



HAL
open science

Ouverture de données qualitatives à caractère personnel Approche éthique, juridique et déontologique

Marie Huyghe, Laurent Cailly, Nicolas Oppenheim

► To cite this version:

Marie Huyghe, Laurent Cailly, Nicolas Oppenheim. Ouverture de données qualitatives à caractère personnel Approche éthique, juridique et déontologique. Véronique Ginouvès; Isabelle Gras. La diffusion numérique des données en SHS Guide des bonnes pratiques éthiques et juridiques, Presses universitaires de Provence, 2018. halshs-02015218

HAL Id: halshs-02015218

<https://shs.hal.science/halshs-02015218>

Submitted on 14 Mar 2019

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.

Ouverture de données qualitatives à caractère personnel

Approche éthique, juridique et déontologique

Marie Huyghe
CITERES (UMR 7324, CNRS)

Laurent Cailly
Université de Tours

Nicolas Oppenheim
Université de Tours

Abstract: *Data sharing is often presented as an ideal in terms of the transparency of the scientific process or the optimization of public-funded data. The main French research institutions (notably INRA, CNRS and CIRAD) therefore encourage their researchers to engage in data sharing movement. In such a context, personal data management becomes of particular importance: it raises ethical, legal and epistemological issues. In this article we are looking at the specific case of a posteriori data sharing, when mutualization had not been presented as an objective while data was being collected. We examine three qualitative data sharing scenarii: in the first one, data are shared between members of the same research program; in the second one, data are more broadly shared, within the scientific community; in the third one, data are being made available in full and open access.*

Cet article propose une réflexion sur la mutualisation et l'ouverture des données qualitatives, menée dans le cadre du programme de recherche « Mutualisation et méta-analyse de corpus qualitatifs sur les mobilités quotidiennes dans les territoires dépendants de l'automobile¹ ». Ce « méta-projet » se donne pour objectif de réinvestir six opérations de recherche portant sur les mobilités et réalisées au sein du laboratoire CITERES (UMR 7324, université François Rabelais de Tours), à partir de la confrontation et du réexamen de l'ensemble des matériaux produits, soit un corpus de plus de 250 entretiens.

Nous faisons l'hypothèse que cette question de la diffusion, de la mutualisation et du retraitement des données qualitatives concerne aujourd'hui un nombre de plus en plus important de chercheurs, en raison des injonctions croissantes au partage des données de la recherche. De fait, l'ouverture et le partage des données (*open data* et *data sharing*) sont présentés comme un idéal à atteindre en termes

1 Ce programme de recherche est mené actuellement à l'UMR 7324 CNRS-CITERES de Tours et financé et suivi scientifiquement par le Forum vies mobiles <http://forumviesmobiles.org>.

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 et en Allemagne.

Dans un tel contexte, la gestion des données à caractère personnel se pose avec une acuité particulière. Elle pose notamment des questions éthique, déontologique et juridique, que nous proposons d'investiguer dans cet article. Plus globalement, nous interrogeons ici la mutualisation et la réutilisation des données *a posteriori*, dans le cas où (comme c'est le cas dans le « méta-projet ») elle n'aurait pas été anticipée dans les différents programmes de recherche initiaux ; il nous semble que ce cas de figure peut concerner de nombreux projets dans la phase actuelle de transition vers l'*open data*.

Nous souhaitons tester trois scénarios d'ouverture des données : dans un cercle restreint entre collègues travaillant sur un même programme de recherche (scénario 1) ; dans un cercle plus large, au sein de la communauté scientifique (scénario 2, dans lequel nous nous plaçons dans le cadre du « méta-projet ») ; enfin, en libre accès (scénario 3). Pour chaque scénario, nous nous interrogerons sur la façon dont il est possible de diffuser les données tout en restant dans un cadre juridique et éthique satisfaisant².

Suivant la loi rien n'est possible... ou presque...

Les données travaillées pendant le « méta-projet » sont issues de 250 entretiens semi-directifs. Ces données sont protégées par un droit d'auteur qui s'applique à la fois à l'enquêteur et à l'enquêté. En outre, une partie des données collectées sont considérées comme des données « à caractère personnel », puisqu'elles permettent une identification directe ou indirecte des enquêtés (en croisant des informations telles que le lieu de domicile, l'emploi, l'âge et le nombre d'enfants par exemple) ; ces informations sont donc soumises à la loi Informatique et libertés de 1978 et au règlement européen sur la protection des données (RGDP).

En conséquence, des règles s'appliquent au moment de la collecte et du traitement des données (retranscription, diffusion et mutualisation notamment) : elles imposent d'informer les enquêtés des finalités de l'enquête et d'obtenir leur consentement pour le traitement qui en sera fait. Ce consentement doit être éclairé : l'enquêteur doit notamment préciser auprès de *qui*, pendant *combien de temps* et *via quel média* les données seront accessibles. Enfin, pour être juridiquement valable, il doit rester une trace du consentement donné par l'enquêteur.

Or, nous devons reconnaître que dans bien des recherches qualitatives un consentement écrit n'est pas collecté. C'est le cas dans le corpus que nous mobilisons : si un « contrat de confiance » liant l'enquêteur et le(s) enquêté(s) a bien été passé oralement pour les 250 entretiens, très peu ont été enregistrés ou

2 Les réflexions présentées dans cet article ont été développées au cours du programme de recherche, et donné lieu à un rapport de recherche (Huyghe *et al.*, 2018).



signés ; par ailleurs, n'ayant pas anticipé la question de la mutualisation ultérieure des données, celle-ci n'a jamais ou rarement été évoquée lors de la passation du contrat de confiance.

