



**HAL**  
open science

## Partage des données d'enquêtes en recherche qualitative : approche juridique, éthique, déontologique et épistémologique

Marie Huyghe, Laurent Cailly, Nicolas Oppenheim

### ► To cite this version:

Marie Huyghe, Laurent Cailly, Nicolas Oppenheim. Partage des données d'enquêtes en recherche qualitative : approche juridique, éthique, déontologique et épistémologique . [Rapport de recherche] CITERES. 2018. halshs-01790983

**HAL Id: halshs-01790983**

**<https://shs.hal.science/halshs-01790983>**

Submitted on 6 Jun 2018

**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.

# **PARTAGE DES DONNEES D'ENQUETES EN RECHERCHE QUALITATIVE : APPROCHE JURIDIQUE, ETHIQUE, DEONTOLOGIQUE ET EPISTEMOLOGIQUE**

Marie Huyghe, Laurent Cailly, Nicolas Oppenheim

## **Introduction**

Ce rapport présente une synthèse des réflexions menées au sein du Métaprojet<sup>1</sup> sur les enjeux du partage des données de la recherche, dans le champ spécifique des études qualitatives portant sur les mobilités quotidiennes et les modes d'habiter. À partir d'une revue de littérature sur les expériences de partage de données menées en France et à l'étranger, nous livrons ici une réflexion programmatique sur les conditions suivant lesquelles pourrait être constitué un outil de partage dans le champ des mobilités spatiales. Pour ce faire, nous mettons en évidence et proposons de questionner la tension entre l'intérêt du partage des données d'enquêtes individuelles – notamment pour l'approfondissement des connaissances sur les mobilités spatiales – et les freins et/ou éléments de blocage importants liés au caractère personnel des données, au rapport singulier des chercheurs aux personnes interrogées (rapport de confiance) et, au-delà, à leurs données d'enquêtes (droits). Dans cette perspective, nous nous proposons d'examiner les conditions du partage des données d'enquête. Celles-ci soulèvent de multiples questions d'ordre éthique, déontologique, juridique et, plus largement méthodologique, qu'il convient de mêler. Par ailleurs, cette réflexion sur le partage des données doit inclure un questionnement très pragmatique sur les implications de sa mise en œuvre pour le chercheur et les difficultés concrètes qu'il peut rencontrer. Celui-ci est soumis à des contraintes de temps ou de moyens, mais il s'inscrit également dans un champ, au sens de Bourdieu, dans lequel la concurrence pour obtenir la reconnaissance symbolique des pairs et/ou du grand public est très importante. Cette concurrence n'est pas sans effets sur la manière dont les chercheurs se positionnent vis-à-vis du partage des données.

De manière générale, la question de la diffusion, de la mutualisation et du retraitement des données qualitatives en SHS concerne aujourd'hui un nombre croissant de chercheurs, en raison des exhortations croissantes au partage des données de la recherche. Dans un contexte international de promotion de l'*open data* et du *data sharing* aujourd'hui relayée par les institutions européennes et les grands instituts de recherche en France, l'ouverture et le partage des données sont présentés comme un idéal à atteindre en termes de capitalisation, de transparence de la démarche scientifique et d'administration scientifique de la preuve ; ils sont censés participer de l'avancée de la recherche en sciences humaines et sociales. Ainsi le CNRS, tout comme l'INRA et le CIRAD (2016), enjoignent les chercheurs à s'engager « *dans le mouvement mondial de partage de données scientifiques ouvertes* » (COMETS, 2015), qui s'est déjà largement imposé en Grande-Bretagne. Malgré ces politiques incitatives, la France se distingue par un retard relatif mis en évidence dans le rapport de Cribier & Feller (2003). Néanmoins, depuis les années 2000, à la suite de grands programmes comme Qualidata au Royaume-Uni, les expériences de partage, généralistes (Quetelet) ou thématiques (Verbatim (Le Roux, 2005), BeQuali, Quetelet, Archethno), se multiplient. Des enquêtes menées auprès de chercheurs des universités de Rennes et Lille révèlent par ailleurs l'intérêt croissant des institutions

---

<sup>1</sup> Le méta-projet, programme de recherche « Mutualisation et méta-analyse de corpus qualitatifs sur les mobilités quotidiennes dans les territoires dépendants de l'automobile » se donne pour objectif de réinvestir six opérations de recherche portant sur les mobilités et réalisées au sein du laboratoire CITERES (UMR 7324, Université de Tours), à partir de la confrontation et du réexamen de l'ensemble des matériaux produits, soit un corpus de plus de 250 entretiens.

locales (universités, laboratoires) pour la conservation et la diffusion des données de la recherche. Ces éléments montrent que le partage des données n'est plus seulement une injonction venue d'en-haut mais – de manière croissante – une pratique de recherche désormais impulsée « par le bas » émanant de chercheurs animés par des objectifs de capitalisation et de réutilisation des matériaux d'enquête.

La genèse de notre recherche est en elle-même révélatrice de cette tendance et constitue une opération volontaire d'appropriation de cette démarche. En effet, à l'origine, différents chercheurs du laboratoire CITERES (Université de Tours) ont manifesté le souhait d'échanger sur leurs thématiques de recherche, en l'occurrence la mobilité quotidienne dans les territoires urbains, périurbains et ruraux. Au sein de ce laboratoire, depuis une dizaine d'années, six programmes de recherche avaient porté sur ces questions, sans que jamais (ou rarement) il n'y ait d'échanges formels sur leur contenu et leurs conclusions respectives. Progressivement, une idée a germé : celle "d'utiliser" (et mutualiser) les données produites par les différents collègues pour aller plus loin dans nos questionnements respectifs ou conjoints.

Ces objectifs scientifiques nous ont amenés à construire un nouveau programme de recherche, le "méta-projet", qui a dépassé rapidement ses ambitions initiales. Au-delà d'une simple mutualisation en interne, il nous a conduits à poser la question des conditions épistémologiques, juridiques et déontologiques d'une mutualisation d'enquêtes qualitatives dans le champ des mobilités quotidiennes. À partir d'une simple réflexion sur la mise en commun entre chercheurs proches, la problématique du partage des données qualitatives à une échelle plus vaste est donc devenue particulièrement centrale dans notre recherche.

Cet objectif (présenté ici) implique de reprendre les grandes questions relatives au partage tout en portant une réflexion propre à nos données d'enquête, en définissant un positionnement et des modalités opératoires spécifiques à nos recherches sur les mobilités quotidiennes. À cette fin, nous reviendrons dans un premier temps sur les objectifs du partage des données qualitatives, à un niveau général et à celui du champ de recherche sur les mobilités. Dans un deuxième temps, nous examinerons les conditions épistémologiques, juridiques et déontologiques de ce partage des données, ce qui nous amènera notamment à réfléchir aux différentes méthodes d'anonymisation des matériaux. Cette réflexion nous permettra d'exposer dans un troisième temps les conditions de diffusion et de partage que nous envisageons dans notre projet, avant d'aborder les facteurs de blocage et les verrous à lever pour promouvoir le partage des données, en particulier à l'intérieur du champ de recherche sur la mobilité.

## **Données, archivage, partage, diffusion, mise en commun... de quoi parlons-nous exactement ?**

### *Données (de la recherche)*

Selon l'OCDE, « les données de la recherche sont définies comme des enregistrements factuels (chiffres, textes, images et sons), qui sont utilisés comme sources principales pour la recherche scientifique et sont généralement reconnus par la communauté scientifique comme nécessaires pour valider des résultats de recherche ». Cette définition est discutée en SHS où les données constituent certes des informations-sources mobilisées ou collectées qui sous-tendent l'analyse et la production de résultats de recherche ; mais ces données sont avant tout "produites" par le regard et le questionnement scientifique du chercheur ainsi que par les procédures d'enquête qu'il met en œuvre. Cette conception, suivant laquelle les données ne sont justement pas "immédiatement données" mais des "données produites" n'enlève en rien à la valeur qu'elles prennent dans le travail des chercheurs. Bien au

contraire, le fait que les données soient “produites” et puissent être considérées comme constitutives d’une recherche, leur confère une valeur intrinsèque et pose la question de leur partage et de leur conservation. Si la valeur des données fait pour l’instant l’objet d’une faible reconnaissance institutionnelle, les enquêtes menées auprès des chercheurs révèlent leur attachement aux données produites. Cet attachement se mesure par exemple à partir de l’angoisse que suscite l’idée de la perte ou du vol des données.

### *Données qualitatives*

La notion de “donnée qualitative”, utilisée ici, relève de l’abus de langage. Il n’y a pas de données strictement qualitatives, mais des données produites et/ou interprétées dans le cadre de démarches qualitatives qui visent à rendre compte, en compréhension, des phénomènes sociaux ou socio-spatiaux. Dans le champ d’analyse des mobilités spatiales, les évolutions conceptuelles ont amené les chercheurs à définir la mobilité - au delà du déplacement dans l’espace - comme un rapport matériel, idéal ou encore potentiel au changement de lieu (Lévy, 2001 ; Kaufmann, 2004). Ce nouveau paradigme a donné aux approches qualitatives de la mobilité une importance considérable et assuré le succès des techniques d’enquête habituellement privilégiées dans ce type de démarche telles les entretiens semi-directifs ou les observations ethnographiques. Personne ne sera ainsi tout à fait dérangé par le fait de considérer - par extension - que le matériau discursif recueilli au cours d’un entretien, ou celui, descriptif, recueilli en situation d’observation, puisse être qualifié de qualitatif.

L’association entre le type de donnée et la démarche est toutefois moins évidente pour d’autres types de matériaux. Les enregistrements GPS de plus en plus communément exploités pour approcher les mobilités spatiales se situent précisément dans ce cas de figure puisqu’ils peuvent être mobilisés dans différentes démarches de recherche. La question d’un traitement qualitatif d’un enregistrement qui par certains aspects est constitué de données mesurables (la quantité de points enregistrés) se pose donc et nous rappelle que c’est bien la démarche (et non la donnée) en elle-même qui confère au matériau son statut ; ainsi, nous considérons ici comme qualitatifs l’ensemble des matériaux produits et/ ou exploités dans le cadre d’une stratégie de recherche visant à saisir en compréhension les phénomènes.

### *Stockage et archivage*

Dans les réflexions actuelles sur la gestion des données de la recherche en sciences humaines et sociales, une première question porte sur la conservation des données. Promue par les grandes institutions de recherche, elle amène à envisager le stockage et, au-delà, l’archivage comme une pratique de recherche à systématiser. Dans les faits, on peut distinguer le stockage (pour une certaine durée, en vue d’un partage et/ou d’une réutilisation des données par exemple) et “l’archivage définitif” qui concerne les données “*présentant un intérêt historique, scientifique ou statistique justifiant qu’elles ne fassent l’objet d’aucune destruction*”. Archiver consiste donc à conserver dans des infrastructures dédiées (telles la plateforme Huma-Num<sup>2</sup> ou le système PAC – Plateforme d’Archivage au CINES) des données qui ne seront ni consultables ni utilisables à court terme, hormis par le chercheur-dépositaire. Les données concernées par l’archivage sont les données brutes, telles que collectées sur le terrain : elles n’ont en particulier pas besoin d’être anonymisées. L’objectif de l’archivage est patrimonial : ces banques d’archivage doivent permettre à terme aux historiens des sciences de « *reconstituer et mieux comprendre les logiques du travail scientifique en SHS ainsi que leurs évolutions* » (Duchesne & Garcia, 2014). Le stockage, quant à lui, répond aux objectifs de réutilisation par le chercheur-dépositaire ou par d’autres chercheurs, à court ou moyen termes. Il

---

<sup>2</sup> <http://www.huma-num.fr/>

implique la plupart du temps une sélection et une transformation du corpus de données : choix de certains documents, anonymisation, etc.

### *Ouverture (ou partage) des données*

L'ouverture (ou partage) des données de la recherche, auxquels nous consacrons ce rapport, diffère de l'archivage *stricto sensu* de par sa temporalité : il vise à mettre à disposition des chercheurs des données à (ré)analyser sans temps de carence ou d'embargo systématiques. En raison de règles juridiques et d'enjeux déontologiques liés au partage des données (cf ci-dessous), celles-ci doivent notamment être impérativement anonymisées avant d'être mises en partage. En termes d'infrastructures, le partage nécessite la création de plateformes de stockage.

### *Réutilisation des données*

Diffusion ou ouverture à la réutilisation ? Opérée par la CNIL, cette distinction est d'importance : d'elle dépendent notamment les règles juridiques qui s'appliquent au chercheur-dépositaire des données. La diffusion désigne l'ouverture des données en *open data*, i.e. librement et sans restriction d'accès ; l'ouverture à la réutilisation désigne le partage des données uniquement dans la sphère académique et à des fins scientifiques.

Enfin, la mutualisation et l'analyse secondaire concernent les différentes réutilisations possibles des données. Dans le cadre du méta-projet, nous opérerons à la fois une mutualisation, i.e. une mise en commun et la constitution d'un « méta-panel » à partir de l'ensemble des données à notre disposition, et une analyse secondaire, i.e. une réanalyse de données produites antérieurement par d'autres chercheurs de l'équipe.

# I. Partage des données de la recherche : historique et objectifs

## 1. Les objectifs du partage de données qualitatives

En Angleterre, les premières réflexions sur le partage des données et l'analyse secondaire remontent à une vingtaine d'années : dès 1995, l'ESRC (Economic and Social Research Council, qui finance 55 % des recherches au Royaume-Uni) a imposé aux chercheurs qu'elle finançait de mettre à disposition les données produites afin qu'elles soient archivées. En réponse à cette demande a été créée Qualidata, première structure d'archivage et de partage de données qualitatives à l'échelle nationale (Corti, 2000), qui a ensuite donné lieu à la création du UK Data Archive ; celui-ci propose aujourd'hui plus de 1000 collections de données (Bishop & Kuula-Luumi, 2017). Depuis le début des années 2000, le partage des données de la recherche est largement encouragé au Royaume-Uni, à la fois par les financeurs (qui exigent de pouvoir accéder facilement ou librement aux données produites, et demandent aux chercheurs des plans de gestion de données qui allient qualité, soutenabilité, accessibilité et disponibilité des données de la recherche) ; par les revues scientifiques (qui demandent à pouvoir accéder aux données utilisées dans les articles proposés par les chercheurs) ; et par le gouvernement britannique (qui souhaite aller vers plus de transparence dans la recherche et vers une meilleure optimisation des données produites). Face à ces demandes et encouragements, le partage des données est permis et soutenu par le développement de politiques, de moyens humains et matériels appropriés.

*“Twenty years ago, most researchers would not have even considered sharing qualitative data; now, at least in the United Kingdom, there is a growing belief that much data can be shared and reused”* (Bishop & Kuula-Luumi, 2017 : 12). Après 20 ans de réflexions sur le sujet, le partage des données et l'analyse secondaire semblent rentrer dans la culture et les pratiques des chercheurs britanniques, l'analyse secondaire de données qualitatives étant de plus en plus considérée comme une méthodologie à part entière. Une analyse menée par Bishop & Kuula-Luumi en 2017 a ainsi montré que 3/5 des collections de données étaient utilisées, majoritairement (64 %) à des fins d'apprentissage et d'enseignement (seulement 15 % à des fins d'analyse secondaire), par des étudiants du supérieur.

En termes de partage des données de la recherche, l'Angleterre semble être une exception européenne, y compris face à des pays relativement avancés sur la question (Finlande, Suisse, Allemagne) dont les infrastructures dédiées et le nombre d'enquêtes proposées en partage sont de bien moindre envergure (par exemple, Bishop & Kuula-Luumi dénombrent seulement 78 collections de données en Allemagne, 10 collections en Suisse et Slovaquie).

