent s'appuyer sur des accords "tacites", ou "implicites", qui ont déjà été illustrés à travers plusieurs citations, auxquelles on pourrait ajouter celle-ci, qui ne donne pas de justification particulière mais constate simplement un état de fait sur les pratiques :

« Je n'utilise pas de formulaires de consentement. Pas pour l'instant. J'ai des accords tacites avec les personnes, sur le fait que ce sera anonymisé. Je floute les photos, je ne mets pas les noms de lieux ou de personnes. Mais ce n'est pas formalisé ». (Jacques, chercheur titulaire, octobre 2020)

Il apparaît que le niveau de connaissance par les chercheurs.euses du cadre légal en France et son évolution depuis 2018 avec l'adoption du RGPD est assez faible, malgré les nombreuses actions de communication et sensibilisation organisées au sein des laboratoires. L'angle mort qui se dévoile dans les témoignages recueillis est celui de la manière dont ils.elles doivent se saisir de ce nouveau cadre : c'est-à-dire, ce qu'ils.elles devraient concrètement faire ou changer dans leurs pratiques.

« Je leur demande s'ils sont d'accord pour que j'enregistre, et à partir du moment où ils ont dit qu'ils étaient d'accord, je lance l'enregistrement.

[Et est-ce que la nouvelle réglementation RGPD, ça change quelque chose pour toi, dans la manière dont tu travailles ou tu recueilles les consentements ?]

Je dois dire que j'ai complètement mis sous le tapis cette question pour le moment, et je vois bien que je vais me faire rattraper par la patrouille un de ces quatre. Il y a des journées de présentation sur RGPD et SHS, je vois bien que c'est des choses auxquelles on ne va pas pouvoir échapper. Mais pour l'instant, je n'ai pas pris le temps de me plonger là-dedans ». (Perrine, chercheuse titulaire, octobre 2020)

Au même titre que Perrine, très rigoureuse et consciencieuse dans le traitement de ses matériaux de recherche et le respect de l'anonymat, peu d'interlocuteurs avaient connaissance de la procédure spécifique concernant le recueil de données à caractère personnel, ni de la teneur de la déclaration de traitement que cela implique auprès du/de la DPD référent.e, alors même que de nombreuses données ethnographiques entrent dans ce champ d'application. D'ailleurs plusieurs interlocuteurs.trices ignoraient l'existence de cette instance, sa fonction et son champ de compétence. Les plus familiers de ces démarches étant les

chercheurs.euses qui ont très récemment participé au dépôt de demandes de financement, ou obtenu un financement pour des projets de grande ampleur (type ANR, ou projet européen) et se sont confrontés au nouvel exercice de rédaction d'un "plan de gestion de données".

Il ne s'agit pas là de pointer du doigt cet état de fait, mais plutôt de mesurer ce qu'il dit du **décalage qui sépare ce cadre juridique entourant la protection des données de recherche à caractère personnel en France, et les pratiques des communautés de recherche engagées dans la collecte et la gestion de données, dont les réticences vis-à-vis de la mise en application de ce cadre juridique se fondent avant tout sur des arguments éthiques et épistémologiques**. Notre enquête a dévoilé plusieurs critères (côté praticien.ne.s) liés à ce décalage, qui mériterait par ailleurs d'être étudié plus en détails pour une caractérisation exhaustive. D'abord un cadre et des démarches qui apparaissent peu lisibles pour eux.elles, notamment en lien avec son évolution récente et la difficulté à identifier les informations importantes et les personnes pouvant venir en appui à des démarches de mise en conformité. Ensuite, le fait de s'être ou non confronté à un cas pratique qui oblige à se pencher sur la question (par exemple une déclaration liées à un programme financé). Le manque de temps également, qui se conjugue à la perception d'une bureaucratisation croissante des démarches administratives liées à l'activité de recherche. Cette dimension rejoint également la tension entre formalisation des démarches, *versus* démarche d'enquête qui privilégie la relation informelle propre à l'approche ethnographique. On observe par ailleurs un effet de réseau : la sensibilisation à la question passe par le fait de voir concrètement des collègues proches mettre ce type de démarches en pratique (on constate cela en particulier dans le cadre de programmes interdisciplinaires). Cet effet de réseau favorise l'échange d'informations sur que faire, comment s'y prendre et qui contacter. Enfin, l'argument éthique qui, dans le discours des chercheurs.euses, apparaît souvent comme prévalent sur le cadre juridique, en tant que le.la chercheur.euse se sent la pleine responsabilité vis-à-vis de ce qui lui est confié et dont il se retrouve dépositaire.