En France, les réflexions autour du partage des données de la recherche sont beaucoup plus récentes. En SHS, malgré un rapport pionnier sur la question (Cribier & Feller, 2003) et quelques expériences au succès d'ailleurs limité, *“les données constituent encore largement un “impensé”* (Serres *et al*, 2017). L'émergence de ces réflexions en France est imputable au contexte de normalisation des pratiques de recherche opérée par les instances européennes qui, au nom de la diffusion et de l'internationalisation des connaissances, assure la promotion du *data sharing*. Elle s'explique également en partie par le contexte économique tendu que connaît la recherche publique, qui implique d'optimiser la gestion des données. En effet, les principaux organismes de recherche publique semblent ne s'être emparés que récemment de la question, au vu des rapports publiés sur les enjeux éthiques du partage des données (CIRAD & INRA, 2016 ; COMETS, 2015). Ceux-ci encouragent les chercheurs à s'engager dans ce « *mouvement mondial de partage de données scientifiques ouvertes* » (COMETS, 2015), ce qui nécessite notamment des formations (« *aux dimensions éthiques de la gestion des données, à l'évolution des règles juridiques concernant le partage responsable de données utilisées* ») ou la création de comités éthiques, par discipline. En outre, le Comité de Concertation pour les Données en SHS (CCDSHS) créé en 2001 auprès des ministères français chargés de l'économie, de

l'emploi, de l'éducation nationale et de la recherche, et chargé de proposer des "*orientations nationales pour une politique publique de données pour la recherche en SHS*", encourage le partage des données de la recherche. Il est notamment à l'origine de la création du GIS Quételet, base d'enquêtes (notamment grandes enquêtes statistiques) en Sciences Humaines et Sociales. Hormis cet exemple, les quelques structures de partage (voire d'archivage) de données qui existent en France sont relativement récentes : on citera par exemple beQuali, banque d'enquêtes qualitatives en Sciences humaines et sociales créée en 2006, ou Isidore, plateforme de recherche permettant l'accès aux données numériques des sciences humaines et sociales créée en 2009. Du côté de l'Agence Nationale de la Recherche, et au regard de ce que fait l'ESRC britannique, la question du partage des données n'est toujours pas une priorité ; dans son appel à projets de 2018, il est simplement mentionné que "*concernant les données de recherche, l'ANR n'impose pas de plan de gestion des données dans l'appel générique 2018. Elle attire cependant l'attention des déposants sur l'importance de considérer la question des données de recherche au moment du montage et tout au long du projet.*" Enfin, du côté des universités et des laboratoires, en l'absence d'obligations chacun s'empare de la question de la gestion des données de la manière et au rythme qu'il le souhaite. La rédaction d'un Plan de Gestion de Données ("*document rédigé au commencement d'un projet de recherche et qui définit ce que les chercheurs feront de leurs données pendant et après le projet, explicitant notamment la mise à disposition des données*") (Commission Européenne, 2013), dont l'intérêt est désormais reconnu, est préconisée : ainsi, elle est par exemple obligatoire pour tous les projets dans le cadre du programme H2020. Néanmoins, dans la recherche française, combien de projets se sont aujourd'hui dotés d'un tel outil ? En revanche, plusieurs universités comme Strasbourg, Lille 3 et Rennes 2 ont réalisé en 2015 et 2016 des enquêtes sur les données de la recherche (Rege, 2015 ; Prost & Schöpfel, 2015 ; Serres et al, 2017) qui ont par exemple conduit à l'élaboration d'un livre blanc sur les données de la recherche dans les thèses de doctorat (Chaudiron *et al*, 2015). Ces travaux récents révèlent que les institutions de recherche, et une partie des chercheurs, s'emparent de la question du partage des données.

Avant de discuter ci-après des enjeux et des modalités du partage de données (et, *a fortiori*, de l'analyse secondaire), il s'agit d'abord d'en réinterroger les finalités et les intérêts supposés. Pour Mauthner & Parry (2009: 298), "*data preservation and sharing lead to "better" science through scientific transparency, scientific innovation and scientific accuracy*".

De fait, le partage des données de la recherche revêt d'une part des intérêts scientifiques. D'après Corti (2011), il permet des réutilisations à des fins de comparaisons (avec d'autres contextes, d'autres périodes, d'autres groupes sociaux, territoires ou cultures) : pour les chercheurs interrogés à Rennes 2 (Serres *et al*, 2017 : 17), le "*souhait de voir se développer d'autres recherches à partir de ses jeux de données*" est d'ailleurs l'une des principales justifications de la mise en partage. L'analyse secondaire permet également de renforcer la validité des analyses en permettant de travailler sur des échantillons plus larges ou diversifiés. Par manque de temps ou parce que le programme de recherche ne le demande pas initialement, les données qualitatives sont rarement exploitées dans leur intégralité : partager et réutiliser ces données ouvre la possibilité d'expurger le matériau ou encore de l'interroger selon de nouvelles questions de recherche, à partir de nouveaux angles de recherche ou méthodologies d'analyse (Cribier & Feller, 2003). En outre, pour Mauthner & Parry (2009), le partage et l'analyse secondaire des données de la recherche renforcent la « *fiabilité, validité, vérifiabilité et plausibilité* » des analyses à travers la « *distance, le détachement et l'objectivité* » du chercheur-secondaire. Enfin, donner à voir les données de la recherche a une fonction d'administration de la preuve (Bertaux, 2010) en rendant possible par des tiers une validation des résultats, par exemple par des *reviewers* ou des jurys de thèse ; certains souhaitent développer la pratique de *l'active citation* (Moravcisk, 2010 : 31), l'idée étant que toutes les citations reproduites dans des travaux qualitatifs soient reliées, par hyperliens, à l'ensemble des données préalablement mises à disposition. Ceci participerait de la

transparence de la recherche en sciences sociales, jugée nécessaire dans un domaine où aucun « *standard de contrôle qualité* » n'existe (Corti, 2000 : 28). Partager les données de la recherche devrait même permettre une réplique des analyses : plus tard, ailleurs, sur une population qui a d'autres caractéristiques, donc de confronter la validité des hypothèses sur d'autres corpus.

L'enquête menée à l'Université de Rennes 2 (Serres *et al.*, 2017 : 13) montre également que les chercheurs qui partagent leurs données avec d'autres chercheurs le font dans un objectif qu'ils qualifient de "social", pour accroître la visibilité de leurs travaux de recherche, pas seulement pour des raisons personnelles, mais pour nouer des contacts et favoriser les coopérations scientifiques. Le partage des données peut alors être considéré comme un élément qui conditionne le développement d'une position, d'un champ et/ou réseau scientifiques.

Les finalités du partage des données sont également d'ordre méthodologique : analyser la façon dont on « *a fait la recherche* » (dont on a identifié un panel d'enquête, collecté des données, rédigé un guide d'entretien, etc.) dans des recherches antérieures peut aider à développer de nouvelles méthodologies ou outils de recherche. Pour Corti (2011), les informations méthodologiques les plus intéressantes proviennent notamment des notes ou carnets de terrain des chercheurs, qui sont rarement accessibles au moment de la publication des résultats (souvent publiés sous la forme de « *polished policy reports* » - Corti, 2011) mais sont donnés à voir en cas de partage des données. On peut également penser que le partage des données met en visibilité la diversité des approches et des méthodes d'analyse utilisées dans les différentes disciplines; et d'une certaine manière, en accentuant les enjeux d'explicitation des méthodes, un vecteur d'innovation.

Partager les données de la recherche est également intéressant en termes pédagogiques, pour l'enseignement des approches et méthodologies qualitatives. Des enseignants ayant utilisé des données existantes dans leurs cours ont ainsi noté que les étudiants ont mieux appris les méthodes qualitatives s'ils pouvaient se baser sur des données réelles et qui les intéressaient (Bishop & Kuula-Luumi, 2017). En outre, faire travailler les étudiants ou leur permettre de travailler à partir de données existantes permet de libérer du temps d'enseignement ou de réflexion pour se concentrer sur la formulation de questions de recherche ou l'analyse des données.

Par ailleurs, partager les données de la recherche et les valoriser de façon optimale est intéressant d'un point de vue financier, en particulier pour les données produites sur fonds publics ; produire des analyses secondaires à partir de données existantes évite en outre de passer du temps (et de l'argent) à en produire de nouvelles (Corti & Bishop, 2005).

Enfin, les chercheurs ayant répondu à l'enquête menée par l'Université de Rennes 2 sont 1/3 (environ 40) à justifier leurs pratiques en évoquant le facteur institutionnel, i.e. l'obligation faite par le financeur de la recherche à partager les données produites. Dans ce contexte, il est évident que l'objectif de partage des données, encore peu promu en France, gagnerait à être appuyé par des dispositifs institutionnels, à la fois dans le sens d'une incitation (par exemple à la mise en œuvre d'un plan de gestion des données dans les programmes scientifiques financés sur fonds publics) et d'une reconnaissance scientifique et/ou symbolique des opérations de recherche fondées sur le partage et la mise en œuvre d'analyses secondaires. La promotion de cet objectif implique toutefois d'en préciser les enjeux et d'en discuter les modalités.

## **2. Objectifs du partage dans le cas spécifique de nos données thématiques et de structuration du champ de recherche sur les mobilités**

Au sein du Méta-projet, nous avons été amenés à réévaluer les enjeux d'une telle démarche et à identifier les éléments qui nous portent à réfléchir et à mettre en œuvre une procédure d'archivage et de diffusion des données dans le champ des mobilités spatiales. Outre les arguments énoncés plus haut quant à la conservation patrimoniale des matériaux recueillis, aux exigences de transparence



scientifique, aux perspectives de capitalisation ou de réutilisation, plusieurs pistes de légitimation de cette démarche méritent d'être développées. Celles-ci sont reliées à un objectif de structuration d'un champ de recherche.

La première porte sur l'intérêt majeur que revêtent l'indexation et la mise en visibilité des matériaux primaires. Depuis une vingtaine d'années, en SHS, les travaux qualitatifs sur la mobilité quotidienne et les modes d'habiter sont nombreux. Les discussions scientifiques et la capitalisation des connaissances s'effectuent au niveau des interprétations et des "résultats". Hormis quelques *verbatim* qui, dans les publications, servent à administrer la preuve, le matériau d'enquête (entretiens, photos, carnets de terrain, etc.) reste dans l'ombre, en "caché" et retombe dans l'oubli peu après la publication des résultats de recherche. Il y a comme une incohérence à considérer que l'enquête *stricto sensu* est une situation de co-production scientifique tout en maintenant ce niveau essentiel de l'élaboration de la connaissance scientifique dans l'ombre de l'arrière-cuisine. Dans le champ qui est le nôtre, combler ce hiatus viserait à reconnaître que la collecte et la publicisation d'enquêtes situées sur les pratiques spatiales et la mobilité, est en soi, un acte de connaissance. De la même manière que le sociologue Ferrarotti (1983) voyait dans les histoires de vie individuelles « *un matériau primaire bien supérieur aux items décharnés, précodifiés, programmés en fonction de l'analyse futur de la méthode quantitative, qui réduit la réalité sociale* », on peut considérer que les matériaux qualitatifs décrivant la gestion des distances par les déplacements et l'expérience ordinaire de l'espace, et au-delà, le rapport aux lieux et les conditions ordinaires de l'habiter, présentent une valeur inestimable et permettent d'accéder à la compréhension des formes de territorialisation de la vie sociale (Barel, 1984) qui justifie à elle seule de les conserver.

Le deuxième enjeu du partage des matériaux d'enquête porte sur la transparence de la démarche, et au-delà, sur l'explicitation des méthodes d'enquêtes. Partager implique de détailler les conditions et les modalités précises suivant lesquelles les matériaux ont été produits ; cet objectif d'explicitation de la collecte scientifique - comme forme "d'expérience" - prend particulièrement sens dans le champ des mobilités où les chercheurs sont généralement peu diserts sur les choix opérés. Les modalités d'enquête, le statut accordé à l'entretien - donc à la parole de l'acteur - ou encore les méthodes utilisées pour l'interprétation ne sont que très rarement développées. C'est d'autant plus étonnant que le champ des mobilités spatiales se caractérise par la mise en avant récurrente par les chercheurs d'innovations méthodologiques (les méthodes mobiles, les entretiens embarqués, etc.).

La logique de diffusion et de partage devrait mettre davantage au jour les contextes épistémologiques et méthodologiques d'une enquête, et ce faisant, permettre, sur de meilleures bases, la discussion autour des résultats. Un tel partage serait extrêmement riche dans le champ des mobilités spatiales, pluridisciplinaire, où les chercheurs n'ont pas nécessairement les mêmes méthodes d'enquête, la même manière de réaliser les entretiens ou un travail ethnographique.

Une autre vertu, très attendue, relève de l'innovation méthodologique. A partir du moment où le matériau primaire acquiert - par le jeu de la diffusion - une valeur intrinsèque, la qualité des procédures de collecte et d'interprétation des matériaux devient prééminente dans l'évaluation d'une recherche. Force est de reconnaître aujourd'hui que - faute d'accès au matériau - cette dimension est un point relativement aveugle de l'évaluation scientifique. Alors que l'accès au matériau ne peut que solidifier la cohérence d'une démarche et, en définitive, consolider l'argumentaire.

Enfin, un dernier point, non exclusif, porte sur les enjeux de la réutilisation. Dans le champ des mobilités, deux enjeux scientifiques donnent une valeur considérable à la possible réutilisation des données d'enquête. La première est liée à la promesse des analyses comparatives entre territoires distincts. Une des questions majeures de la recherche périurbaine, par exemple, est celle de la diversité des contextes résidentiels et de leurs inférences sur les modes d'habiter. Suivant les géotypes de résidence, les systèmes de mobilité n'obéissent ni aux mêmes structures ni aux mêmes stratégies. La comparaison des enquêtes, dans divers contextes socio-spatiaux, doit permettre de mieux comprendre

l'articulation entre territoire et construction des modes de vie. Une deuxième piste, tout aussi féconde, tient aux promesses de l'approche diachronique. La réflexion croissante sur les changements de comportements (en relation, par exemple, avec les enjeux environnementaux globaux) rend déterminant le fait de pouvoir mener une nouvelle enquête sur un territoire ou dans un espace sur lequel une enquête a déjà eu lieu. Un archivage et une diffusion des données-enquêtes pourraient démultiplier les démarches de recherche fondées sur la revisite.

En définitive, le partage et la réutilisation des données dans le champ des mobilités spatiales sont riches de promesses. Il convient dès lors d'en discuter les conditions, en commençant par mettre en débat le statut des données et la possible réutilisation des données d'enquête.

### 3. Un positionnement épistémologique qui permet la diffusion des données

Les conceptions épistémologiques de la recherche en sciences sociales prédéterminent fortement le positionnement des chercheurs vis-à-vis de la réutilisation des données. En effet, ces derniers n'accordent pas tous le même statut aux matériaux d'enquête ni la même importance aux conditions dans lesquelles ils ont été collectés. Cette dimension épistémologique du partage des données a suscité de nombreux débats en Angleterre dont Mauthner et Parry (2009) font état. Pour certains chercheurs, la réutilisation des données est source d'objectivité scientifique et confère à l'analyse un surcroît de scientificité en raison de la distance qu'elle induit vis-à-vis du matériau et des contextes de production. C'est le cas, dans la recherche anglo-saxonne, du *foundationnalisme* qui considère que les données sont relativement indépendantes de leur contexte de production : dès lors, plus les « ré-utilisateurs » sont éloignés du terrain d'origine, plus ils sont supposés objectifs et impartiaux pour traiter les données<sup>3</sup>.

Dans le paysage des sciences humaines et sociales, aujourd'hui, cette position est toutefois très marginale. La construction sociale de la réalité et la dimension intersubjective de la connaissance, à laquelle le monde de la recherche n'échappe pas, sont mieux appréciés (Berger et Luckmann, 1966). Dans ces conditions, la posture suivant laquelle les données sont indissociables de leur contexte de production est devenue prééminente. Dans cette perspective, travailler avec moins de données et sans éléments de contexte suffisants produit des analyses beaucoup moins stables. La perte des contextes émotionnels, sensibles et intersubjectifs liés à la situation d'enquête affaiblit considérablement la valeur des matériaux et compromet la réutilisation des données d'enquête. À partir de ce constat, deux postures se dessinent :

- une conception relationnelle suivant laquelle la donnée n'existe que par ses contextes de production : il n'est alors pas possible de séparer une donnée du contexte dans lequel celle-ci a été produite ; à défaut d'avoir vécu la situation d'enquête, il n'est pas envisageable de comprendre et de donner sens au matériau. Ce dernier est, en quelque sorte, incommensurable. L'idée même de diffusion et de réutilisation du matériau est vaine et n'a aucun sens ;

- une conception constructiviste, proche de la précédente mais plus ouverte. Suivant celle-ci, tout processus de recherche, qu'il s'agisse d'une analyse primaire ou secondaire des données, est déterminé par un contexte spécifique d'analyse : l'analyse première des données n'est pas plus légitime que l'analyse secondaire. Le chercheur secondaire analyse les données dans un contexte scientifique différent, ce qui produit des résultats de recherche qui ne seront pas nécessairement les

---

<sup>3</sup> « This view is echoed by Bornat (2005: 13) who argues that when using archived studies the 'interview texts become decontextualised so the secondary analyst is working with less data. This may allow a more 'objective approach' » (Mauthner & Parry, 2009)

mêmes (mais n'invalidant pas nécessairement les premiers résultats). Par exemple, des chercheurs qui analyseraient maintenant des entretiens des années 1960 seraient sûrement plus attentifs aux questions liées au genre que les chercheurs primaires. Il n'en demeure pas moins que la relecture, si elle est orientée par des prismes nouveaux, ne peut se dispenser de prendre en considération le contexte initial de l'enquête. Celui-ci reste déterminant pour fixer la valeur et définir le cadre de ré-analyse du matériau. Suivant cette posture constructiviste, il faut donc s'assurer de médiatiser le contexte de l'enquête, en donnant accès à la méthodologie, à la manière dont les matériaux ont été collectés, à la description des situations, pour permettre la réutilisation.