Un autre témoignage éclairant à ce sujet est sans doute le nôtre, prenant appui sur l'observation participante de nos propres démarches de mise en conformité avec le RGPD dans le cadre de cette enquête - donc en tant que chercheuses ethnologues, recueillant des entretiens (enregistrés) qui impliquent un traitement de données personnelles (la voix) et de nombreuses données identifiantes. Nous nous considérons comme relativement familiarisées avec ce type de démarches, du fait de notre thématique de recherche qui nous a sans doute amenées à aborder de manière plus poussée que la moyenne des chercheurs.euses cette caractérisation juridique des données ethnographiques en amont d'une recherche. A l'aube de cette enquête, nous avons donc cherché à nous mettre en conformité, notamment selon l'argument qu'il aurait été mal venu qu'une enquête de ce type ne le soit pas, d'autant plus que des DPD devaient être interrogés. Notre première épreuve a été d'identifier le DPD référent pour formuler nos demandes : vérifier le modèle de consentement que nous avons produit en nous inspirant d'un modèle utilisé pour une autre enquête, et faire une déclaration de traitement. Après plusieurs contacts au sein de notre unité d'affiliation pour identifier le DPD référent, nous découvrons que notre question n'est pas si évidente : on ne sait pas qui est référent comme nous avons plusieurs tutelles, et on ne sait pas vraiment à qui s'adresser pour poser la question. Finalement, nous apprenons par un collègue qui a déjà reçu une demande de ce type que c'est le DPD de l'Université de Bretagne Occidentale qui devrait traiter notre demande, mais qu'à cette date le poste est vacant. Nous nous replions sur le CNRS, seconde tutelle, sans être sûrs de savoir si nous devons contacter le service central ou la délégation régionale. Nous passons donc par la délégation régionale pour nous informer, et découvrons finalement que c'est bien le CNRS qui est référent pour notre unité⁵¹. L'échelon régional fait suivre notre demande au service national qui la traitera avec peu d'allers-retours, puisqu'elle est assez simple. Il faut alors compléter et retourner le "formulaire

⁵¹ Le DPO de chaque tutelle reste le référent pour le personnel des tutelles si besoin, mais il faut néanmoins déclarer un DPO référent pour l'Unité et c'est la démarche menée dans le cadre de l'enquête PARDOQ, par laquelle le DPO CNRS était sollicité pour la première fois au niveau de l'unité, qui a conduit à ce que le CNRS se propose comme DPO référent de l'unité (après constat de l'absence de déclaration de référent unité).

d’inscription au registre” pour déclarer le traitement de données. A notre surprise, le directeur d’unité (unité de recherche interdisciplinaire avec peu de chercheurs.euses “qualitaticien.nes”) n’a jamais rempli ce type de formulaire, et s’étonne de se trouver impliqué dans ce projet en tant que “responsable de traitement”. Une fois ces démarches effectuées, nous remarquons au cours de l’enquête que nous devenons un “modèle”, c’est-à-dire une ressource pour savoir quelles démarches faire ou quels modèles de documents utiliser. Notre formulaire de consentement est par exemple retransmis par nous-mêmes ou des collègues comme format à adapter pour d’autres types de projets. Mais au-delà de cela, nous devenons vite identifiées comme une ressource sur le sujet, un peu, selon la métaphore qui nous est venue, comme le collègue qui maîtrise vaguement un logiciel de SIG et à qui tout le monde vient demander de l’aide pour faire ses cartes. Par ailleurs, il n’était pas rare qu’au cours des entretiens que nous avons réalisés, l’enquêtrice échange à un moment autour de questions que se pose un.e interlocuteur.trice sur ce type mise en conformité. C’est même ce motif qui a amené à un des entretiens spontanés que nous avons réalisés, sollicités par une chercheuse à la fois intéressée par l’enquête et qui avait diverses questions à nous poser sur le thème même de l’enquête (suite à un appel ouvert sur une liste de diffusion). Cette logique de don/contre-don nous a paru tout à fait naturelle, à la limite près que nous ne nous sentions pas dépositaires de la compétence ni de l’expertise nécessaires pour répondre à ce type de questions, et avons peur de délivrer des informations erronées. Notre stratégie consistait donc simplement à indiquer où trouver les bonnes ressources, et témoigner de notre propre expérience. Une chercheuse nous a recontactées quelques semaines après son entretien pour nous poser des questions plus spécifiques sur les démarches à réaliser dans le cadre de la mise en œuvre d’un programme de recherche pour lequel elle venait d’obtenir un financement, et d’autres collègues, de nos réseaux personnels, qui savent vaguement que nous travaillons autour de ce sujet nous consultent régulièrement, telles des expertes du sujet, pour relire des plans de gestion des données, ou savoir quelles déclarations ils.elles doivent faire, comment prendre contact avec un DPD et quel niveau d’informations ils.elles doivent lui fournir.

Finalement, ce que ce retour d’expérience a mis en lumière est **le besoin de s’adresser et de s’identifier à un interlocuteur proche** (du point de vue du type de démarche et de données concernées), en capacité de **témoigner d’une procédure concrète, et des implications concrètes qu’elle peut avoir** : comment prépare-t-on le recueil de consentement formalisé, à qui s’adresse-t-on pour le faire vérifier, qu’est-ce que cette déclaration de traitement change concrètement dans ce qu’on fait avec les données, qu’est-ce que le cadre légal nous oblige à faire et comment on fait dans la réalité? Dans le même temps, elle a fait apparaître la **difficulté de caractériser chaque situation**, parce que notre cas n’était jamais transposable tel quel au cas du.de la collègue, et la réponse juridique ne pouvait être formulée que par un.une spécialiste. Dans ces décalages, la question qui se posait était donc : où trouver l’information concrète ? A cela, nous répondions de consulter le DPD qui est en capacité de réaliser ces aiguillages fins, et qu’il nous semblait qu’il ne fallait pas uniquement envisager le contact avec le DPD comme le dépôt des documents formels et déjà parfaitement complétés (ce qui était souvent le cas), mais comme un échange et un accompagnement pour déterminer comment se mettre en conformité en fonction des particularités de la recherche envisagée. C’est alors que nous avons pris la mesure de l’engorgement de travail pour ces dernier.e.s si tous.toutes les chercheurs.euses décidaient demain de se mettre en conformité.

A RETENIR

- **La référence aux exigences éthiques en matière de protection des enquêtés.e.s préexiste et l’emporte, dans la pratique, sur la référence à des cadres juridiques peu maîtrisés, a fortiori en ce qui concerne leurs évolutions récentes.**
- **L’attention portée au droit local, en particulier au droit coutumier, prévalant sur les terrains d’enquête constitue une spécificité disciplinaire.**
- **La formalisation écrite du consentement est rare et considérée par de nombreux ethnographes comme difficilement applicable, notamment parce qu’elle réduit la sphère de l’informel qui constitue un fondement épistémologique dans la pratique ethnographique.**

4.3.4. Rapport à l'injonction d'ouverture

Les résultats d'enquête que nous avons présentés montrent que la question du partage des matériaux ethnographiques est une dimension constitutive des pratiques de recherche des chercheurs.euses que nous avons interrogés. **Il existe, du point de vue des chercheurs.euses, un enjeu de partage des matériaux de recherche qu'ils.elles produisent en dehors du processus de publication académique classique**, et qui s'inscrit en priorité dans des motifs de transmission (fabrique d'archives), de restitution, de travail collaboratif à l'échelle d'une équipe scientifique, de diffusion des connaissances et de recherche création. Néanmoins, ces pratiques du partage, les enjeux et les formes qui les caractérisent, s'inscrivent en décalage avec le cadrage proposé par le mouvement de l'ouverture des données : **sauf de rares exceptions, les matériaux ethnographiques ne sont pas publiés sur des entrepôts de données en accès libre**, pour une série de raisons que nous nous sommes attachées à décrire. Des raisons qui s'appréhendent à la fois selon l'épistémologie de la discipline, l'éthique des praticien.ne.s et les spécificités des matériaux produits.

Les témoignages que nous avons rassemblés montrent par ailleurs que les praticien.ne.s sont relativement à l'aise avec les pratiques de partages qu'ils.elles mettent en oeuvre : leurs formes, leurs périmètres et leurs objectifs de diffusion apparaissent faire sens dans leur démarche plus large de production et de diffusion de savoirs fondés sur l'ethnographie. Ainsi, la manière dont le partage est questionné ne se positionne pas autour d'un enjeu à ouvrir davantage de données librement, mais davantage autour de la capacité à mieux mettre en valeur certains projets de partage en cours ou envisagés : trouver des outils plus adaptés, disposer de ressources pour déployer un projet, avoir du temps pour se consacrer à ses données en dehors de leur unique analyse (nous l'avons vu par exemple avec l'enjeu de transmettre une vidéothèque ou celui de valoriser des collections d'artéfacts considérés comme trop invisibles).

Il apparaît donc en pratique que les **chercheurs.euses abordent le partage selon une démarche que l'on pourrait qualifier d'ancrée et artisanale** : à partir de leurs corpus, de leur expérience, de leur éthique de recherche, de leurs préoccupations vis-à-vis de la diffusion de leur ethnographie, ils conçoivent des projets de mises en partage d'une sélection de matériaux de recherche, auprès d'une sélection d'acteurs, en restant soucieux d'accompagner ces partages, quels qu'en soient leurs formes. Cette démarche s'inscrit dans une conception inverse à celle portée par le mouvement d'ouverture des données, qui implique une standardisation et une systématisation du partage des données dès lors que toute contre-indication à ouvrir les données est levée. Le rapport au partage et le sens des pratiques embarquées dans ces deux conceptions contrastées apparaissent donc en décalage. Comment alors les chercheurs.euses ethnographes perçoivent-ils.elles le mouvement de l'Open Research Data? Dans quelles mesures questionnent-ils ces écarts et appréhendent-ils les changements de paradigmes que porte ce mouvement?

L'Open Data : un cadre peu connu par les ethnographes

Un constat émerge de notre enquête : les implications concrètes portées par le mouvement sont peu connues (voir inconnues) des praticien.ne.s que nous avons interrogés, et peu d'entre eux.elles y ont déjà été confrontés en pratique. Seuls deux chercheurs.euses ont déjà été concernés par "une demande" d'ouverture de leurs données, tous deux dans le cadre de leur participation à un projet international financé par l'ERC. Dans le premier cas, le dépôt des données sur un entrepôt a été réalisé par la porteuse du projet, et le chercheur impliqué n'a, à son niveau, participé qu'à une mise à disposition interne au projet d'une série de matériaux qu'il avait produits. Dans le second cas, la configuration est similaire : le porteur de projet gèrera le dépôt global en fin de programme, et la chercheuse interrogée, productrice d'une partie des données collectivisées, n'est pas directement impliquée dans ce processus et a peu de visibilité sur la manière dont il sera mis en œuvre. Dans le premier cas, les données correspondent à des séries standardisées issues de protocoles de recension, et dans le second, des corpus massifs de données "scrapées" (absorbées) sur des réseaux sociaux et anonymisées. Par ailleurs, le seul récit que nous avons recueilli d'une expérience de publication sur la plateforme Zenodo d'un corpus de données associée à une publication scientifique relate une démarche volontaire prenant appui sur une sensibilité aux valeurs de l'Open Access, alors que le chercheur en question explique ne pas connaître les recommandations liées au

mouvement de l'Open Data, et ne jamais avoir été confronté à une demande de ce type émanant d'un financeur ou d'une revue par exemple. Il apparaît ainsi qu'à l'automne 2020, période de notre enquête, rares, voire exceptionnels sont les cas de chercheurs.euses qui ont reçu ou répondu à une incitation à "ouvrir leurs données" au sens de leur publication dans un entrepôt dédié, soit par le biais de revues scientifiques, de financeurs, ou de leurs tutelles.