La conception relationnelle est très forte dans la tradition sociologique, et encore plus dans les travaux ethnographiques et anthropologiques. L'ethnographie narrative, par exemple, insiste sur la singularité subjective de la rencontre entre un chercheur et ses enquêtés, ce qui rend très difficile l'analyse sans la compréhension très fine de la situation d'enquête. Mais, à l'intérieur même de ces disciplines le positionnement des chercheurs vis-à-vis de la nature de leurs données dépend très fortement du modèle scientifique dans lequel ils s'inscrivent et des méthodes qu'ils utilisent. Ainsi, les chercheurs en sciences sociales ont un rapport différent à leurs données selon qu'ils utilisent les méthodes qualitatives ou quantitatives. On peut penser que les quantitativistes sont moins sensibles à la question du contexte de production des données. S'il ne faut pas exagérer l'opposition entre ces deux types de méthodes, les quantitativistes ont souvent tendance à ignorer le contexte de production des données obtenues par questionnaire, notamment le fait que les conditions de passation ne sont pas neutres sur les résultats obtenus (Hugrée & Kern, 2008). À l'inverse, les qualitativistes auront plus tendance à s'inscrire dans la conception relationnelle lorsqu'ils travaillent à partir de l'étude approfondie d'un cas et non par un raisonnement hypothético-déductif (Béliard & Eideliman, 2008). Dans cette dernière démarche, la prise en compte du contexte est donc essentielle (en ethnographie plus que par entretiens par exemple).

La perspective constructiviste constitue une voie médiane retenue ici. Au regard des tensions inhérentes à l'éloignement du chercheur ré-utilisateur vis-à-vis de la situation d'enquête initiale, nous soutenons que la contextualisation des données d'enquête permet d'approcher les contextes de collecte initiaux et, ce faisant, de donner une valeur scientifique à l'analyse secondaire. « *L'ajout de contextes permet de recréer les conditions de l'être là-bas et de compenser la perte de présence* » (Fielding, 2004)<sup>4</sup>. Dès lors, on peut gérer la tension entre l'intérêt de l'analyse secondaire et les difficultés épistémiques qu'elle pose en fournissant une documentation épaisse susceptible d'éclairer les contextes de production des données. Nous rejoignons ici la posture suivie par beQuali qui considère que « *la mise à disposition ne se limite pas aux données de terrain mais repose sur une conception exhaustive de l'enquête* » et produit un document de contextualisation intitulé « l'enquête sur l'enquête » pour chaque enquête mise à disposition.

D'un point de vue épistémologique, il nous semble donc possible de partager les données. Ce premier verrou levé, il faut dès lors étudier les conditions éthiques et juridiques de leur mise en partage.

---

<sup>4</sup> Fielding écrit : « *Si le débat sur les questions épistémologiques liées à l'analyse secondaire nous dit quelque chose, il est très important que les documents archivés contiennent autant d'informations sur le contexte de la collecte de données originales que possible.* »

## II. Enjeux éthique, déontologique et juridique liés à la diffusion des données qualitatives

Au cœur des questionnements sur le partage des données qualitatives se trouve le contrat de confiance, objet juridique et déontologique.

Passé au début de l'entretien et/ou tout au long du terrain ethnographique, ce contrat de confiance fixe avec l'enquêté les conditions dans lesquelles les données sont produites (le contexte de la recherche), les conditions dans lesquelles elles seront utilisées (par qui, quand, dans quel objectif) et, en particulier, partagées (*via* quel média, pendant combien de temps, auprès de quel public). Si, en pratique, l'instauration de ce contrat prend des formes et obéit à des rigueurs variables, elle répond à une prescription méthodologique forte qui est rappelée dans la plupart des manuels d'initiation aux techniques d'enquête. Dans les faits, cette exigence de transparence sur les objectifs de l'enquête et l'utilisation des données constitue une condition du bon déroulement et même de la qualité de l'enquête, car cette confiance est nécessaire pour libérer la parole ou voir la présence du chercheur acceptée au quotidien. Plus largement, il engage le sérieux et le crédit apportés au travail d'enquête. Paradoxalement, à l'heure où les données personnelles sont captées sans beaucoup de protection par les grands opérateurs de la téléphonie et de l'internet, le chercheur se doit de tenir une position très rigoureuse sur la protection des données personnelles qu'il traite, notamment parce qu'il engage un rapport de confiance très personnel qui, pour être fécond doit être débarrassé de toute logique d'intérêt. En vertu de ce contrat de confiance conclu par les deux parties (même en l'absence de preuves, écrites ou orales), des obligations juridiques s'imposent, en particulier à l'enquêteur : celui-ci est notamment tenu de protéger l'enquêté, son anonymat et la confidentialité de ses propos, et de respecter les conditions d'utilisation que l'enquêté a acceptées initialement. La portée de ce contrat va néanmoins au-delà : il symbolise les impératifs éthiques et déontologiques qui s'imposent à l'enquêteur. C'est un engagement moral, à la fois vis-à-vis de la personne interrogée, mais aussi au regard de la communauté des chercheurs, puisque le respect d'un tel principe conditionne la possibilité des enquêtes futures. Plus généralement, elle a des incidences sur le crédit scientifique apporté aux SHS dans la société, dans sa capacité à faire valoir et à respecter une déontologie de la donnée.

Comme nous le verrons par la suite, il peut y avoir une tension entre le respect de ce contrat de confiance et les possibilités de traitement (analyse, diffusion) des données. Comment alors respecter les impératifs éthiques et juridiques qui s'imposent au chercheur-producteur et au chercheur-réutilisateur des données ?

### **1. Le chercheur doit se positionner entre normes éthiques...**

En matière de mutualisation des données de la recherche, existe-t-il une déontologie commune à l'ensemble des chercheurs ? Assurément, non. Pour autant, un certain nombre de principes permettent de réfléchir aux enjeux déontologiques du partage des données, autour de la mise en tension et de la conciliation entre deux principes : l'intérêt scientifique du partage *vs* la protection des enquêtés et des données d'enquête.

En sciences dures, les questionnements posés sur les méthodes et objectifs de la recherche scientifique ont mené à l'élaboration d'une "éthique de la recherche", définie comme "*un ensemble de règles ayant pour objet de soumettre l'activité scientifique au respect de valeurs jugées plus hautes que la liberté du chercheur*" (Conseil National du Développement des Sciences Humaines et Sociales, 2001).

En SHS, on peut se référer aux publications du COMETS, comité d'éthique créé en 1994 par le CNRS ; le guide "Pratiquer une recherche intègre et responsable" publié en 2017, est censé "*fournir une aide à tous ceux [qui] sont confrontés à des dilemmes éthiques*". Concernant le traitement de données personnelles, il rappelle quelques obligations, notamment : "s'assurer de la confidentialité des données", "permettre l'information des personnes concernées par l'étude" ou encore "soumettre à l'autorisation de la CNIL les traitements informatiques de données personnelles qui présentent des risques particuliers d'atteinte aux droits et aux libertés". En somme, il rappelle simplement les principes juridiques qui s'imposent lors du traitement de données de la recherche (respect de la confidentialité et de l'anonymat, consentement des personnes). Concernant les principes éthiques, le guide rappelle une nouvelle fois les grands principes généraux sur lesquels "la société fonde sa confiance en la recherche : honnêteté, intégrité, responsabilité (ce qui signifie qu'en cas de "faute", le chercheur sera jugé à l'aune de sa profession). Le CNDSHS ajoute les principes de liberté (en particulier liberté d'expression), d'indépendance, de tolérance, d'objectivité et de rigueur... et précise que la déontologie du chercheur (sa liberté de chercheur) ne concerne que ce qui n'est pas explicitement interdit par la loi.

Ces principes fournissent des balises pour réguler le travail de chercheur, en particulier les principes de responsabilité et d'honnêteté vis-à-vis des enquêtés. Ces principes restent néanmoins très généraux ; il manque un code de déontologie qui décline ces grands principes en règles de conduite communément admises pour encadrer le partage des données.

Certaines pistes sont lancées à l'image du projet de charte déontologique initié il y a quelques années par l'association française de sociologie, dont une partie portait sur la question de la réutilisation des données. Celle-ci n'a pas été adoptée, ce qui montre notamment la difficulté à élaborer un consensus écrit autour de règles déontologiques à respecter. Elle stipulait que la réutilisation des données était un objectif souhaitable, mais pas automatique, mais ne donnait guère de précisions sur les procédures à mettre en œuvre pour arriver à cette fin : "*Les sociologues prévoient le partage de données comme partie intégrale d'un plan de recherches dès lors que le partage de données est possible. Les sociologues rendent leurs données disponibles à la fin d'un projet de recherche ou après une publication, sauf si des accords de propriété avec des employeurs, ou des financeurs, excluent une telle accessibilité, ou s'il est impossible de protéger la confidentialité des données et l'anonymat des participants à la recherche lors d'un partage. Les sociologues qui ne placent pas leurs données dans des archives publiques conservent celles-ci, avec la documentation concernant la recherche, par leurs propres moyens pendant une période suffisante après publication ou diffusion des résultats. Les sociologues doivent être en mesure de s'expliquer sur l'origine de leurs données et sur les méthodes utilisées pour les obtenir. L'utilisation des données recueillies (questionnaires, retranscription d'entretiens, notes d'entretien, documents divers etc.) est sous le contrôle exclusif des sociologues qui doivent veiller à ce qu'il n'y ait pas de manipulation de résultat et qu'elles soient archivées pour être accessibles à d'autres collègues. Ceux-ci peuvent demander à en évaluer la pertinence, ou à les réutiliser pour une autre recherche*".

Dans le cadre de notre réflexion, ce projet de charte retient notre intérêt car il considère les objectifs d'archivage et de mutualisation comme constitutifs d'une éthique de la recherche ainsi sous-tendue par des valeurs de conservation (patrimoniale) des données, de transparence de la démarche scientifique ou encore de la capitalisation de données possiblement réutilisables. Simultanément, il réaffirme l'impératif éthique de protection des données personnelles, confidentialité et anonymat. Implicitement, ce projet de charte nous conforte dans l'idée qu'il faut mettre en regard ces deux plans éthique et juridique et étudier en détail les conditions et les modalités d'une conciliation.

## 2. ... et impératifs juridiques

Selon la législation actuelle sur la protection des données (issu de la loi Informatique et Libertés du 6 janvier 1978), la CNIL impose des règles au moment de la collecte, du traitement et, *a fortiori*, du partage des données collectées au cours des entretiens, en raison du caractère “personnel” d’une partie de ces données (qui permettent une identification directe ou indirecte des enquêtés). Par ailleurs, selon leur degré “d’originalité”, les entretiens peuvent également être protégeables au nom de la propriété intellectuelle, auquel cas enquêteur et enquêté pourraient jouir d’un droit d’auteur ; pour arguer du caractère original d’un entretien, il faudrait expliciter les choix libres et créatifs de l’auteur.

L’application à partir du 25 mai 2018 du règlement européen d’avril 2016 sur la protection des données personnelles modifie en profondeur la législation sur la protection des données de la recherche (Soubiran, 2017). Ce règlement poursuit trois objectifs : renforcer la protection des personnes, notamment à l’ère du tout-numérique ; responsabiliser les acteurs traitant des données (chercheurs notamment) ; crédibiliser la régulation, ce qui passe notamment par un renforcement des sanctions.

### *a. Question du consentement et de l’information*

Jusqu’à l’entrée en vigueur du règlement européen, la législation impose de solliciter le consentement des enquêtés, consentement qui doit être “éclairé” : l’enquêteur doit notamment préciser auprès de qui, pendant combien de temps et *via* quel média les données seront accessibles, quels sont les intérêts et les risques de la diffusion. Pour être juridiquement valable, il doit rester une trace (écrite ou orale) du consentement donné par l’enquêté.

Dans notre situation, étant donné que les objectifs de mutualisation et de diffusion n’étaient pas envisagés au moment de la passation des entretiens, nous ne disposons pas de ce consentement.

Plusieurs voies peuvent être suivies suivant les situations : dans le cas de la diffusion (au sens de la CNIL, i.e. en *open data*), la CNIL impose de recontacter les enquêtés pour solliciter leur consentement. Ce recontact pose des questions de méthode : comment aborder des individus que nous avons pour certains interviewés il y a de nombreuses années ? Juridiquement, a-t-on le droit de conserver les coordonnées de ces enquêtés pendant aussi longtemps ? Quid du droit à l’oubli (selon la CNIL, les coordonnées des enquêtés ne peuvent pas être transmises ni gardées au-delà du temps de la recherche, sauf accord express de l’enquêté) ? En outre, certains chercheurs soulignent le risque que la sollicitation du consentement à la diffusion nuise à la relation avec l’enquêté (Bishop, 2013) ; ce faisant, on court notamment le risque que les enquêtés refusent la diffusion.

Dans le cas d’une demande de réutilisation à des fins scientifiques, les règles fixées par la CNIL sont *a priori* moins contraignantes. A condition que les matériaux restent dans la sphère académique, le contrat de confiance passé initialement avec les enquêtés constituerait une condition suffisante pour diffuser les matériaux qualitatifs.

À partir de mai 2018, la situation change : en particulier, les données de recherche, collectées et traitées à des fins scientifiques, jouissent d’un statut particulier qui limite grandement les impératifs juridiques.

Concernant les données à caractère personnel (qui révèle l’origine raciale ou ethnique, les opinions politiques, les convictions religieuses ou philosophiques ou l’appartenance syndicale de l’enquêté ; qui concerne la vie sexuelle ou l’orientation sexuelle d’une personne physique ; les données génétiques ou biométriques permettant d’identifier une personne physique de manière unique) : leur traitement est

interdit, exception faite des traitements à des fins archivistiques dans l'intérêt public, à des fins de recherche scientifique ou historique ou à des fins statistiques (article 9).

L'article 6 précise en outre qu'il est en principe impératif de recueillir le consentement des personnes pour collecter ou traiter des données à caractère personnel... exception faite des *“traitements nécessaires à l'exécution d'une mission d'intérêt public”*. En tout état de cause, même dans les cas où le règlement n'impose pas, de manière exceptionnelle, de recueillir le consentement préalable des personnes, il est impératif que ces dernières soient informées de manière « loyale et transparente au regard de la personne concernée » (article 3), l'information communiquée devant être *“concise, compréhensible et aisément accessible”*. Contrairement à la situation actuelle, le chercheur n'est donc plus juridiquement obligé d'obtenir le consentement des enquêtés pour traiter les données si *« le traitement est nécessaire à l'exécution d'une mission d'intérêt public ou relevant de l'exercice de l'autorité publique dont est investi le responsable du traitement »* ; une simple information éclairée suffit.

En outre, le règlement précise que *“la charge de la preuve du consentement incombe au responsable de traitement”*. Le responsable de traitement est défini comme *« la personne physique ou morale, l'autorité publique, le service ou un autre organisme qui, seul ou conjointement avec d'autres, détermine les finalités et les moyens du traitement »* (article 4) ; en pratique, dans le cas d'une recherche scientifique le responsable de traitement peut être le représentant légal de l'entité dont dépend le chercheur. Ceci signifie que le responsable du traitement doit s'assurer d'être en mesure de démontrer que le traitement et l'information sont effectués conformément au règlement. À chaque nouveau traitement (lors des ré-analyses), le responsable du traitement devra respecter en tout état de cause les dispositions du règlement. Ceci pose une nouvelle fois des questions de méthode : ceci signifie notamment que, pour que les chercheurs secondaires puissent informer les enquêtés du traitement qui est fait de leur données, les données soient accompagnées des coordonnées des enquêtés... ce qui semble totalement incompatible avec l'impératif d'anonymat.

Concernant le droit à l'oubli, l'article 17 impose au responsable du traitement des données d'effacer dans les meilleurs délais les données à caractère personnel qui *“ne sont plus nécessaires au regard des finalités pour lesquelles elles ont été collectées”* ; une nouvelle fois, la règle ne s'applique pas aux données utilisées à des fins scientifiques : cette exception rend légale le partage et la réutilisation à des fins scientifiques de données de la recherche (et leur archivage dans la durée).