Des écarts de conceptions

Il existe par ailleurs **une confusion entre ouverture des publications et ouverture des données dans l'interprétation des praticien.nes**. Cette confusion ne porte pas en soi sur la proximité et les racines communes à ces deux mouvements imbriqués de la Science Ouverte, fondées sur des valeurs communes. Elle reflète davantage une spécificité de recherche qualitative, et du rapport à l'exposition des données dans ce contexte : pour une partie des ethnographes interrogés, en particulier les profils "auteurs", **ouvrir ses données revient à ouvrir ses publications en les déposant sur une archive en libre accès**. Cette conception rejoint les discours que nous avons évoqués dans la partie 4.3.1 sur le rapport aux données : celles-ci sont partagées *via* les extraits qui en sont présentés dans une publication, *via* les descriptions que peuvent en faire les chercheurs, qui considèrent de fait que leur travail consiste précisément à rendre intelligible leur ethnographie dans le cadre de la publication scientifique. Une question illustre particulièrement ce décalage ; elle a été posée par un chercheur anthropologue à l'occasion d'un atelier scientifique autour du partage des données, et à l'issue d'une série de retours d'expériences et échanges divers à ce sujet : "Pourquoi les publications ne suffisent-elles plus?" (Ethnologue, le 27/04/2021, atelier "Du terrain au web" organisé au LESC). Ainsi, il ne s'agit pas tant d'une confusion liée à la connaissance et l'interprétation de ce qu'est le mouvement d'ouverture des données, que de la question du rapport plus profond aux données ethnographiques dans la construction des savoirs scientifiques, et de ce que leur publicisation "brute" au cours de la recherche change à la construction de ces savoirs.

Un autre niveau de confusion touche la question de **la tension entre ouverture et protection des données**. Nous avons vu en partie 4.3.3 que le cadre juridique de la protection des données personnelles était assez mal connu, et que l'éthique de recherche primait sur son application. Le mouvement d'ouverture des données ne concerne pas les données à caractère personnel ou les données sensibles : c'est ce qui est énoncé sous le principe "fermé autant que nécessaire". Ainsi, le mouvement d'ouverture des données, nous l'avons vu en partie 2, est distinct de celui de la protection des données, bien que tous deux aient connu des évolutions concomitantes dans les années récentes (2016-2018) et soient par ailleurs relativement imbriqués. Les praticien.ne.s tendent à percevoir ce couple comme un ensemble de contraintes qui s'impose progressivement à eux dans un processus de bureaucratisation croissante de la recherche. **Cette confusion autour de l'appréhension de ces deux mouvements distincts est renforcée par le fait que l'essentiel des corpus ethnographiques apparaît en zone floue dans cette polarité entre ouvert et fermé, et positionnés davantage du côté de la protection** (sans pour autant que le cadre juridique de la protection des données à caractère personnel soit pleinement connu, compris ou approprié). Nous l'avons largement vu dans la teneur des témoignages qui évoquent l'impossible ouverture des entretiens par exemple : or, les entretiens sont par essence des données protégées au titre du RGPD, en particulier lorsque la voix est enregistrée et que la personne est identifiable (ce qui est le but dans un entretien ethnographique situé), il n'est donc en théorie pas question d'ouvrir ce type de données, sans consentement explicite. Le fait est que les données citées comme les plus importantes pour les chercheurs.euses dans leurs recherches (observations et entretiens) sont aussi les plus identifiantes. Ainsi, lorsque cet enjeu d'ouverture des données est abordé dans l'enquête que nous avons menée, il fait écho de manière privilégiée avec ces corpus "coeur" de l'ethnographie particulièrement sujets à des mesures de protection. A l'inverse, les données qui se prêtent plus facilement à une mise en partage sont celles qui sont considérées comme plus secondaires dans l'analyse, ou illustratives, comme des photographies non identifiantes, ou certaines archives. Finalement **le caractère marginal de l'ethnographie vis-à-vis de l'ouverture des données se comprend à l'aune de cette difficulté à la fois à qualifier le statut de ces données selon le principe "aussi ouvert que possible, fermé autant que nécessaire", et du fait que les données les plus essentielles à la recherche sont aussi les plus**

sujettes à protection. Cela s'inscrit de surcroît dans un contexte où la lisibilité des deux cadres afférents à ces deux dimensions (mouvement d'ouverture et mesures de protection) reste mauvaise dans l'expérience des praticien.ne.s.

L'apparente réticence à l'ouverture des données est à contraster avec une adhésion assez forte des praticien.ne.s interrogé.e.s **aux valeurs du partage et un attachement à la notion de "service public"**, selon laquelle ces chercheurs.euses, financés sur fonds publics, se considèrent avant tout comme responsables de produire et diffuser des connaissances scientifiques. Si de nombreuses réserves sont émises lorsque nous posons la question : "une ouverture de vos données vous paraît-elle possible et souhaitable ?", l'idée de porter une réflexion à ce sujet apparaît comme une perspective qui intéresse les chercheurs.euses. Ce, à la condition que cette réflexion ne soit pas menée sous l'égide d'une mise en conformité avec un cadre imposé, mais **davantage sur le fond, selon les questions éthiques, épistémologiques, déontologiques et politiques que pose ce projet d'ouverture**, et *in fine*, dans le respect des spécificités de la disciplines sur ces différents plans. C'est ce dont témoigne Emmanuel dans cet extrait, ouvert à cette réflexion :

« Par contre l'idée qu'on y réfléchisse n'est pas de trop, je ne suis pas contre l'idée de réfléchir, de savoir comment tu vas gérer les données, de savoir quelles données tu vas mettre en archives ouvertes ou tu vas partager dans un cercle restreint, etc. Pourquoi pas, je n'ai rien contre, je ne suis pas contre le partage, au contraire. Tous mes résultats de recherche, moi je suis un fêru du partage des articles, chapitres. Toutes mes publications sont partagées en accès libre, sur HAL. Et le CNRS l'oblige plus ou moins. Par contre je pense que les données elles-mêmes, il ne faut pas d'obligation. Enfin, nos domaines en sciences sociales en particulier, il ne faut pas d'obligation, il faut que ce soit au cas par cas. Mais qu'on y réfléchisse, pourquoi pas ? Je n'ai rien contre. Ça voudrait dire à l'intérieur de chaque projet, chaque donnée, voir un peu ce qui est jouable, ce qu'on peut diffuser ou pas ». (Emmanuel, chercheur titulaire, octobre 2020)

Emmanuel souligne la question du rapport à l'injonction institutionnelle, qui apparaît entrer en tension avec la démarche scientifique. Jean, chercheur honoraire, met également en cause cet enjeu de soumission à une forme d'autorité externe, problématique à son sens, face à une démarche scientifique qui intègre historiquement l'enjeu du partage des matériaux de recherche :

*« Le souci d'ouvrir des matériaux de terrain a toujours fait partie des préoccupations de l'ethnologue, c'est presque une raison d'être de la démarche scientifique, il ne faut pas confondre cela avec la révolution technologique qu'on vit actuellement. **Le risque, c'est de vouloir imposer ou de codifier ces formes d'ouverture dans des formats qui n'ont pas de sens, ni pour le chercheur, ni pour les autres qui les utiliseraient.** L'ouverture des matériaux de terrain c'est un phénomène constant, qui a toujours existé, mais qui est fait différemment selon les différents chercheurs : tu prends [untel], ou [unetelle], ou moi, on voit bien qu'on pratique chacun une ethnographie différente qui est très liée à notre personnalité. Et il est vrai que les ethnologues peuvent être excessivement jaloux de leurs matériaux, parce qu'ils sont très engagés dans leurs matériaux. [...] Par exemple les carnets de terrain, les partager ne serait pas possible parce que ce ne serait pas lisible, et ça n'intéresse personne, sauf peut-être dans le cas de certains « grands hommes » ou « grandes femmes » qui font eux-mêmes l'objet de thèse, où là ça peut être valorisé. [...] Le problème, je trouve c'est qu'il y a une espèce de soumission à ces injonctions, renforcée par les financements, c'est de l'institutionnel pour moi ça, ce n'est pas de l'évolution de la recherche". (Jean, chercheur honoraire, octobre 2020)*

A RETENIR :

- **Le niveau de connaissance, d'appropriation et de maîtrise du cadre d'Open Data est faible. Dans ce contexte, les pratiques de partage se construisent de manière logique en dehors de ce cadre.**
- **Les données ethnographiques les plus importantes pour ceux et celles qui les collectent sont aussi celles qui posent le plus de difficultés de qualification juridique et qui, de ce fait, peuvent être invisibilisées dans le processus d'ouverture.**
- **L'institutionnalisation de la gestion des données s'inscrit pour les personnes enquêtées dans un contexte plus large de raréfaction et de concentration des ressources, et perd de ce fait en légitimité.**
- **L'intérêt pour les enjeux épistémologiques, éthiques, et politiques associés à l'ouverture des données est beaucoup plus prononcé que la préoccupation de conformation au droit et aux politiques des tutelles.**

Conclusion générale

L'ethnographie est une pratique de plus en plus légitime et partagée au sein des sciences humaines et sociales. Qu'elle intervienne en complément d'autres méthodes de recherche ou de façon exclusive, elle constitue une référence commune pour de nombreux.euses chercheurs.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 ethnographiques. 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.

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 dans ce rapport, 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, 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 types de configuration, les enjeux scientifiques que les chercheurs.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 oeuvre 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.e.s. L'histoire de la pratique ethnographique est marquée par la lente construction de conventions collectives le plus souvent informelles qui vise à réguler le risque pour les enquêté.e.s 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.e.s et de leurs droits, puisqu'elles reposent autant sur la protection de l'indépendance des chercheurs.euses, que sur les pratiques et les choix de techniques littéraires *ad hoc* permettant, suivant les cas, une anonymisation très rigoureuse, une protection sourcilieuse 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 chercheurs.euses pratiquant les méthodes qualitatives et l'ethnographie, et sur la place qu'y occupent 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. A cet égard, le développement des outils et infrastructures numériques constitue, dans ce domaine comme ailleurs, une transformation dans les pratiques dans lesquelles les chercheurs.euses enquêté.e.s 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.ne.s 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 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 d'ouverture 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 défini 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 chercheurs.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.e 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 chercheurs.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 présentés dans le cadre de ce rapport convergent avec les principaux enseignements de la littérature scientifique en études sociales des sciences, présentés en première partie de ce rapport : si les valeurs portées par le mouvement de l'*Open Science*, comme l'impératif catégorique de protection des enquêté.e.s 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.le.s 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 chercheurs.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**. Nous espérons donc que ce rapport contribuera à renforcer l'intelligibilité de ces processus, au premier chef pour ceux et celles qui en sont les chevilles ouvrières.