Le nouveau règlement européen modifie également les jeux d'acteurs impliqués dans la protection des données. Ainsi, dans l'objectif de responsabiliser les acteurs traitant des données (notamment les chercheurs), ceux-ci ne seront plus nécessairement tenus de faire de déclarations à la CNIL. En revanche, une analyse d'impact devra être effectuée si le traitement est *« susceptible d'engendrer un risque élevé pour les droits et libertés des personnes physiques »*, ce qui pourra conduire à une consultation préalable de la CNIL dans certains cas (cf. articles 35 et 36 du règlement).

Le règlement institue par ailleurs un nouvel acteur, le Délégué à la Protection des Données, qui sera *“chargé de mettre en œuvre la conformité au règlement [...] au sein de l'organisme qui l'a désigné s'agissant de l'ensemble des traitements mis en œuvre par cet organisme”* ; le DPO sera notamment chargé de tenir à jour un registre des traitements (article 30). Cela signifie que la gestion des données de la recherche sera mise en œuvre dorénavant à l'échelon local, avec des effets pour le moins incertains : assouplissement des contraintes et facilitation de la diffusion dans certains cas ; application plus ferme des règlements dans d'autres.

## *b. Question de l'anonymat et de la confidentialité*

### *i. La confidentialité des données d'enquête : une règle d'or des méthodes qualitatives*

Les enquêtes qualitatives menées en SHS collectent des données ayant trait à la vie privée. Elles enregistrent et analysent des informations personnelles, des paroles énoncées, des actes avoués qui relèvent, pour une large part, de l'intime, et sont relatés sous couvert d'anonymat. Dès lors, de la collecte à la diffusion des résultats, elles sont soumises à une exigence de confidentialité : elles doivent garantir aux personnes rencontrées que leurs propos – et les informations qui y sont contenues – ne seront pas répétés dans leur intégralité, doivent empêcher qu'une personne extérieure à l'enquête puisse reconnaître les personnes y ayant participé, et doivent aussi éviter que ces dernières puissent se reconnaître les unes les autres. Confidentialité et anonymat sont ainsi les deux faces de la même médaille et doivent assurer une dissociation entre les paroles (ou les actes), d'une part, et l'identité de la personne enquêtée, de l'autre.

## ii. Les spécificités des corpus et de la démarche

Au-delà des principes généraux, les conditions de mise en œuvre du principe de confidentialité et d'anonymat dépendent beaucoup des situations d'enquêtes, de la méthode, du thème et des terrains choisis. Aude Béliard et Jean-Sébastien Eidelimann (2008) insistent sur la nécessité d'appréhender la mise en œuvre d'une politique de confidentialité de manière contextuelle, en fonction des enquêtes et des traitements envisagés. Ils nous invitent, par là-même, à réfléchir aux spécificités des matériaux que nous collectons, et dans notre cas, que nous souhaitons archiver et diffuser.

Plusieurs spécificités de nos corpus renforcent les enjeux de confidentialité.

En premier lieu, d'un point de vue thématique, les matériaux qualitatifs portant sur la mobilité quotidienne abordent de manière transversale de nombreuses dimensions de la vie ordinaire. Souvent précédées d'un récit retraçant le parcours biographique, ces enquêtes amènent les personnes interrogées à relater l'ensemble de leurs déplacements et, ce faisant, à exposer divers « plis » de leur existence. De ce fait, elles peuvent avoir le sentiment de se livrer « en totalité », de « se mettre à nu ». Sous couvert d'investiguer les pratiques de déplacement, l'entretien livre ainsi un accès direct aux différentes faces de l'intime et se révèle, dans les faits, assez intrusif. L'approche biographique, généralement associée, peut renforcer ce sentiment. En comparaison d'une enquête qui, par exemple, ne porterait que sur le rapport au travail, des liens peuvent être tissés entre les différentes sphères de la quotidienneté. Ce décloisonnement entre les différentes facettes de l'identité individuelle explique que les enjeux de confidentialité y sont particulièrement importants.

En deuxième lieu, les enquêtes de mobilité portent sur des terrains de « proximité » où la distance entre les enquêtés et les enquêteurs est faible. Dans la plupart des corpus, le recrutement des enquêtés a été réalisé de proche en proche, ce qui facilite la tâche mais augmente considérablement le risque d'interconnaissance entre chercheurs et enquêtés, entre enquêtés eux-mêmes, entre enquêtés et lecteurs de l'enquête. Un autre aspect renforce considérablement les risques d'identification liés à l'interconnaissance : il s'agit de la territorialisation des enquêtes, i.e. le regroupement des enquêtes dans un même périmètre. Ce risque est poussé à l'extrême dans une partie des corpus (programme UPHA) où l'investigation a été menée à l'échelle du quartier ; dans ce cas, non seulement les personnes sont très facilement identifiables, mais les propos relatifs aux pratiques de proximité, aux rapports de voisinages ou aux politiques menées par la commune, exposent beaucoup les enquêtés.

Une troisième spécificité de nos matériaux est le caractère particulièrement identifiant des itinéraires et des lieux pratiqués. L'approche individuelle et qualitative des systèmes de mobilité amène, pour tirer parti d'un cas, à conserver pour l'analyse des localisations précises, des temporalités ; elle invite également à reprendre sans les déformer les « énoncés » en matière de qualification des lieux et des itinéraires fréquentés. Associées la plupart du temps à la description d'activités personnelles (« *Je prends tous les matins le train à la gare de Montlouis à 7h58 pour me rendre au travail dans le quartier des Deux-Lions* »), les pratiques de mobilité ainsi relatées singularisent fortement les



personnes et, en l'absence de précautions, les rendent potentiellement repérables. Il en va ainsi, notamment, des personnes dont la mobilité est relativement atypique et « visible », comme les personnes qui, en périurbain, circulent à vélo. Cette possibilité de reconnaissance est bien évidemment beaucoup plus forte dans les territoires peu denses, périurbains et ruraux, que nous investiguons principalement, que dans le cœur des grandes métropoles, où la « masse » préserve mieux l'anonymat.

iii. L'anonymisation : une garantie « nécessaire » mais très relative

L'objectif de préservation de la confidentialité des matériaux d'enquête achoppe sur les difficultés à anonymiser les matériaux dans leur globalité. En effet, si l'on s'en tient à une définition *stricte* de l'anonymat, l'anonymisation est non seulement impossible à mettre en œuvre – puisqu'il faudrait « brouiller » la totalité des données personnelles (qui constituent par exemple le corps même de l'entretien) – et stérile, puisqu'elle contrevient à l'objectif d'une méthodologie qualitative et compréhensive qui vise à tisser des réseaux de signification entre les informations et/ou les éléments de discours recueillis. L'anonymisation dans la définition stricte que retient la CNIL n'est donc ni envisageable ni souhaitable.

Si toutefois l'on considère, dans une conception plus ouverte de l'anonymisation, qu'elle obéit à un gradient, plusieurs techniques pourraient être mises en œuvre afin de préserver – de manière relative – l'anonymat.

On ne discutera ici que de l'anonymisation des entretiens et traces-GPS, dont nous disposons dans le méta-projet.

Dans le cas des entretiens semi-directifs, la première technique est une technique assez simple et banale de pseudonymisation que l'on peut réaliser en modifiant les prénoms, ne gardant que les initiales ou les remplaçant par un numéro. L'usage d'un pseudonyme est désormais la méthode la plus utilisée en SHS depuis le milieu des années 1980 puisqu'elle préserve l'idée que l'enquêté est une personne « entière », incarnée, actrice de ses actes et de son récit. Remarquons que le choix d'un pseudonyme est déjà une altération de la qualité d'un entretien car le prénom (choisi par les parents) est emprunt d'une symbolique qui impose – si on veut en garder toute la signification – de travailler sur le choix d'équivalents symboliques (Coulmont, 2017).

Néanmoins, l'introduction d'un pseudonyme n'abaisse que faiblement le risque d'identification d'un enquêté pour le lecteur qui disposerait d'un entretien, puisqu'il demeure très facile de reconnaître un individu par recoupement des autres informations personnelles (domicile, emploi, âge, nombre d'enfants).

Cela nous amène à envisager une deuxième méthode susceptible d'abaisser plus nettement le risque (au sens de la probabilité) d'identification. Issue des techniques de randomisation (altération des données individuelles), elle consiste à introduire sur certaines données clés « un brouillage » : taire le nom de l'employeur et le métier en ne retenant que la PCS détaillée, donner une fourchette d'âge plutôt que l'âge exact, dissimuler l'adresse, changer le nom de la rue, etc.

Ce brouillage des données pose néanmoins un problème considérable d'analyse pour les chercheurs travaillant sur la dimension spatiale des pratiques : en géographie ou en sociologie urbaine, il est extrêmement difficile d'envisager de brouiller les lieux. Ainsi, si l'on prend l'exemple de Tours, habiter au « Sanitas » ou aux « Prébendes » (pourtant deux quartiers contigus) est particulièrement signifiant puisque ces deux quartiers balisent dans l'ordre des représentations collectives les deux extrémités de l'espace social tourangeau. Changer le « Sanitas » par un autre quartier populaire n'est pas envisageable car c'est le seul quartier de grands ensembles situé dans l'hyper-centre, et gommer cela reviendrait à ne plus rien comprendre des pratiques. Ne pas évoquer le « nom » en retenant «

quartier d'habitat social situé en centre-ville » reviendrait à cultiver inutilement l'art du secret de polichinelle.

Concernant les traces-GPS, il est autant, voire plus primordial que pour les données textuelles, de les anonymiser de façon optimale. Une trace GPS est constituée d'un ensemble de points géo-localisés (définis par leurs latitude, longitude et altitude), enregistrés à une fréquence suffisamment faible (quelques secondes) pour renseigner sur les lieux traversés et arrêts effectués par la personne suivie, à l'échelle d'une ou plusieurs journées généralement. Prise dans sa totalité, une trace constitue donc une mine d'informations extrêmement précises qui permettent, en les recoupant, d'identifier facilement l'enquêté. Il suffit notamment de repérer le lieu de résidence (où l'enquêté passe la nuit) ou le lieu d'emploi (où il passe la plupart de ses journées) pour obtenir l'identité de l'enquêté. Il est ensuite possible d'obtenir des informations relativement sensibles, liées par exemple aux pratiques religieuses ou politiques de l'individu (Primault *et al*, 2015).

Diverses méthodes d'anonymisation ont été développées. Certaines jouent sur l'altération des lieux pratiqués (*spatial distortion*). Andrés *et al* (2013) ont ainsi développé le concept de « *geo-indistinguishability* » qui consiste à protéger la localisation des points d'arrêt des individus en les déplaçant de manière aléatoire au sein d'un cercle de rayon  $r$  : plus le rayon  $r$  est élevé, plus le niveau de confidentialité sera élevé. Abul *et al* (2010) proposent une autre technique consistant à remplacer la trace GPS (polyligne dans un espace tridimensionnel) par un volume cylindrique : ce ne sont plus uniquement les points d'arrêt mais la totalité des points enregistrés qui sont alors déplacés de manière aléatoire. On peut également aller plus loin avec cette méthode de distorsion spatiale, en « dé-territorialisant » totalement la trace, i.e. en séparant les points enregistrés de leur environnement ; une telle méthode conviendrait par exemple aux analyses statistiques ou de modélisation, qui chercheraient simplement à opérer des calculs (de distances parcourues, de nombre de points d'arrêts, etc.) sans analyser les résultats au regard du contexte des déplacements. Enfin, d'autres méthodes reposent sur une altération du temps (*time distortion*) plutôt que l'espace : Primault *et al* (2015) proposent de gommer les changements de vitesse, de manière à invisibiliser les points d'arrêts tout en gardant en revanche une précision d'informations sur les lieux traversés/pratiqués (à l'exception des points de départ et d'arrivée, qui sont également brouillés).

Ces différentes méthodes d'anonymisation reposent sur du brouillage, spatial ou temporel : l'ajout de « bruit » va inévitablement altérer l'utilité et les possibilités d'analyse des données. Certaines méthodes d'anonymisation sont donc inenvisageables pour des chercheurs qualitatifs travaillant sur les mobilités spatiales quotidiennes ; ainsi, il est par exemple impensable de « dé-territorialiser » la trace puisqu'il est absolument indispensable de comprendre les traces (et les pratiques qu'elles donnent à voir) par rapport à leur environnement et au territoire dans lequel elles se déroulent. De même, il nous est nécessaire de connaître avec précision les points d'arrêt parcourus au cours des différents trajets quotidiens, à partir desquels on peut comprendre les logiques qui président aux déplacements : la méthode de *time distortion* n'est donc pas du tout adaptée à nos méthodes d'analyse. Comme pour les entretiens semi-directifs, un brouillage relatif des lieux de résidence et des lieux de travail (dans un périmètre assez restreint qui serait à définir), mais aussi des dates de collecte, permettrait d'abaisser le risque de reconnaissance, tout en conservant la valeur analytique de la trace. Il n'en demeure pas moins que toute altération des données a des conséquences sur les possibilités de réutilisation.

#### iv. Une anonymisation qui contraint les possibilités de réutilisations

Avant de diffuser ses données, un chercheur doit donc les anonymiser, selon une technique ou une autre permettant de garantir la confidentialité des enquêtés et de leurs propos tout en garantissant leur intérêt scientifique.

Mais l'intérêt... pour qui ? On l'a dit, il est nécessaire de ne pas brouiller trop de données pour conserver une possibilité d'analyse, pour pouvoir continuer à « tisser des réseaux de signification entre les informations et /ou les éléments de discours recueillis ».

On l'a dit aussi, certaines données auront plus d'importance dans certaines disciplines que dans d'autres : un géographe aura besoin de connaître les lieux pratiqués avec précision et ne se contentera pas d'un nom de commune, mais pourra se satisfaire d'une catégorie de PCS à propos de la profession de l'enquêté ; au contraire, il sera nécessaire pour un sociologue de conserver des informations précises sur les emplois pratiqués, alors qu'une information plus vague sur les lieux fréquentés restera satisfaisante.

Toute anonymisation, tout brouillage ou suppression d'informations, limitera nécessairement les analyses qui pourront être faites par d'autres chercheurs : a fortiori, en fonction des procédures selon lesquelles elles auront été anonymisées, les données mises en partage ne pourront pas être utiles pour des chercheurs de toutes disciplines.

Ainsi, des entretiens anonymisés par un géographe pourraient se révéler stériles pour un sociologue. Ceci pourrait constituer une réelle difficulté dans le cas de la création d'une base de données sur les mobilités spatiales : ce champ d'étude regroupe en effet des chercheurs qui viennent de diverses disciplines, abordent la question sous différents angles et ne cherchent *a priori* pas les mêmes informations dans les données de la recherche.

### **Conclusion**

Quelle que soit la technique d'anonymisation retenue, il est nécessaire de rappeler qu'elles permettent d'abaisser sensiblement le risque d'identification mais n'excluent en aucun cas une possible identification par un tiers qui pourrait reconnaître un individu par recoupement d'autres informations que l'on ne peut pas randomiser sans stériliser le matériau d'enquête. Cette précision prend tout son sens dans cas du méta-projet. D'une part parce que lorsqu'on travaille en partie sur des territoires de faible densité, où certains hameaux ne sont peuplés que de quelques habitants qui sont donc facilement identifiables en croisant peu d'informations. D'autre part parce que l'approche qualitative mise en œuvre implique de travailler à l'échelle individuelle sur des « cas », en croisant des informations issues des divers matériaux à notre disposition (entretiens semi-directifs, traces-GPS, voire photo, carnets de pratiques, etc.). Ce faisant, on multiplie les informations à caractère personnel et on élève d'autant le risque d'identification des individus.

Le chercheur qui souhaite traiter (et en particulier partager) des données à caractère personnel est soumis à des impératifs juridiques, qui lui imposent d'avoir le consentement de l'enquêté – et en tout état de cause de l'informer de manière transparente et loyale conformément aux dispositions du règlement européen – et d'anonymiser ses propos. Dans le cas où le consentement de l'enquêté n'a pas pu être obtenu, mais surtout en raison de l'anonymisation nécessairement partielle des matériaux, le chercheur doit se référer à certaines valeurs éthiques partagées par la profession.

Plus généralement, il nous semble qu'aucun règlement ne doit se substituer à l'exercice du jugement moral du chercheur : même lorsqu'il est « couvert » juridiquement, le chercheur doit toujours interroger ses actes au regard de l'éthique de sa profession et de sa propre déontologie.

### III. Modalités de partage et de diffusion – prises de positions

Nous proposons ci-dessous de définir des modalités concrètes pour le partage des données de la recherche, en termes de périmètre de diffusion (qui peut avoir accès aux données ?) et de conditions de mise à disposition. Ces modalités et le protocole de partage diffèrent selon qu'il s'agit de données déjà existantes, ce qui suppose une gestion *a posteriori* (gestion que nous appliquons pour les données analysées dans le méta-projet) ; ou de données qui seront produites dans le cadre de futurs projets.

Ces modalités que nous présentons sont guidées par un objectif : arbitrer au mieux, dans un système en tension, entre le respect du contrat de confiance passé initialement avec l'enquêté au moment de la collecte des données et l'intérêt scientifique du partage des données. D'un côté, en raison des enjeux juridiques (qui lient le chercheur à l'enquêté) et déontologiques (qui lient le chercheur à l'enquêté, mais également au reste du monde académique) explicités *supra*, il nous semble essentiel de ne pas trahir et de préserver au maximum ce contrat de confiance ; de l'autre, parce que nous soutenons l'idée qu'il est bénéfique pour la recherche de partager les données de la recherche, il nous semble essentiel de favoriser et de faciliter ce partage.

#### 1. Périmètre de partage des données

Cette question du périmètre de partage recouvre trois questions : quelles données de la recherche partager ? Avec qui les partager ? À quelle fin, dans quel objectif de réutilisation ?

Nous faisons et assumons le choix de restreindre ce périmètre de partage : en effet, il nous semble que c'est un élément de garantie du contrat de confiance passé initialement entre l'enquêteur et l'enquêté. Ce faisant nous rejoignons les positions d'autres structures de partage de la donnée (Quetelet, Archethno, beQuali, Qualidata). Nos choix sont guidés d'une part par les valeurs éthiques indissociables de la recherche, d'autre part par les valeurs de compétences (méthodologiques, des réutilisateurs) et d'échange (entre les réutilisateurs et les chercheurs-primaires).

##### a. Quelles données ?

Afin de préserver l'anonymat des personnes enquêtées et la confidentialité de leurs propos, nous proposons de définir des niveaux de confidentialité et des restrictions d'accès en fonction du type de données. Ce procédé est également appliqué par d'autres structures de partage : Qualidata par exemple définit trois niveaux d'accès, en distinguant "*open data*", "*safeguarded data*" et "*controlled data*". Ainsi, les données que nous qualifions de « brutes » et « travaillées » ne sont pas accessibles.

Type de données	Description	Niveau d'accès
Présentation des recherches	Éléments de présentation des programmes de recherche (date, financeur(s), problématiques, méthodologie employée, principaux résultats)	Accès libre, sans contrôle d'accès
Métadonnées individuelles	Ensemble des caractéristiques de l'individu (ex.: tranche d'âge, situation familiale, statut)	Accès libre, sans contrôle d'accès

	d'occupation, ancienneté dans le logement)	
Données brutes (enregistrements, traces GPS, photos, carnets de terrain)	Données collectées directement sur le terrain par le chercheur primaire	Non accessibles, hormis par le chercheur primaire ; archivées dans une plateforme d'archivage
Données travaillées (retranscriptions)	Données brutes retravaillées pour permettre leur analyse	Non accessibles par quiconque ; uniquement accessibles par le chercheur primaire ; archivées dans une plateforme d'archivage
Données anonymisées (retranscription anonymisées, traces GPS anonymisées)	Données anonymisées	Soumises à contrôle d'accès ; accessibles sous conditions : faire une demande d'accès, signer une charte de réutilisation

### b. Quel public ?

Qui peut avoir accès aux données mises en partage ?

Le principal prérequis relève des compétences méthodologiques permettant d'analyser les données de recherche. D'un point de vue général, le chercheur secondaire doit être en mesure de comprendre les procédures et les outils mis en œuvre par le chercheur primaire et d'en cerner les implications. Il doit également, à partir de ce matériau résultant de la mise en œuvre d'une première stratégie de recherche, redéfinir sa propre méthode de traitement des données. Pour parvenir à ses fins, dans un cadre scientifique rigoureux, il doit donc faire montre d'un haut niveau de maîtrise des procédures et des outils mis en œuvre dans les recherches qualitatives. En effet, analyser des données issues d'approches qualitatives (entretiens semi-directifs, carnets de recherche ethnographiques) ou des traces-GPS nécessite des connaissances méthodologiques fines, et *a fortiori* une formation spécifique. Ainsi, pour exploiter des entretiens semi-directifs par exemple, il ne suffit pas de piocher quelques *verbatim* parlants permettant d'asseoir une analyse, mais il est nécessaire de comprendre les entretiens dans leur globalité tout en tenant compte de leur contexte de production, de la position et des caractéristiques du/des locuteur(s). De même, analyser un carnet de terrain produit par un anthropologue nécessite d'avoir suivi des enseignements méthodologiques.

En l'absence de ces connaissances, il nous semble que les données ne pourront pas être analysées avec rigueur.

Des limites doivent donc être posées, au-delà desquelles on considèrera que les individus n'ont pas les compétences requises. Néanmoins, juger des compétences méthodologiques de chacun est compliqué : si elles s'acquièrent traditionnellement par le biais de formations spécifiques (dans l'enseignement supérieur notamment), elles peuvent également se développer au cours de diverses expériences de vie, ou au contact d'individus disposant de ces compétences. Alors, où et comment poser les limites ? À défaut de mieux, il nous semble que le statut institutionnel constitue l'indicateur le plus satisfaisant même s'il peut paraître limitatif : pour ces raisons, nous proposons de limiter le partage des données aux docteurs (qu'ils soient ou non rattachés à une institution de recherche - cf. ci-dessous), post-

doctorants, doctorants et étudiants en master (sous la direction d'un enseignant-chercheur), qui ont bénéficié et validé une formation méthodologique qui nous semble nécessaire à l'analyse secondaire. Notre position nous rapproche de celle adoptée par les autres structures de partage de données ; Qualidata va même plus loin, en limitant l'accès aux *controlled data* (données personnelles) aux seuls chercheurs accrédités par l'ONS (*Office for National Statistics*), ayant donc suivi une formation spécifique.

Le contexte actuel de la recherche française amène de plus en plus de jeunes docteurs à quitter le monde universitaire : on assiste ainsi à une diversification des acteurs de la recherche et au développement, en SHS, d'une recherche "indépendante", opérée par des structures privées sans nécessaire connexion avec la recherche académique, qui peut se prévaloir d'apporter des "connaissances à portée générale". Dans ce cas, quelles pourraient être les conditions d'ouverture des données ?

À l'impératif méthodologique évoqué supra, un deuxième impératif déontologique est lié au "désintérêt" du chercheur, au sens de désintéressement économique. L'accès aux données sera donc refusé à tout demandeur qui tirerait un quelconque bénéfice financier de la réutilisation des données : nous rejoignons ici la position du réseau Quetelet qui exclut toute utilisation commerciale des données ou servant "*les intérêts particuliers d'institutions publiques ou privées*". De fait, les données ayant été collectées ou produites dans un objectif scientifique, il nous semble qu'elles ne doivent pas servir des intérêts particuliers, notamment commerciaux. D'une part, car l'utilisation à des fins commerciales de matériaux produits sur fonds publics peut être entrevue comme une opération d'externalisation des coûts et, réciproquement, de privatisation des bénéfices, sauf à imaginer une rétribution en amont ou une contribution en aval à la recherche publique. D'autre part, car l'utilisation commerciale abolit les principes d'indépendance, de désintéressement et de neutralité axiologique qui sont collectivement pensées comme des valeurs fondamentales de la recherche.

Il nous semble donc impossible de partager des données de la recherche avec des chercheurs travaillant pour des entreprises privées ou à but commercial, pour qui ce critère de désintérêt serait difficilement respecté.

### c. *A quelle fin, dans quel objectif ?*

Lors de la collecte des données, les chercheurs sollicitent le consentement des enquêteurs pour utiliser les données à *des fins scientifiques*, en précisant qu'ils agiront dans le cadre d'un programme de recherche particulier. Cette finalité scientifique, sous-tendue par l'idée de participation au "bien commun" que constitue la production scientifique, constitue un ressort de l'engagement de l'interviewé dans la situation d'enquête et lui apporte une rétribution symbolique en échange du temps investi. Dès lors, dans une perspective déontologique, cette finalité scientifique doit impérativement être préservée puisqu'elle est fondatrice de la situation d'enquête et du contrat de confiance passé avec l'enquêté. En outre, cette finalité scientifique n'est pas neutre juridiquement, puisqu'elle nous permet notamment de conserver ces données (souvent à caractère personnel), et de ne pas s'astreindre au principe du droit à l'oubli.

Pour respecter ce contrat de confiance, et en particulier ne pas trahir les modalités d'utilisation présentées initialement par le chercheur, il nous semble donc primordial de contrôler les objectifs de réutilisation des données que poursuivent les demandeurs.

Ainsi, l'accès aux données sera donné uniquement aux demandeurs justifiant d'un véritable projet scientifique. Nous rejoignons ici la position du réseau *Quetelet*, qui limite l'accès aux données aux projets visant la "*production ou reproduction, dans un but de validation, de connaissances nouvelles de portée générale*". Ils précisent également que les "*résultats [doivent être] publics et libres de*

*diffusion [...] et être destinés à des publications scientifiques.*” Qualidata adopte la même posture, en exigeant des demandeurs (uniquement des chercheurs dans leur cas) un projet détaillé, co-signé par le responsable de leur laboratoire.

Qualidata exige des chercheurs demandant un accès aux données de « suivre une formation spécifique ». Pour notre part, nous préconiserons au demandeur (en plus de signer la charte de réutilisation détaillée ci-dessous) de s’entretenir avec le chercheur-producteur des données ou un responsable du programme de recherche.

Un autre critère essentiel pour donner accès aux données concerne le contrôle par les pairs, principe déontologique fondateur et fédérateur du monde de la recherche. En effet, pour faire œuvre scientifique, une recherche doit assurer la transparence de sa démarche, des méthodes et des matériaux mobilisés, et recourir aux instruments de diffusion scientifique qui permettront à cette recherche d’être évaluée (*reviewed*) et discutée. Cette exigence doit garantir le fait que le demandeur s’inscrit bel et bien dans une démarche de recherche et participe à la construction d’un savoir commun.

Enfin, l’accès aux données sera donné aux demandeurs justifiant d’un projet de réutilisation des données à des fins pédagogiques, dans le cadre d’enseignements en licence, master ou doctorat. Plusieurs travaux soulignent l’intérêt de travailler sur des enquêtes déjà réalisées pour assurer une formation critique aux méthodes de collecte. Cet objectif nous semble particulièrement louable concernant les méthodologies qualitatives qui – de notre avis – ne sont pas toujours enseignées avec autant de systématisme que les méthodes quantitatives. Par ailleurs, rendre possible par la diffusion une « critique méthodologique » des méthodes de collecte nous semble particulièrement fécond et à même de porter dans la formation à la recherche une montée en puissance d’un niveau de réflexivité global sur les méthodes.

#### Deux cas-limites :

##### Cas des structures assurant des missions de service public

Un point de débat porte sur le partage des données avec ou en vue d’une utilisation par des institutions publiques ou parapubliques à des fins d’intérêt général : association, agence d’urbanisme, institut d’aménagement, collectivité et, cas limite, opérateur privé assurant des missions de service public. Dans ce cas, si les utilisateurs sont compétents et inscrivent leur réflexion dans un objectif visant l’intérêt commun, on peut légitimement s’interroger sur l’ouverture des données. Cela pourrait d’ailleurs rejoindre le souhait d’un certain nombre de chercheurs que leurs travaux aient une plus grande utilité sociale et “servent” la société civile.

Ce sujet a néanmoins suscité de longues discussions au sein de l’équipe du méta-projet, qui nous semblent symptomatiques de réticences ou de points de blocage.

En premier lieu, des interrogations ont porté sur la capacité de ces institutions à maintenir un niveau de rigueur méthodologique suffisant dans la ré-analyse des données de la recherche, en raison notamment de contraintes temporelles ; or, l’appropriation puis l’analyse de ces matériaux (des données, de leur contexte de production) est un exercice très chronophage. À priori, l’appropriation des résultats publiés (articles, colloques, synthèses) semble plus commode et nous pouvons nous interroger sur l’intérêt que pourrait accorder ces institutions à un matériau source.

D’autre part, on peut s’interroger sur la compatibilité entre la réutilisation de données par ce type de structures et les valeurs de « désintéressement » mentionnées précédemment. Diverses expériences (comme l’étude *Familles en campagne* menée en 2015 par l’agence d’urbanisme de l’agglomération de St Nazaire – ADDRN, 2015) montrent que certaines recherches menées dans des bureaux d’études sont finalement assez proches des exigences et objectifs de travaux universitaires. Néanmoins, on ne peut pas tout à fait exclure que d’autres travaux soient fortement orientés par « la commande », par les

systèmes de représentation, les attentes ou les enjeux formulés par la « commande publique », rendant difficile le travail de déconstruction des cadres d'analyse que présuppose le travail scientifique d'objectivation. Notons par ailleurs que ceci ne préjuge en rien de la qualité ni de la valeur du travail fourni, pour les commanditaires, mais aussi parfois pour les chercheurs.

On s'interroge également sur le critère déontologique de « contrôle par les pairs » évoqué ci-dessus : si certains chercheurs salariés de ce type de structures (travaillant dans les départements de recherche d'EDF ou d'Orange Lab par exemple) respectent ce critère en publiant les résultats de leurs travaux et en les soumettant à leurs pairs lors de colloques et journées d'étude, il nous semble néanmoins peu probable qu'un salarié de telles structures puisse être totalement indépendant et libre dans ses recherches et ne subisse aucune pression sur les résultats à produire.

Enfin, un autre problème provient des incertitudes sur la possible mise en œuvre de relations de réciprocités entre ces institutions et le monde académique. Si les données issues du monde académique pourraient servir ces études, sommes-nous certains qu'en retour, ces institutions seraient à même d'être elles-mêmes dans une logique de partage et d'échange, en dehors d'un cadre ponctuel et contractuel très précis ? Les expériences dont nous avons connaissance ne semblent pas aller dans ce sens.

En raison de l'ensemble de ces interrogations et du peu de visibilité sur la manière de faire la recherche dans ces institutions, il nous semble judicieux d'user d'un principe de précaution : ainsi, nous n'envisageons pas d'ouvrir les données de la recherche à quelque structure assurant des missions de service public.

#### Cas de l'open-data

Partager les données en *open data*, i.e. laisser les données en libre accès total, reviendrait à n'exercer aucun contrôle ni sur le profil des personnes pouvant utiliser ces données, ni sur la réutilisation qui en serait faite.

Plusieurs raisons justifient notre profonde opposition à une telle diffusion des données.

En premier lieu, on retrouve ici les blocages liés aux compétences requises et aux finalités de réutilisation évoquées supra. D'une part, il ne nous semble pas que le « grand public » ait les compétences scientifiques lui permettant d'analyser avec rigueur les données partagées. Rappelons-le, ces données ont été collectées dans des cadres méthodologiques précis et complexes, en articulation avec une problématique de recherche ; ouvrir largement les données pourrait laisser entendre que celles-ci peuvent être immédiatement accessibles et traitées, sans méthode, mais aussi sans « critique des sources », c'est-à-dire sans connaissance précise des conditions et des raisons pour lesquelles elles ont été collectées. D'une certaine manière, l'ouverture des données reviendrait à donner l'*illusion* – fortement dénoncée en médecine par exemple – suivant laquelle les données de la recherche sont appropriables par tous, sans condition ; et, d'une certaine manière, à jeter sur la construction scientifique un relatif discrédit, considérant que les données de la recherche seraient une *production comme une autre* à laquelle quiconque pourrait donner sens sans coût d'accès. Il ne s'agit pas pour autant d'ériger le simple citoyen en « idiot culturel » mais d'affirmer que l'exploitation des données de la recherche (à distinguer des résultats) n'a de valeur analytique qu'à l'intérieur d'une démarche scientifique régulée par des règles, des pratiques et des savoirs spécifiques dont la maîtrise est nécessaire pour pouvoir leur donner sens. Une autre question est posée, celle du dévoiement par rapport aux objectifs de connaissance portés initialement par l'enquête. Si elles étaient en libre accès, les données pourraient être utilisées dans n'importe quel objectif, sans procédure de contrôle. Rien ne pourrait encadrer ou réguler l'usage commercial de ces données. On ne peut exclure par ailleurs qu'elles soient utilisées à « mauvais escient », dans un objectif de voyeurisme ou de délation par exemple.