Synthèse des résultats

1. 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.
2. 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).
3. 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).
4. 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'innovation, dans des cadres non standardisés et peu formalisés.
5. L'enquête PARDOQ montre en particulier l'engagement et le rôle moteur à titre individuel de certain.e.s chercheurs.euses et ingénieur.e.s 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.
6. 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 chercheurs.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.
7. Pour les chercheurs.euses engagé.e.s 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.
8. Quelle que soit la qualité des ressources externes dont il.elle dispose, le.la chercheur.euse/collecteur.trice 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.
9. Ces questions sont plus largement discutées et débattues dans les communautés d'ingénieur.e.s (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.e.s aux côtés des chercheurs.euses, pour des raisons historiques, budgétaires et/ou organisationnelles.
10. 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.

Bibliographie

Abélès, Marion, et Marie-Dominique Mouton. 2008. « Sauvegarder le terrain des ethnologues: un défi relevé en commun ». *Gazette des archives* 212 (4): 89-99.

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](https://doi.org/10.4000/cdst.2061).

Augé, Marc. 2013. *Un ethnologue dans le métro*. Fayard/Pluriel.

Beurier, Anne-Gaëlle. 2019. « L'intermédiation à l'épreuve de la pratique - Analyse des activités de conception et d'intermédiation au sein du programme de recherche coopérative 'Paroles et chemins de l'agriculture littorale' (2016-2019) ». Mémoire de Master 2 Gouvernance de la Transition, Ecologie et Société. AgroParisTech.

Biagioli, Mario. 2012. « From Ciphers to Confidentiality: Secrecy, Openness and Priority in Science ». *The British Journal for the History of Science* 45(2):213-33. doi: [10.1017/S0007087412000088](https://doi.org/10.1017/S0007087412000088).

Bigot, Jean-Édouard. 2018. « Instruments, pratiques et enjeux d'une recherche numériquement équipée en sciences humaines et sociales ». PhD Thesis, Compiègne.

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>.

Borgman, Christine L. 2012. « The Conundrum of Sharing Research Data ». *Journal of the American Society for Information Science and Technology* 63 (6): 1059-78. <https://doi.org/10.1002/asi.22634>.

Bourdaloie, Hélène. 2014. « Ce que le numérique fait aux sciences humaines et sociales. Epistémologie, méthodes et outils en questions ». *tic&société*, n° Vol. 7, N° 2 (mai). <https://doi.org/10.4000/ticetsociete.1500>.

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*, n° #19 (décembre): 183-98. <https://doi.org/10.4000/traces.10968>.

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*, n° 12 (septembre). <https://doi.org/10.4000/pds.5127>.

Cefai, Daniel, et Daniel Cefai. 2010. *L'engagement ethnographique*. Ed. de l'École des hautes études en sciences sociales.

Cefai, Daniel, et Paul Costey. 2009. « Codifier l'engagement ethnographique ». Remarques sur le consentement éclairé, les codes d'éthique et les comités d'éthique. *Vie Idées*.

Chartron, Ghislaine. 2014. « Open access et SHS : Controverses ». *Revue européenne des sciences sociales. European Journal of Social Sciences*, n° 52-1 (mai): 37-63. <https://doi.org/10.4000/ress.2658>.

———. 2018. « L'Open science au prisme de la Commission européenne ». *Education et sociétés* n° 41 (1): 177-93.

Chartron, Ghislaine, et Joachim Schöpfel. 2017. « Open access et Open science en débat ». *Revue française des sciences de l'information et de la communication*, n° 11 (août). <http://journals.openedition.org/rfsic/3331>.