Une deuxième ligne d'arguments prend en considération le point de vue de l'interviewé. Au départ, ces données n'ont pas été produites ni collectées dans l'objectif d'être diffusées en libre accès. Les enquêtés se sont exprimés dans un cadre défini, à finalité scientifique et ont livré leur parole à un enquêteur qui s'est engagé à garantir la confidentialité de leurs propos. En vertu du respect du contrat de confiance, cette finalité doit être respectée. Si l'objectif de diffusion en ligne avait été mentionné dès le départ comme c'est le cas par exemple dans le projet "Raconter la vie" lancé par l'historien Pierre Rosanvallon, il est par ailleurs très probable que les enquêtés ne se seraient pas exprimés de la même manière, voire n'auraient pas tenu les mêmes discours. En effet, l'idée même que le discours recueilli puisse être médiatisé modifierait en profondeur la signification de l'enquête. Certains biais, telle l'illusion biographique (Bourdieu)<sup>5</sup>, seraient démultipliés et pourrait affaiblir la dimension analytique et réflexive de l'entretien, en renforçant considérablement la propension à l'héroïsation du quotidien et à la mise en scène de soi ; ou encore, à l'élaboration d'une parole militante ; il est alors probable que la matériau produit s'éloigne des objectifs compréhensifs de mise en intelligibilité des pratiques par les discours qui légitiment la démarche d'enquête par entretien.

En définitive, l'objectif de diffusion des données qualitatives en *open data*, bien qu'il puisse séduire en première lecture, rencontre de multiples blocages, liés au respect de l'intégrité de la démarche scientifique, à la protection des enquêtés et des données personnelles ou encore aux contradictions évidentes entre les objectifs « réflexifs » et les objectifs de médiatisation.

Pour résumer nos propos, à un tel niveau de réflexion sur le partage des données de la recherche, nous préconisons d'adopter une position plutôt restrictive, qui nous semble la plus à même de garantir le respect des enquêtés et l'intérêt scientifique/intérêt général du partage. Pour l'ensemble des raisons exposées ci-dessus, nous préconisons donc de restreindre l'accès aux données aux demandes émanant du monde de la recherche, et uniquement dans le cas de projets à visée scientifique ou pédagogique ; de ne pas répondre favorablement aux demandes émanant des mondes de l'entreprise ou des structures assurant des missions de service public.

En plaçant la réflexion à un niveau plus général, on peut également arguer que, s'il est certain que les recherches en SHS peuvent être utiles à l'intérêt général et valorisées auprès des différentes structures œuvrant en ce sens, il ne nous semble pas forcément évident que la réutilisation des données proposées en partage (entretiens anonymisés notamment) soit le procédé le plus efficace. Les analyses synthétiques produites par les chercheurs (producteurs et/ou ré-utilisateurs de ces données) et les échanges "publiés" qu'il peut y avoir au sein de la communauté scientifique, sont mieux à même de servir, sans doute, les objectifs de ces structures.

Dans le cas où une demande d'accès aux données serait rejetée, nous préconisons donc de proposer plutôt une discussion avec le chercheur-producteur autour des résultats qu'il a produits. La base de données servirait donc dans ce cas d'outil de rapprochement et de mise en connexion entre les sphères académique et non-académique.

## **2. Protocole de mise à disposition**

### *a. Quelle structure pour gérer la mise à disposition des données ?*

La mise à disposition des données de la recherche implique trois types d'acteurs : les chercheurs-primaires (ou chercheurs-producteurs de données), les chercheurs-demandeurs, et une structure qui

---

<sup>5</sup> Qui amène aujourd'hui les chercheurs à traiter les entretiens avec discernement

assurera le lien, indispensable pour les projets de grande ampleur (à l'instar de Qualidata, beQuali, Quételet).

Une question importante est celle de la répartition des rôles entre cette structure médiatrice et le chercheur-primaire. Qui assure l'anonymisation des données ? Qui s'assure du consentement des enquêtés ? Qui, enfin, décide d'accorder ou non l'accès aux données ? Derrière cette répartition, plusieurs questions : celle des responsabilités endossées par chacun, notamment en rapport aux enjeux juridiques de la diffusion des données ; celle du lien du chercheur à son matériau : un chercheur acceptera-t-il de confier la responsabilité morale et le respect du contrat de confiance à un tiers ? ; celle, enfin, de l'équilibre entre institutions de la recherche et chercheur, entre collectif et individuel : comment assurer une gestion des données de la recherche qui soit démocratique tout en prenant en compte le caractère spécifique de chaque matériau ?

Les structures de partage existantes ont opté pour des fonctionnements différents, qui induisent divers enjeux.

Qualidata choisit de faire endosser les principales responsabilités au chercheur-primaire, à qui est confié le soin d'anonymiser ses données et d'obtenir le consentement des enquêtés. Ensuite, la structure gère seule la mise à disposition des données ; on peut néanmoins supposer que les règles très strictes fixées en termes de profils et objectifs de réutilisation limitent les cas de demandes problématiques. Les sondages réalisés dans les universités de Rennes et Lille et les échanges informels que nous avons pu avoir avec divers chercheurs, qui laissent apparaître des réticences sur la diffusion des données au-delà du cercle académique, nous font penser que cette structuration est susceptible de rencontrer une adhésion assez large de la part des chercheurs.

beQuali opte pour le travail partenarial : pendant le travail « d'édition » des données, le chercheur-primaire est consulté pour la phase d'anonymisation des données, ainsi que pour la réalisation de « l'enquête sur l'enquête ». Ils sont en revanche seuls responsables de la sollicitation des enquêtés pour obtenir leur consentement à la mise en partage des données. Enfin, beQuali prend en charge la procédure d'instruction des demandes d'accès.

On pourrait également imaginer des structures au fonctionnement plus ouvert, en particulier au sujet de l'ouverture envers les acteurs extérieurs au monde académique ; il nous semble qu'un tel fonctionnement encourt un risque d'exclusion des chercheurs les plus réticents (pour différentes raisons) : cela imposerait alors la nécessité d'intégrer la possibilité d'un veto du chercheur primaire.

Comment nous positionnons-nous ? Dans la perspective d'assurer le succès d'une base de données auprès des chercheurs, nous nous rapprochons de l'exemple de Qualidata et de sa position plutôt restrictive en termes d'ouverture des données. En termes de fonctionnement « administratif » de la base, nous préconisons de s'inspirer de l'exemple de beQuali en formant un conseil scientifique impliquant des techniciens assurant la gestion quotidienne de la base, un référent CIL, et des chercheurs ayant une bonne connaissance des données mises en partage et pouvant instruire les demandes d'accès aux données en toute indépendance. Géré ainsi par un collectif indépendant (voire élu démocratiquement), l'accès aux données ne dépend donc pas uniquement du bon vouloir ou des inimitiés du chercheur primaire (qui pourraient le conduire à interdire l'accès à certains chercheurs). Une question demeure, non tranchée parmi les auteurs de ce rapport : *quid* des demandes d'accès qui seraient controversées au sein du conseil scientifique ? Le chercheur primaire doit-il rester décisionnaire final, en conservant un droit de veto ? Au contraire, doit-on considérer que, une fois ses données confiées à la structure de partage, le chercheur s'efface au profit du collectif ?

Quoi qu'il en soit, nous préconisons d'impliquer le chercheur-primaire autant que possible pendant le processus de partage : au cours de l'intégration de son enquête dans la base (à l'image des « enquêtes

sur l'enquête » de beQuali) et également (voire surtout) une fois la demande acceptée, en encourageant les échanges entre chercheurs primaire et réutilisateur. Sur ce dernier point, notre point de vue semble largement partagé par la communauté scientifique (Dargentas et al, 2007 ; Duchesne, 2017 ; Bishop, 2014). Les échanges entre chercheurs primaire et secondaire limiteraient le problème du « *not having been there* » (Heaton, 2008) en étant « *bénéfiques pour la compréhension totale des contextes de recherche initiaux et pour la meilleure compréhension des résultats produits lors d'une analyse secondaire* » (Dargentas et al, 2007) ; ils permettraient également de limiter les risques ou craintes évoquées par les chercheurs primaires de critiques de la part du chercheur secondaire envers le chercheur ou ses travaux.

Au-delà de ces préconisations, bénéficié de retours d'expérience de partage (encore rares) permettra de faire évoluer les modalités de partage.

### b. *Information ou consentement*

Le règlement européen entrant en application en mai 2018 impose au chercheur d'informer l'enquêté de l'utilisation qui va être faite des données collectées, et de lui laisser la possibilité de s'y opposer ; il n'est en revanche pas tenu (juridiquement) d'avoir une preuve du consentement de l'enquêté.

Ce nouveau règlement semble alléger les contraintes juridiques qui pèsent sur les épaules du collecteur de données et simplifier les démarches à mettre en œuvre pour utiliser ou partager les données de la recherche. Néanmoins, ceci ne doit pas conduire le chercheur à négliger les enjeux déontologiques que nous avons évoqués *supra* et ses *devoirs* envers l'enquêté.

Par respect pour l'enquêté et en raison des valeurs déontologiques qui régissent la recherche, il nous semble primordial d'informer le plus précisément possible l'enquêté des conditions dans lesquelles son discours sera utilisé et/ou partagé ; plus, d'obtenir un consentement, véritablement éclairé, de la part de l'enquêté. Ainsi, pour des structures de partage qui envisageraient de partager librement les données au sein de la sphère universitaire ou avec des structures privées, voire en *open data*, nous considérons qu'il serait de leur devoir de chercheur de solliciter l'approbation de l'enquêté et de se conformer à sa réponse.

Le règlement européen précise également que c'est à la personne qui "traite" les données d'informer ou d'obtenir le consentement des enquêtés ; a priori, cela concerne également les chercheurs qui réutiliseraient les données plusieurs années après leur collecte. Il nous semble qu'une telle mesure est techniquement impossible à mettre en œuvre (comment recontacter des enquêtés plusieurs années après ? en particulier, comment faire recontacter des enquêtés par un chercheur-secondaire, avec qui les enquêtés n'ont jamais eu aucun contact ?) ; plus, il nous semble qu'elle n'est pas réellement souhaitable : à partir du moment où un enquêté a accepté de donner sa parole à la science, nous pensons qu'il devrait bénéficier d'un droit à l'oubli.

### c. *Modalités d'anonymisation ou de « désidentification »*

Trois scénarii de protection des personnes pourraient être envisagés, qui impliquent différents enjeux en termes de respect du contrat de confiance et de garantie de l'intérêt scientifique des données.

1<sup>er</sup> scénario – Anonymisation optimale, mais nécessairement partielle : Le matériau livré aux demandeurs est anonymisé au départ par les soins du chercheur-primaire (producteur des données) selon ses propres critères et ses besoins. Ce scénario préserve le contrat de confiance, car c'est l'auteur des enquêtés qui anonymise une fois pour toutes, et de manière irréversible ; néanmoins,

l'anonymisation restreint la diversité et la profondeur des analyses qui peuvent être faites puisque certaines données « manquent ».

2<sup>ème</sup> scénario – Anonymisation au cas par cas : L'anonymisation est faite par un tiers à chaque requête d'un demandeur, en fonction de ses besoins et de ses demandes, ce qui permet « d'épargner » les données-clés : l'anonymisation est faite de telle façon que l'anonymat est conservé, tout en permettant le traitement par le demandeur. Ce scénario, plus ouvert que le premier, permet d'adapter au mieux le matériau ; il implique néanmoins deux choses : d'une part, une structure importante, en termes financiers et de personnel, pour assurer l'anonymisation à la demande ; d'autre part, de faire circuler des matériaux non anonymisés auprès d'acteurs intermédiaires, ce qui constitue une entorse au respect du contrat de confiance.

3<sup>ème</sup> scénario – Non anonymisation : Le matériau est livré totalement “brut” (non anonymisé) au demandeur, avec lequel un contrat redéfinit les règles de confidentialité à suivre pour la valorisation ultérieure des matériaux. Ce scénario maintient une grande latitude dans le traitement, mais ne respecte absolument pas le contrat de confiance passé initialement avec l'enquêté, ni la nécessaire protection des personnes et des données personnelles. Dans ces conditions, il n'est pas envisageable d'un point de vue déontologique de soutenir ce scénario.

Pour les raisons évoquées précédemment, c'est le 1<sup>er</sup> scénario qui nous semble le plus équilibré. Nous préconisons donc qu'avant tout partage de leurs données (dans les conditions précisées ci-dessus), les chercheurs anonymisent leur matériau, si possible dès leur collecte, et dans tous les cas, avant toute diffusion. Selon quelle méthodologie ? En se référant à leurs propres besoins, i.e. en supprimant ou brouillant les données qui ne leur sont pas indispensables. Dans le cas d'une base de données thématique sur les mobilités quotidiennes, nous préconisons ainsi, pour les entretiens, de pseudonymiser l'ensemble des noms propres, en particulier celui des enquêtés et de leurs enfants ; de remplacer la profession par une catégorie de PCS ; de brouiller l'adresse exacte du lieu de résidence et du lieu d'emploi. Pour les traces GPS, nous préconisons de brouiller les lieux de résidence et d'emploi, qui sont les plus informations les plus identifiantes ; mais de ne brouiller ni les autres lieux d'arrêts ni les trajets.

Ce scénario permet selon nous de concilier les enjeux juridique et déontologique avec les intérêts scientifiques du partage ; c'est également celui qui nous semble le plus rationnel au vu du temps que peuvent raisonnablement consacrer les chercheurs et les structures de partage à l'anonymisation. Il n'en demeure pas moins que l'anonymisation reste extrêmement chronophage, en particulier lorsqu'elle est réalisée sur des données existantes (ce qui nécessite de se « replonger » dans des données qui peuvent avoir été collectées longtemps auparavant). Bien que théoriquement nécessaire, elle peut alors constituer un vrai frein au partage et rebuter les chercheurs qui souhaiteraient diffuser leurs données. Alors, pourrait-on envisager un protocole d'anonymisation « *a minima* » pour les données existantes, en considérant que le temps passé depuis la récolte des données constitue une protection supplémentaire des enquêtés et de leurs paroles ? Pour ces données existantes, pourrait-on se contenter d'une pseudonymisation associée à une anonymisation des données sensibles ?

#### d. Contrat de réutilisation, charte d'usage de la donnée

Nous voudrions revenir dans cette sous-partie sur notre positionnement vis-à-vis d'un certain nombre de questions que pose la rédaction d'un contrat de réutilisation des données.

i. La première question est celle du choix des différentes parties signataires du contrat de réutilisation. Deux options pourraient être envisagées : d'une part, un contrat entre le chercheur primaire et la personne souhaitant réutiliser ses matériaux d'enquête ; d'autre part, un contrat entre la personne souhaitant réutiliser les matériaux d'enquête et une institution, à laquelle le chercheur primaire délèguerait la tâche de s'assurer du respect de conditions déontologiques et méthodologiques de la réutilisation des données.

C'est cette dernière option qui a été suivie par la banque de données qualitatives beQuali : elle a pour avantage de faciliter la réutilisation des données, notamment dans les cas où le chercheur primaire ne participe plus, pour diverses raisons, à la sphère scientifique (décès, retraite, mutation, etc.). Pour notre part, nous pensons que l'absence de participation à l'opération de collecte des matériaux est propice à l'affranchissement de certains garde-fous. Aussi, le dispositif le mieux à même de garantir et de faire circuler le contrat de confiance initialement lié avec l'enquêté (voir *infra*) nous semble être un contrat de réutilisation rappelant une éthique (des règles) d'utilisation des matériaux, qui serait co-signé par la structure de partage, le chercheur réutilisateur et le chercheur primaire, lorsque celui-ci est encore en vie. Ce contrat tri-partite engagerait les chercheurs réutilisateurs, et apparaîtrait également pour le chercheur primaire comme l'expression et la réaffirmation de l'engagement moral qu'il a pris vis-à-vis de ses enquêtés. En outre, faire signer ce contrat au chercheur primaire impose évidemment de le contacter, parfois plusieurs années après qu'il ait déposé ses données : ceci donne donc une indication au chercheur primaire de l'usage qui est fait de ses données, ce qui semble apprécié d'après l'expérience de L. Bishop (2014 : 173).

Une question importante se pose dans le cas des recherches collectives, dont la spécificité peut bloquer l'exploitation ou la diffusion des données (Serres *et al.*, 2017b : 44). La contractualisation doit-elle se faire avec le coordinateur scientifique du projet, quand bien même celui-ci n'a pas participé de manière intensive à la collecte des matériaux ? Avec l'ensemble des chercheurs ayant participé au projet, et surtout à la collecte des matériaux ?

ii. La seconde question, concomitante, est celle des échanges prévus dans le contrat entre le chercheur primaire et les ré-utilisateurs des données. Dans le cas de beQuali, si le chercheur primaire peut « demander à être informé régulièrement de l'utilisation des données déposées et diffusées », il n'est pas tenu systématiquement au courant des demandes de réutilisation. Comme nous l'avons plusieurs fois mentionné, nous pensons au contraire que le contrat de réutilisation devra viser à favoriser les échanges entre le chercheur primaire et les réutilisateurs de ses données, afin de s'assurer notamment d'une bonne compréhension des conditions de production des données et de favoriser les collaborations scientifiques entre producteurs et réutilisateurs des données (même si certains outils, tels que l'Enquête sur l'enquête réalisée par beQuali, participent de cette bonne compréhension du contexte).

iii. Par ailleurs, est-il nécessaire de partager un même positionnement épistémologique vis-à-vis de la réutilisation des données (voir *infra*, 1.c) pour contractualiser ? Si la restriction de la diffusion des données à des chercheurs partageant le même positionnement théorique que le chercheur primaire nous semble contraire aux objectifs de notre projet et au fonctionnement de la sphère scientifique, nous pensons qu'il est important de rappeler dans le contrat de réutilisation les principes épistémologiques de base sur lesquels nous nous appuyons (voir *infra*). A l'instar de beQuali, qui insiste fortement sur l'importance du module « enquête sur l'enquête » dans la compréhension des données mises à disposition, nous pensons fondamental de rappeler dans le contrat de réutilisation que les données ne peuvent être dissociées de leur contexte de production et qu'il est recommandé de se référer au module explicitant les conditions et objectifs scientifiques des recherches menées.

iv. Quelles règles déontologiques doivent apparaître dans le contrat ? En nous inspirant du contrat de réutilisation des dispositifs beQuali et du centre Quetelet, devront figurer dans le contrat le souci de ne pas chercher à lever l'anonymat des enquêtés et à protéger cet anonymat lors de la diffusion des résultats, le respect de la confidentialité des données (ne pas céder les données à un tiers et assurer une protection matérielle des données) et du droit d'auteur du chercheur primaire, ainsi que l'engagement à utiliser les données exclusivement dans une finalité de recherche. Certaines règles déontologiques pourront être propres à des projets de réutilisation des données dans un cadre strictement pédagogiques, notamment l'engagement de ne pas diffuser les matériaux bruts aux étudiants et/ou de leur faire signer eux-mêmes une charte de réutilisation.

v. La dernière question est celle du respect et de la valorisation du travail du chercheur primaire. Le contrat devra bien mentionner que le chercheur primaire reste le seul propriétaire des matériaux d'enquête, associés dans l'idéal à un DOI (voir *supra*). Comme dans le dispositif beQuali, le réutilisateur des données devra s'engager à « *citer précisément la référence de l'enquête, son titre et le chercheur producteur dans toutes les productions scientifiques (publication, communication, enseignement) réalisées à partir des documents de l'enquête* ». Si une possibilité d'embargo provisoire de publications est également prévue dans ce dispositif, lorsque les matériaux d'enquête sont déposés peu de temps après les premières exploitations, nous pensons qu'une citation obligatoire de l'article *princeps* du chercheur primaire, comme cela est pratiqué en épidémiologie, constitue également une forte marque de respect vis-à-vis du travail de ce chercheur, et donc une incitation supplémentaire à la mise à disposition des données. En effet, comme nous allons le voir dans la section suivante, il existe de nombreuses craintes de chercheurs d'être dépossédés de leurs données, expliquant la faible mise à disposition des matériaux d'enquête.

### **3. Modalités de partage dans le cas d'une mutualisation**

Nous avons explicité dans les deux parties précédentes nos positionnements au regard de la diffusion et du partage des données de la recherche, en termes de philosophie et de mise en œuvre concrète du partage. Nous nous sommes placés dans un contexte de diffusion unilatérale, de *formal data sharing* (Heaton, 2008) où les données sont récupérées par des chercheurs réutilisateurs sur une structure de stockage (et peuvent, s'ils le souhaitent, rentrer ensuite en contact avec les chercheurs primaires).

Nous souhaitons maintenant aborder le cas spécifique de la mutualisation, i.e. de la mise en commun de données. Cette situation, qui est notamment celle que nous connaissons dans le cadre du méta-projet<sup>6</sup>, représentait en 2008 la majorité des expériences d'analyse secondaire, comme le montre une analyse menée par Heaton (2008) sur 65 études. Il nous semble en outre souhaitable de favoriser ce type de partage, qui participe du développement de la recherche en réseaux.

Dans cette situation, les chercheurs sont à la fois réutilisateurs et primaires, fournisseurs de certaines des données partagées. La principale différence que nous établissons entre *partage* et *mutualisation* réside donc dans la réciprocité dans les échanges. Cette réciprocité ainsi que le travail en commun

---

<sup>6</sup> Pour rappel, l'ambition initiale du méta-projet (qui est à l'origine de toutes les réflexions développées précédemment) était de mutualiser les données produites au cours de six programmes de recherche sur les mobilités quotidiennes dans les territoires ruraux, périurbains et urbains, menés depuis une dizaine d'années au laboratoire CITERES (Université de Tours). Les objectifs de ce méta-projet rejoignent ceux exposés au début de ce rapport ; ce sont en particulier des objectifs scientifiques : valider ou invalider certains résultats à partir d'un panel de données plus large et optimiser des données qui seraient « non valorisées » en les analysant au prisme de nouvelles questions de recherche. Ce méta-projet regroupe six chercheurs et un cartographe ayant tous participé à un ou plusieurs des six programmes de recherche, et travaillant tous (sauf un) à l'Université de Tours.

associé à la mutualisation impliquent une forme de contrôle entre chercheurs, qui limitent les risques de dévoiement des données.

En termes de mise en œuvre concrète du partage, quelles différences établissons-nous entre diffusion et mutualisation ?

En premier lieu, il nous semble primordial de conserver le principe de contractualisation et d'engagement entre les chercheurs, qui signeraient alors un contrat de mutualisation reprenant les principaux éléments du contrat de réutilisation. (Insérer contrat de réutilisation signé par les membres du méta-projet).

Se pose ensuite la question du respect du contrat de confiance initial : peut-on considérer qu'il s'étend aux autres chercheurs participant au projet de mutualisation ? Concernant la resollicitation des enquêtés : nous pouvons considérer que le consentement initialement donné par les enquêtés au chercheur primaire pour l'analyse des matériaux collectés est toujours valable dans le cadre d'une mutualisation qui associe le chercheur primaire ; à ceci s'ajoute le fait que, dans une situation de mutualisation, on conserve bien les objectifs scientifiques fixés initialement. Aussi, nous pensons qu'il n'est pas nécessaire de recontacter les enquêtés (ce qui, comme nous l'avons déjà mentionné, peut de toute façon s'avérer difficile à faire). En revanche, l'impératif de protection des enquêtés (*via* l'anonymisation des données) nous semble devoir être conservé, y compris dans le cadre d'une mutualisation. Et ce d'autant plus que la mutualisation se fait en local (ce qui est notamment le cas du méta-projet), que les possibilités de ré-identification des enquêtés sont les plus fortes et ses implications les plus problématiques. Concernant le méta-projet, en raison du caractère chronophage de l'anonymisation, nous allons néanmoins nous orienter vers une anonymisation *a minima*, i.e. une pseudonymisation associée à une suppression des informations les plus sensibles (conviction religieuses et données de santé notamment). Ce faisant, nous avons bien conscience de ne pas respecter parfaitement le principe de protection des enquêtés que nous venons d'évoquer.

Dernier point concernant la mutualisation. On peut supposer que la connaissance des données existantes, qui se traduit par le souhait d'initier un travail en commun et de mutualiser des données, est souvent permise par une interconnaissance entre chercheurs. Est-ce que ces relations d'interconnaissance ont un impact sur les modalités concrètes du partage ? Autrement dit, qu'est-ce que ces relations de confiance changent par rapport à de la diffusion avec un chercheur-réutilisateur anonyme ?

Ainsi, les sept chercheurs travaillant sur le méta-projet se (re)connaissent, ont tissé des relations de confiance et se côtoient (voire travaillent ensemble) quasi-quotidiennement ; ceci pourrait nous conduire à être plus laxistes dans la gestion de nos données et le respect du contrat de confiance : à envisager par exemple de mutualiser des données non anonymisées, considérant qu'aucun d'entre nous ne va chercher à nuire aux divers enquêtés. Néanmoins, nous pensons que cette confiance partagée ne doit rien changer aux protocoles de diffusion et de mutualisation évoqués précédemment, car l'interconnaissance ne justifie en rien de s'affranchir des règles énoncées plus haut. Elle peut au contraire conduire à des malentendus entre chercheurs, avec des interprétations différentes de ce qui sera fait des données partagées ou du degré de diffusion de ces données vers l'extérieur. Un contrat de réutilisation des données sera ainsi signé entre les sept chercheurs du méta-projet.

## IV. Quels facteurs de blocage à la diffusion ? Comment les dépasser ?

### 1. Le jeu de la diffusion des données en vaut-il la chandelle pour le chercheur ?

Les enquêtes menées à l'Université Rennes II et Lille III, ainsi que le retour d'expérience britannique montrent que la plupart des chercheurs adhère aux objectifs scientifiques de mise à disposition des données à leurs pairs et à son importance dans la mise en débat de leurs résultats (voir *infra*, I.1.) : dans l'enquête menée à Rennes, seuls 27% des répondants déclarent s'opposer à la pratique. Néanmoins, le nombre de chercheurs qui diffusent leurs données qualitatives reste mesuré. A Rennes, 46% des chercheurs-répondants (soit une soixantaine) avaient "déjà partagé leurs données avec d'autres chercheurs", et 24% "déjà déposé leurs données en libre accès" ; dans l'enquête de Lille, ils étaient 34% (soit environ 70) à déclarer "partager leurs données avec des collègues. Comment l'expliquer ?

Le nombre mesuré de chercheurs diffusant leurs données s'explique tout d'abord par des craintes vis-à-vis de la mise en pratique concrète de cette diffusion (Bishop, 2013) ; autrement dit, par des contraintes objectives à sa mise en œuvre, liées à un manque de temps dans un contexte d'accroissement des tâches administratives (la publication d'articles, ouvrages, communications étant prioritaire par rapport au partage des données), à une impossibilité juridique (de partager des données à caractère personnel par exemple) et à un manque de structures idoines.

Il s'explique également par un certain nombre de perceptions négatives, de peur évoquées par les chercheurs : en particulier, celle d'être dépossédés, juridiquement et scientifiquement, de leurs données.

Cette crainte de la dépossession s'explique en partie par la spécificité du recueil des matériaux par les chercheurs de sciences sociales utilisant les méthodes qualitatives. Ce recueil repose sur une interaction, plus ou moins prolongée, entre un chercheur et des enquêtés. Cette interaction peut durer quelques heures, dans le cas d'un entretien, ou plusieurs mois, voire années, dans le cadre du travail ethnographique. Afin de rendre possible et de nourrir cette interaction, le chercheur s'engage dans un travail de persuasion, d'écoute, d'échanges, de dévoilement d'une partie de sa personnalité. Cet engagement entraîne un lien affectif important du chercheur à ses matériaux d'enquête (Serres et al, 2017 : 9). Au-delà de la crainte de ne pas respecter le contrat de confiance lié avec les enquêtés lors de la recherche (voir *infra*, I.2.), ce rapport, souvent intime, des chercheurs à leurs matériaux d'enquête nourrit une appréhension à mettre ses données à disposition d'autres chercheurs (Bishop, *ibid*).

Cette appréhension est nourrie par la crainte que les ré-utilisateurs plagient (Serres et al, 2017 : 18), ou critiquent, scientifiquement ou éthiquement, la manière dont les données ont été collectées. Une majorité de chercheurs s'appuyant sur des méthodes qualitatives semble ainsi partager une certaine ambivalence vis-à-vis de la transparence des démarches scientifiques (Bishop, *ibid*) : la plupart sont favorables à cette transparence (ne serait-ce que pour éviter les cas de fraude scientifique, de falsification, de sélection ou d'exagération des données recueillies sur le terrain), mais ils craignent dans le même temps de livrer aux regards extérieurs les « coulisses de la recherche » et leurs « secrets de fabrication », en raison des dommages potentiels pour les personnes enquêtées, ainsi que pour leur propre réputation. Cependant, les exemples anglo-saxons de réutilisation de matériaux d'enquête pouvant nuire, des années après leur collecte, à la réputation de l'enquêteur (Bishop, *ibid*) semblent au contraire montrer une bienveillance des nouveaux chercheurs vis-à-vis de leurs aînés : dans une logique kantienne, ils souhaitent traiter l'enquêteur avec la bienveillance et la compréhension du contexte d'enquête initial, dont ils souhaiteraient eux-mêmes bénéficier si leurs propres matériaux d'enquête étaient diffusés. Néanmoins, il est difficile de nier qu'en diffusant ses matériaux d'enquête,



le chercheur prend le risque de donner à voir toute l'ambiguïté des relations qu'il peut entretenir avec les enquêtés, ambiguïté qui est généralement cachée dans les articles ou les ouvrages scientifiques tirés de ses recherches. Un bon exemple en est l'absence de mention, et donc *a fortiori* d'analyse méthodologique, dans la quasi-totalité des travaux sociologiques de l'ensemble des éléments d'enquête ayant quelque chose à voir avec la sexualité, alors que la relation d'enquête peut être considérée comme un contexte particulier favorisant l'érotisation des échanges : le chercheur « *adresse à un autre homme ou à une autre femme une demande de relation particulière ; cette demande est souvent insistante et, pour être acceptée, elle conduit généralement l'enquêteur-trice à se faire séduisant-e (...) Une fois que l'entretien est accepté, très souvent l'enquêteur-trice s'efforce de le situer dans un endroit calme, privé ou au moins isolé : la demande de retrait par rapport au regard d'autrui est plus que fréquente, elle peut être interprétée comme une demande d'exclusivité (ce qu'elle est en partie), d'intimité (ce qu'elle est aussi)* » (Clair, 2016). Cet exemple des possibles non-dits autour de la sexualité dans le travail de terrain est un cas extrême. Néanmoins, il est emblématique d'une gêne plus générale que pourrait ressentir le chercheur à la publicisation de sa relation avec les enquêtés: les réactions qu'il a pu avoir à certains propos, la manière dont il a pu orienter la discussion pour obtenir des informations, les moments où il s'est éloigné de son rôle de chercheur, etc.

Le chercheur ne prend-il pas des risques en donnant à voir toutes les facettes de son engagement auprès des enquêtés ? Pourquoi prendre ce risque, alors qu'il n'existe actuellement en France aucune rétribution symbolique dans le champ de la recherche à la mise à disposition des données de recherche ?

En effet, le déficit de reconnaissance professionnelle à la diffusion des données constitue un second facteur de blocage, comme l'a notamment mis en évidence l'enquête sur les données de la recherche menée à l'Université Rennes 2 (Serres et al, 2017 : 8). En aménagement, géographie et sociologie, la publication des données de la recherche et leur diffusion ne sont pas reconnues comme des « publications » dans le CV des chercheurs par les laboratoires de recherche français et les principales instances d'évaluation (HCERES notamment). La mise à disposition des données ne contribue donc nullement à valoriser les carrières des chercheurs et enseignants-chercheurs de ces disciplines, alors qu'en histoire par exemple la constitution de catalogues de données est considérée comme une contribution importante à la recherche. Cette absence de reconnaissance professionnelle est d'autant plus préjudiciable dans des disciplines qui ont pu être marquées par une division du travail entre des étudiants et/ou des jeunes chercheurs réalisant les enquêtes de terrain, et des professeurs plus âgés écrivant en leur nom propre à partir de ces enquêtes. Elle s'inscrit également dans un contexte de pénurie de recrutement et de financement de la recherche par projet, qui contribue à accentuer les hiérarchies internes au champ des chercheurs. L'absence d'incitation à la mise à disposition des données peut amplifier la crainte de chercheurs d'être des « soutiers » de la recherche, investissant beaucoup de temps et d'énergie dans la collecte des données, au détriment parfois de sa vie privée (Serres et al, 2017b: 42), alors que d'autres profiteraient de la mise à disposition pour s'épargner cette phase essentielle et chronophage de la recherche scientifique. Cette crainte est sans doute d'autant plus justifiée dans le champ des études urbaines, où les « essayistes » sont légion. Cette crainte pourrait également expliquer l'influence de l'âge dans les pratiques de mise à disposition des données : ce sont les chercheurs de plus de 50 ans qui sont le plus favorables à cette mise à disposition (Serres et al, 2017b: 75), peut-être parce qu'ils ont des positions plus établies dans le champ (ainsi peut-être qu'un temps plus important à y consacrer).

Enfin, mentionnons un troisième facteur de blocage mentionné par les chercheurs interrogés : un frein épistémologique lié à la définition même des « données de recherche » collectées, i.e. leur caractère construit entraînant l'impossibilité scientifique de réutiliser des données sans avoir participé à leur

collecte, par méconnaissance du contexte de production de ces données (Serres et al, 2017b : 18 ; voir supra, 1.c.).