- Cohn, Clarice. 2020. « Retornando sua coleção de material audiovisual: os dilemas de uma antropóloga ». *Journal de la Société des Américanistes* 106 (106-2): 105-27.
- Dassa, Michèle. 2019. « Comment caractériser les inconduites au niveau des publications des différentes communautés SHS? Eléments de réflexion ».
- Datson, Lorraine J, et Peter Galison. 2010. *Objectivity*. Zone Books - MIT.
- Denis, Jérôme. 2018. *Le travail invisible des données: Eléments pour une sociologie des infrastructures scripturales*. Transvalor, Presses des mines.
- Denis, Jérôme, et Samuel Goëta. 2016. « "Brutification" et instauration des données. La fabrique attentionnée de l'open data ».
- Deschamps, Florence. s. d. « La parole et le droit, de l'insouciance à la vigilance ». In *La diffusion numérique des données en SHS. Guide des bonnes pratiques éthiques et juridiques*.
- Desquilbet, Loic, Sabrina Granger, Boris Hejblum, Arnaud Legrand, Pascal Pernot, Nicolas P. Rougier, Elisa de Castro Guerra, et al. 2019. *Vers une recherche reproductible*. Édité par Unité régionale de formation à l'information scientifique et technique de Bordeaux. Unité régionale de formation à l'information scientifique et technique de Bordeaux. <https://hal.archives-ouvertes.fr/hal-02144142>.
- 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.
- Drucker, Johanna. 2011. « Humanities approaches to graphical display ». *Digital Humanities Quarterly* 5 (1): 1-21.
- 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>.
- Freund, Julien. 1978. « De l'interprétation dans les Sciences sociales ». *Cahiers internationaux de sociologie*, 213-36.
- Gaillard, Gérald. 1993. « Conditions de l'ethnologue ». *Journal des anthropologues* 53 (1): 97-105. <https://doi.org/10.3406/jda.1993.1793>.
- 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>.
- Geertz, Clifford. 1996. *Ici et Là-bas. L'anthropologue comme auteur*. Métailié.
- Ginouès, Véronique, et Isabelle Gras. 2018. *La diffusion numérique des données SHS : guide des bonnes pratiques éthiques et juridiques*. Presses de l'Université de Provence, Coll. « Digitales ». Aix-en-Provence.
- Gitelman, Lisa. 2013. « Raw data is an oxymoron ». CERN Document Server. MIT Press. 5 mars 2013. <https://doi.org/0262518287>.
- Gitelman, Lisa, et Virginia Jackson. 2013. « Introduction. Raw data is an oxymoron ». 1-15.
- Hackett, Edward J., John N. Parker, Niki Vermeulen, et Bart Penders. 2017. « The social and epistemic organization of scientific work ». *The handbook of science and technology studies* 4: 733-64.
- Huyghe, Marie, Laurent Cailly, et Nicolas Oppenheim. 2018a. « Partage des données d'enquêtes en recherche qualitative: approche juridique, éthique, déontologique et épistémologique ».
- . 2018b. « Le méta-projet, retour d'expérience d'une analyse secondaire ». <https://halshs.archives-ouvertes.fr/halshs-01922702>.
- InSHS, 2021. « Les sciences humaines et sociales et la protection des données à caractère personnel dans le contexte de la Science Ouverte : Guide pour la recherche ».

- Jarrige, François, Stefan Le Courant, et Camille Paloque-Bergès. 2018. « Infrastructures, techniques et politiques ». *Tracés. Revue de Sciences humaines*, n° 35: 7-26.
- Jaton, Florian, et Dominique Vinck. 2016. « Processus frictionnels de mises en bases de données ». *Revue d'anthropologie des connaissances* Vol. 10, n° 4 (4): 489-504.
- Katz, Jack. 2001. « From how to why: On luminous description and causal inference in ethnography (Part I) ». *Ethnography* 2 (4): 443-73.
- . 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>.
- Latour, Bruno. 1993. « Le topofil de Boa-Vista ». *Raisons pratiques* 4: 187-216.
- Le Marec, Joëlle, et François Mairesse. 2017. *Enquête sur les pratiques savantes ordinaires. Collectionnisme numérique et environnements matériels*. Le bord de l'eau. Lormont. <https://www.decitre.fr/livres/enquete-sur-les-pratiques-savantes-ordinaires-9782356875143.html>.
- Leonelli, Sabina, Brian Rappert, et Gail Davies. 2017. « Data Shadows: Knowledge, Openness, and Absence ». *Science, Technology, & Human Values* 42 (2): 191-202. <https://doi.org/10.1177/0162243916687039>.
- Levain, Alix, Sandrine Dupé, Valérie VIAUD, Florence Revelin, Marine Legrand, Quentin Toffolini, Hervé Squidant, Virginie Parnaudeau, Marianne Cerf, et Rodéric Bera. 2021. « Paroles et chemins de l'agriculture littorale. Rapport final. » Research Report. INRAE LISIS ; CNRS AMURE ; INRAE SAS. <https://hal.archives-ouvertes.fr/hal-03202550>.
- Levin, Nadine, et Sabina Leonelli. 2017. « How Does One "Open" Science? Questions of Value in Biological Research ». *Science, Technology, & Human Values* 42 (2): 280-305. <https://doi.org/10.1177/0162243916672071>.
- Masclat De Barbarin, Marie. 2018. « Préface - La diffusion numérique des données en SHS - Guide de bonnes pratiques éthiques et juridiques ». In *La diffusion numérique des données en SHS - Guide de bonnes pratiques éthiques et juridiques*, édité par Véronique Ginouvès et Isabelle Gras, 9-16. Digitales. Presses universitaires de Provence. <https://hal-amu.archives-ouvertes.fr/hal-01916809>.
- 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>.
- Millerand, Florence. 2011. « Le partage des données scientifiques à l'ère de l'e-science : l'instrumentation des pratiques au sein d'un collectif multidisciplinaire ». *Terrains & travaux* n° 18(1):215-37.
- Mohia, Nadia. 2008. *L'expérience de terrain: pour une approche relationnelle dans les sciences sociales*. La Découverte.
- Molinié, Antoinette, et Marie-Dominique Mouton. 2008. « L'ethnologue aux prises avec les archives - Introduction ». *Ateliers d'anthropologie. Revue éditée par le Laboratoire d'ethnologie et de sociologie comparative*, n° 32 (août). <https://doi.org/10.4000/ateliers.1093>.
- Moore, Samuel A. 2017. « A Genealogy of Open Access: Negotiations between Openness and Access to Research ». *Revue Française Des Sciences de l'information et de La Communication*, n° 11 (août). <https://doi.org/10.4000/rfsic.3220>.
- 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.
- Mouton, Marie-Dominique. 2001. « Archiver la mémoire des ethnologues ». *Gradhiva: revue d'histoire et d'archives de l'anthropologie* 30 (1): 67-72.