## 2. Comment dépasser ces blocages ?

Afin de dépasser ces blocages, des réponses peuvent être apportées à trois niveaux pour développer une culture de la diffusion parmi les chercheurs travaillant sur les questions de mobilité.

Une première condition nécessaire, mais non suffisante, est, logiquement, la mise en œuvre de politiques de gestion des données à l'échelle nationale ou à celle plus locale des laboratoires de recherche. Celle-ci suppose un effort financier, afin que les chercheurs puissent être accompagnés dans leur démarche d'archivage par des professionnels. Cet accompagnement permettra de rassurer les chercheurs sur le devenir de leurs matériaux d'enquête et l'investissement temporel nécessaire à la mise à disposition des données dans des conditions éthiques satisfaisantes. Développer une culture du partage des données demande également une évolution de la manière de faire la recherche actuellement (en tout cas en France) : le temps important nécessaire à la mise en partage (notamment à l'anonymisation) des données n'est aujourd'hui pas prévu parmi l'ensemble des tâches qui incombent aux chercheurs. Plus largement, le temps de valorisation des résultats de la recherche (*a fortiori* le temps de valorisation des matériaux, des données) est rarement prévu lors de l'élaboration des projets : ainsi l'écriture d'articles, ou plus encore de livres collectifs, se fait souvent « en plus », à l'issue des projets, pendant les temps libres des chercheurs. Pour favoriser la valorisation des matériaux de recherche, il est donc impératif de « considérer » ce temps de valorisation, de le prévoir et de le formaliser au moment de l'écriture des projets de recherche.

Une seconde série de réponses se situe au niveau de la valorisation du partage des données dans la carrière des chercheurs et des enseignants-chercheurs par les institutions de gestion de la recherche publique en France. Comment assurer une reconnaissance des matériaux partagés, et donc aux chercheurs qui les ont collectés ? Cette reconnaissance pourrait passer par l'attribution d'un DOI (Digital Object Identifier, dont l'identifiant reste invariant au fil du temps) aux matériaux mis à disposition des autres chercheurs. On pourrait alors imaginer que la citation de ce DOI soit autant valorisée dans le CV des chercheurs que celle de leurs articles de recherche.

Une dernière piste serait de sensibiliser les chercheurs aux bénéfices qu'ils pourraient tirer de la réutilisation des données d'autres chercheurs, et donc à l'importance de mettre eux-mêmes à disposition leurs matériaux d'enquête. Cette sensibilisation pourrait s'opérer dans le cadre de recherches collectives, dans les projets des laboratoires de recherche, mais aussi plus en amont dans la formation des futurs chercheurs, en développant la réutilisation des données dans des cadres pédagogiques, comme par exemple dans les enseignements de master. C'est une voie qui est suivie actuellement par la banque d'enquêtes qualitatives beQuali, dont un des objectifs explicites est de favoriser la réutilisation des enquêtes dans le cadre des enseignements en sciences humaines et sociales ([http://bequali.fr/media/ckeditor/uploads/2016/07/12/fiche\\_bequali\\_enseignement.pdf](http://bequali.fr/media/ckeditor/uploads/2016/07/12/fiche_bequali_enseignement.pdf)).

Cette sensibilisation des chercheurs aux bénéfices de la mutualisation et de la réutilisation des données nous semble pouvoir être menée à une plus petite échelle dans le champ des chercheurs francophones travaillant sur la thématique des mobilités. Ce champ est en effet constitué d'un nombre « raisonnable » de chercheurs, issus de différentes disciplines. Il n'est pas traversé par de fortes oppositions théoriques, et de personnes, entre chercheurs. Il est également structuré en partie par le groupe de travail « mobilités spatiales et fluidité sociale » (MSFS) de l'Association internationale des sociologues de langue française (AISLF). Les chercheurs de ce champ ont ainsi l'habitude de se retrouver lors de colloques scientifiques, sont tenus au courant des travaux de recherche par le biais de la liste de diffusion de MSFS, et une partie d'entre eux a participé ensemble à des recherches collectives. Une

partie des membres de ce groupe de travail a par ailleurs participé à la rédaction d'une charte de la désexcellence, dont un des principes est de « considérer les fruits de recherches financées (en tout ou en partie) par les pouvoirs publics comme ceux de l'ensemble de la société » ([http://lac.ulb.ac.be/LAC/charte\\_files/Charte\\_Desexcellence\\_1-1.pdf](http://lac.ulb.ac.be/LAC/charte_files/Charte_Desexcellence_1-1.pdf)). Toutes ces caractéristiques du champ nous laissent penser à un possible intérêt des chercheurs travaillant sur la thématique des mobilités envers une réflexion collective sur la mise en commun de différents matériaux d'enquête, sous réserve d'un dispositif de mutualisation satisfaisant au niveau méthodologique et déontologique. Cet intérêt des chercheurs sera renforcé si notre méta-projet sur les mobilités périurbaines porte ses fruits au niveau de nouveaux résultats scientifiques permis par la mutualisation des données. En effet, un discours général de sensibilisation à la question de partage des données est d'autant plus efficace lorsqu'il s'appuie sur des projets pilotes démontrant les avantages scientifiques de ce partage (Serres et *al*, 2017b: 48). Les deux volets de notre projet, mutualisation et réanalyse des données, ne peuvent ainsi être séparés.

## Conclusion et ouverture

L'ouverture et la mise en partage des données de la recherche, en particulier des données qualitatives, posent des questions épistémologique, éthique et juridique. Depuis une vingtaine d'années, une littérature relativement abondante, en particulier anglophone, s'est attachée à répertorier et traiter ces enjeux : la lecture de cette littérature fait ressortir quelques éléments marquants, que nous avons tenté de présenter dans le présent livrable. On retiendra d'une part les tensions, voire les incompatibilités, qui peuvent exister au sujet de la mutualisation des données, entre, d'une part, les intérêts scientifique, méthodologique ou financier du partage et, d'autre part, le respect du contrat de confiance et des conditions d'utilisation définies lors de la collecte des données. Cette littérature montre ainsi par exemple l'impossibilité de satisfaire l'impératif d'anonymisation, tout en maximisant les possibilités de ré-analyse. Elle souligne également la contradiction entre ce même respect de l'anonymat des enquêtés et les objectifs « d'information » mentionnés dans le règlement européen de 2016.

Les chercheurs, qu'ils soient producteurs ou réutilisateurs des données, essaient de faire avec ces tensions : malgré l'absence de charte déontologique traitant spécifiquement des enjeux liés à l'ouverture et au partage des données de la recherche, ils doivent réussir à maintenir un équilibre entre leurs valeurs éthiques, la déontologie de leur profession, et les impératifs juridiques qui s'imposent à eux. C'est à eux d'arbitrer sur ce qu'ils peuvent/doivent/se donnent le droit de faire avec leurs données ; à eux de décider si, et selon quelles modalités, ils souhaitent les partager. A ces tensions vient s'ajouter le principe de réalité auquel sont confrontés les chercheurs, en termes de contraintes temporelles ou de manque de moyens humains et infrastructurels facilitant le travail de stockage et d'archivage des données ; s'ajoutent également des choix à opérer entre des intérêts parfois incompatibles (intérêt scientifique *vs* non reconnaissance institutionnelle du partage par exemple).

Tout en reconnaissant l'intérêt de mettre en partage les données de la recherche, il faut donc reconnaître que la concrétisation de ces objectifs (soutenus par les principales structures du monde de la recherche et partagés par de nombreux chercheurs au vu de diverses enquêtes menées dans des universités françaises) n'est pas simple. Comment alors favoriser le partage des données pour qu'il soit scientifiquement pertinent, tout en étant réalisé dans des conditions éthique et juridique acceptables ?

A l'instar de Qualidata, BeQuali ou Quételet, structures de partage de données de la recherche, nous prôtons une position résolument restrictive. Restrictive en termes de données partagées, en proposant d'imposer des restrictions d'accès pour certaines données plus sensibles. Restrictive en termes de mise à disposition, en proposant de limiter le partage aux seuls demandeurs proposant un projet scientifique sérieux, disposant de compétences leur permettant d'analyser sérieusement les données, et dont on pourra être sûr qu'ils sont soumis à un « contrôle par les pairs » et ne sont pas motivés par des intérêts purement économiques. Cette position restrictive est selon nous essentielle pour garantir le respect des enquêtés et l'intérêt scientifique du partage.

En outre, favoriser le partage des données nécessite selon nous un véritable soutien institutionnel, comme le montre l'exemple britannique : d'une part en termes de moyens humains, d'infrastructures techniques et donc évidemment de soutien financier ; d'autre part en termes de politiques générales d'administration de la recherche. En effet, pour que le partage de données se développe, il faut avant tout qu'une culture du partage et de l'analyse secondaire émerge et se propage parmi les chercheurs et les étudiants, ce qui suppose des politiques de promotion volontaristes à l'échelle des universités et des

laboratoires. On peut espérer que l'institution des nouveaux « délégués à la protection des données » au niveau des laboratoires participe du développement de ces politiques.

Au final, ce livrable a posé un certain nombre de jalons visant à définir les conditions générales de la mise en partage des données de la recherche ; il a également défini des conditions de mutualisation spécifiques au méta-projet, selon lesquelles nous pensons pouvoir nous affranchir des contraintes liées au consentement des enquêtés à voir leurs entretiens ré-exploités. Dans de telles conditions, il s'agit désormais de définir les conditions méthodologiques dans lesquelles nous pourrions mettre en œuvre une analyse secondaire des matériaux à notre disposition.

## Bibliographie

- Abul O, Bonchi F. & Nanni M., 2008. “*Never Walk Alone: Uncertainty for Anonymity in Moving Objects Databases*” in Proceedings of the 2008 IEEE 24th International Conference on Data Engineering. IEEE Computer Society: 376–385.
- Andrés M. E., Bordenabe N. E., Chatzikokolakis K., & Palamidessi C., 2013. “*Geo-indistinguishability: Differential Privacy for Location-based Systems*” in Proceedings of the 2013 ACM SIGSAC Conference on Computer & Communications Security. ACM: 901–914.
- Barel Y., 1984. *La société du vide*, Seuil, Paris
- Béliard A. & Eideliman J.S., 2008. Au-delà de la déontologie. Anonymat et confidentialité dans le travail ethnographique. *La Découverte. Politiques de l'enquête. Epreuves ethnographiques*: 123-141
- Berger P.L. & Luckmann T., 1966, trad. Fr. 1986. *La construction sociale de la réalité*, Ed. Armand Colin, Coll. Références
- Bertaux D., 2010. *Le récit de vie*. 128p. Paris: Colin.
- Bishop L., 2013. “Reply to: Natasha Mauthner and Odette Parry’s “Open access digital data sharing: principles, policies and practices””, *Social Epistemology review and reply collective* 2 (8): 71-78
- Bishop L., 2014. “Re-using qualitative data: A little evidence, on-going issues and modest reflections”. *Studia Sociologica*, 3(214): 167-176.
- Bishop L. & Kuula-Luumi A., 2017. “Revisiting qualitative data reuse: a decade on”, *SAGE Open* Special Issue: 1-15
- Chaudiron S., Maignant C., Schöpfel J., Westeel I., 2015. Les données de la recherche dans les thèses de doctorat - Livre blanc, Rapport de recherche, Université Lille 3, 15p. [Disponible sur <http://hal.univ-lille3.fr/hal-01192930>]
- CIRAD & INRA, 2016. *Avis 8 sur les enjeux éthiques et déontologiques du partage et de la gestion des données issues de la recherche*. [Disponible sur : <http://institut.inra.fr/Missions/Promouvoir-ethique-et-deontologie/Avis-du-comite-d-ethique/Avis-8-Partage-et-gestion-des-donnees-issues-de-la-recherche#>].
- COMETS, 2015. *Les enjeux éthiques du partage des données scientifiques*. [Disponible sur : [https://www.google.fr/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&ved=0ahUKEwidmdjwqJnTAhVDAcAKHY\\_xC8YQFgghMAA&url=http%3A%2F%2Fwww.cnrs.fr%2Fcomets%2Fspip.php%3Farticle123&usq=AFQjCNFj6-N6XoViURqR7my1svYrjU3UeA&sig2=Z\\_NF44xliGrV\\_mn3zehUnw](https://www.google.fr/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&ved=0ahUKEwidmdjwqJnTAhVDAcAKHY_xC8YQFgghMAA&url=http%3A%2F%2Fwww.cnrs.fr%2Fcomets%2Fspip.php%3Farticle123&usq=AFQjCNFj6-N6XoViURqR7my1svYrjU3UeA&sig2=Z_NF44xliGrV_mn3zehUnw)].
- Corti L., 2000. “Progress and problems of preserving and providing access to qualitative data for social research - The international picture of an emerging culture”, *Forum: Qualitative Social Research*, 1(3), Art. 2
- Corti L., 2011. “The european landscape of qualitative social research archives: methodological and practical issues”, *Forum : Qualitative social research* 12 (3) Art.11
- Corti L. & Bishop L., 2005. “Strategies in teaching secondary analysis of qualitative data”, *Forum: Qualitative Social Research*, 6 (1) Art. 47
- Coulmont B., 2017. « Le petit peuple des sociologues. Anonymes et pseudonymes dans la sociologie française », *Genèses*, 2017/2 (107) : 153-175
- Cribier F. & Feller E., 2003. *Projet de conservation des données qualitatives des sciences sociales recueillies en France auprès de la « société civile*. CNRS-EHESS, LASMAS, UMR.
- Dargentas M., Le Roux D., Salomon A.C. & Brugidou M., 2007. « Sur les perspectives de la recherche qualitative en France : capitalisation et réutilisation d’entretiens de recherche », *Recherches qualitatives* Hors Série (3)

- Duchesne S. & Garcia G. (2014) “beQuali : une archive qualitative au service des sciences sociales”, in Cornu M., Fromageau J. & Muller B., *Archives de la recherche. Problèmes et enjeux de la construction du savoir scientifique*. L’Harmattan, p. 35–56.
- Duchesne S., 2017. « De l’analyse secondaire à la réanalyse. Une innovation méthodologique en débats », *Recherches qualitatives Hors-Série* (21) : 7-28
- Ferrarotti F., 1983. *Histoire et histoires de vie, la méthode biographique dans les sciences sociales*, Librairie des Méridiens
- Fielding, N., 2004. “Getting the most from archived qualitative data: Epistemological, practical and professional obstacles”, *International Journal of Social Research Methodology*, 7(1) : 97-104.
- Hugrée C. & Kern A.L., 2008. « Observer des téléenquêteurs, Les paradoxes de la rationalisation de la production statistique », *Genèses*, 3 (72) : 102-118
- Kaufmann V., 2004. « La mobilité comme capital ? » in Montulet B. & Kaufmann V. (dirs.). *Mobilités, fluidités... libertés ?*, p. 25-41. Bruxelles : Presses de l’Université Saint-Louis.
- Le Roux D., 2005. *Verbatim, une expérience de capitalisation dans le monde de l’entreprise*, Colloque « l’analyse secondaire en recherche qualitative », Grenoble
- Lévy J., 2001. Préface, in Pinson D. & Thomann S., *La maison en ses territoires*, L’Harmattan
- Mauthner N. S. & Parry O., 2009. “Qualitative data preservation and sharing in the social sciences: on whose philosophical terms?”, *Australian Journal of Social Issues*, 44(3) : 291-307
- Moravcsik A., 2010. “Active Citation: A Precondition for Replicable Qualitative Research”, *PS: Political Science & Politics* 43(1) : 29-35.
- Prost H. & Schöpfel J., 2015. Les données de la recherche en SHS. Une enquête à l’Université de Lille 3 : Rapport final, Rapport de recherche, Université Lille 3, 37p. [Disponible sur <http://hal.univ-lille3.fr/hal-01198379>]
- Rege A. (2015). *Retour sur l’enquête sur les pratiques de publication scientifique et de production de données de la recherche*, Université de Strasbourg, Projet AOC, COPIL 27 mars 2015.
- Serres A., Malingre M.L., Mignon M., Pierre C., Collet D., 2017. *Donnée de la recherche en SHS, Pratiques, représentations et attentes des chercheurs : une enquête à l’Université Rennes 2*, Synthèse des résultats, 23p. [Disponible sur <https://hal.archives-ouvertes.fr/hal-01635186>]
- Serres A., Malingre M.L., Mignon M., Pierre C., Collet D., 2017b. *Donnée de la recherche en SHS, Pratiques, représentations et attentes des chercheurs : une enquête à l’Université Rennes 2*, 161p.
- Soubiran T., 2017. *La règlementation relative aux données à caractère personnel en sciences sociales*, Journées Data-SHS, 11-12 décembre 2017