———. 2018. *Dématérialisation et valorisation des matériaux de terrain des ethnologues: l'archiviste face aux questions éthiques*. Presses universitaires de Provence.

Olivier de Sardan, Jean-Pierre. 2004. « La rigueur du qualitatif. L'anthropologie comme science empirique ». *Espace Temps* 84 (1): 38-50. <https://doi.org/10.3406/espat.2004.4237>.

———. 2008. *La Rigueur du Qualitatif: Les Contraintes Empiriques de l'Interprétation Socio-anthropologique*.

Olivier de Sardan, Jean-Pierre. 1995. « La politique du terrain. Sur la production des données en anthropologie ». *Enquête. Archives de la revue Enquête*, n° 1 (octobre): 71-109. <https://doi.org/10.4000/enquete.263>.

———. 1996. « La violence faite aux données. De quelques figures de la surinterprétation en anthropologie ». *Enquête. Archives de la revue Enquête*, n° 3 (novembre): 31-59. <https://doi.org/10.4000/enquete.363>.

Pasquetto, Irene V., Ashley E. Sands, et Christine L. Borgman. 2015. « Exploring Openness in Data and Science: What Is "Open," to Whom, When, and Why? » *Proceedings of the Association for Information Science and Technology* 52 (1): 1-2. <https://doi.org/10.1002/pra2.2015.1450520100141>.

Pasquetto, Irene V., Ashley E. Sands, Peter T. Darch, et Christine L. Borgman. 2016. « Open Data in Scientific Settings: From Policy to Practice ». In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, 1585-96. San Jose California USA: ACM. <https://doi.org/10.1145/2858036.2858543>.

Pouyllau, Stéphane. 2018. « Éditorial - La diffusion numérique des données en SHS - Guide de bonnes pratiques éthiques et juridiques ». In *La diffusion numérique des données en SHS - Guide de bonnes pratiques éthiques et juridiques*, édité par Véronique Ginouvès et Isabelle Gras, 299-302. Digitales. Presses universitaires de Provence. <https://hal-amu.archives-ouvertes.fr/hal-01974527>.

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>.

Revelin, Florence, Anne-Gaëlle Beurier, Hervé Squidant, et Alix Levain. 2019. « Parchemins : espace d'expérimentation, espace d'enquête sur l'ouverture des données qualitatives. » In *Journée d'étude MSH Bretagne - Données qualitatives : partager, archiver, ré-analyser. Où en est-on côté chercheur ?* Rennes, France. <https://hal.archives-ouvertes.fr/hal-03202753>.

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

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>.

Shapin, Steven, Simon Schaffer, Thierry Piélat, et Simon Schaffer. 1993. *Léviathan et la pompe à air: Hobbes et Boyle entre science et politique*. Ed. La Découverte.

Simukovic, Elena, Maxi Kindling, et Peter Schirmbacher. 2013. « Umfrage zum Umgang mit digitalen Forschungsdaten an der Humboldt-Universität zu Berlin », 56. <https://doi.org/10.18452/13568>.

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*, n° 9 (mars). <http://journals.openedition.org/ctd/3791>.

Stérim, Anne-Laure. 2018. *Diffuser des données de la recherche dans le respect du droit et de l'éthique*. Presses universitaires de Provence. <https://doi.org/10/document>.

Stérim, Anne-Laure, et Camille Noûs. 2019. « Ouverture des données de la recherche: les mutations juridiques récentes ». *Tracés. Revue de Sciences humaines*, n° 19: 37-50.

- Thévenot, Laurent. 1986. « Les investissements de forme ». *Conventions économiques* 29: 21-71.
- Vapnarsky, Valentina. 2019a. « Des communautés sources aux communautés d'experts: l'expérience SAWA (Savoirs Autochtones Wayana-Apalai de Guyane) ». *Culture et recherche*, n° 140: p-71.
- . 2019b. Petites aventures et grands défis de la restitution patrimoniale interculturelle: quelques réflexions à partir d'une expérience wayana (Guyanes). Presses Universitaires Paris Nanterre.
- . 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>.
- Weil, François, et Chantal Cases. 2012. « La structuration de la recherche en sciences humaines et sociales », 55.
- Weller, Jean-Marc. 2019. « La fabrique des données. À propos de l'ouvrage de Jérôme Denis, Le Travail invisible des données. Elements pour une sociologie des infrastructures scripturales, Paris, Presses des Mines, 2018. » In *Gérer et Comprendre. Annales des Mines*.
- Wilkinson, Mark D., Michel Dumontier, IJsbrand Jan Aalbersberg, Gabrielle Appleton, Myles Axton, Arie Baak, Niklas Blomberg, et al. 2016. « The FAIR Guiding Principles for Scientific Data Management and Stewardship ». *Scientific Data* 3 (mars): 160018. <https://doi.org/10.1038/sdata.2016.18>.
- Wyatt, Sally. 2017. « Making Policies for Open Data: Experiencing the Technological Imperative in the Policy World ». *Science, Technology, & Human Values* 42 (2): 320-24. <https://doi.org/10.1177/0162243916687645>.
- Zeitlyn, David. 2019. « Des archives pour l'anthropologie : futurs possibles et passés contingents. L'archive comme substitut anthropologique ». *Ateliers d'anthropologie. Revue éditée par le Laboratoire d'ethnologie et de sociologie comparative*, mars. <https://doi.org/10.4000/ateliers.10817>.