En outre, l'ouverture des données de la recherche est de plus en plus posée comme une injonction, mais aussi comme un horizon souhaitable, puisqu'elle permet de mieux exploiter, voire d'optimiser, les matériaux de la recherche ; or, jusqu'à récemment, la question de la réexploitation ou de la mutualisation des données n'était que rarement anticipée dans les projets. Désormais, certains laboratoires encouragent les chercheurs à aborder cette question, grâce notamment à des modules de formation (IIST 2014).

Dans ces conditions, qu'est-il possible de faire avec des données protégées, en l'absence d'un consentement écrit donnant l'autorisation de réutiliser les données ? Quelles procédures doit-on mettre en œuvre pour rester dans un cadre juridique et éthique acceptable ?

Obtenir le consentement des enquêtés

Pour traiter les données à notre disposition tout en restant dans un cadre légal, il n'y a d'autres possibilités que d'obtenir l'accord des enquêtés : pour un traitement « simple » des données (leur retranscription par exemple, qui dans les faits a déjà été réalisée), mais également pour leur mutualisation et leur diffusion auprès d'autres personnes.

Dans un premier temps, dans l'hypothèse où nous parviendrions à recontacter les enquêtés sans difficulté, se pose la question des conditions concrètes de la passation de ce nouveau contrat de confiance. L'objectif, dans notre situation, est *a minima* de solliciter l'accord des enquêtés pour mutualiser les entretiens au sein de l'équipe de chercheurs du « méta-projet » et plus largement au sein de la sphère universitaire (scénario 2). Plusieurs questions se posent alors :

Que présenter aux enquêtés ? La question de la diffusion et de la mutualisation des données peut être compliquée à comprendre, en particulier leurs objectifs, intérêt et enjeux pour l'enquêté. Nous proposons de présenter ces points, ainsi que les détails des trois scénarios de diffusion, sur un formulaire qui sera transmis aux enquêtés et où ils devront choisir à quelle modalité de diffusion ils consentent.

Qui doit se charger de contacter les enquêtés ? Le plus facile serait que ce soit l'enquêteur initial ; néanmoins, comment faire lorsque les entretiens ont été réalisés par des étudiants ou par des chercheurs qui ne font pas partie de l'équipe de recherche ?

Enfin, par quel moyen ? Par téléphone, en face à face après prise de rendez-vous, par courrier ? Il faut peser les risques : plusieurs années après la passation des entretiens, il est possible que les enquêtés refusent d'accorder de nouveau du temps pour recevoir l'enquêteur ; *a contrario*, un appel téléphonique permet de gagner du temps, mais n'est pas forcément suffisant pour amener les enquêtés à accorder leur consentement pour la mutualisation des données. Un premier appel explicatif, suivi de l'envoi du formulaire à signer par courrier, pourrait convenir.

La CNIL indique une exception à l'obligation d'obtenir le consentement des enquêtés, lorsque « la démarche est manifestement impossible ou implique la

mobilisation de moyens disproportionnés par rapport à l'intérêt qu'elle présente³ ». D'après E. Masson (2017), cette exception ne serait cependant sans doute pas applicable dans le cas de nos données : *a priori*, il ne serait pas jugé « impossible » de recontacter une centaine de personnes dont nous connaissons (pour la plupart) les coordonnées.

Néanmoins, cette disposition ouvre une « zone grise » : peut-on traiter les données si l'on prouve qu'il a été tenté de recontacter les enquêtés mais sans y parvenir (lorsque les personnes ont déménagé, changé de numéro de téléphone ou d'adresse mail) ou lorsque nous n'avons pas obtenu de réponse ? Ce cas de figure nous amène à examiner différentes conditions de circulation des données.

Déontologie de la circulation des données

Anonymiser les entretiens

Le respect de la confidentialité des données personnelles et de l'anonymat font partie des règles déontologiques élémentaires de l'entretien semi-directif. Lorsque le chercheur stocke ou exploite des entretiens, il s'engage auprès de la personne interrogée à ce que les informations qui y sont contenues ne soient pas divulguées. Lorsque, pour le besoin de la démonstration scientifique, il cite des extraits d'entretien dans une publication, la règle veut (ou voudrait) qu'un travail rigoureux d'anonymisation soit effectué.

Dans le projet de mutualisation que nous poursuivons, l'anonymisation relative des entretiens pourrait constituer une condition suffisante pour assurer la circulation des matériaux entre chercheurs dans le cas où l'accord de l'enquêté n'a pas pu être obtenu. C'est cette solution qui a également été retenue par les chercheurs de beQuali, principale banque de données qualitatives en sciences humaines et sociales (Duchesne et Garcia 2014).

Cet objectif achoppe toutefois sur les difficultés à anonymiser un entretien dans sa globalité. En effet, si l'on s'en tient à une définition *stricte* de l'anonymat⁴, l'anonymisation est non seulement impossible à mettre en œuvre – puisqu'il faudrait « brouiller » la totalité des données personnelles qui constituent le corps même de l'entretien –, mais 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 une définition stricte n'est donc ni envisageable ni souhaitable.

Si toutefois l'on considère, dans une conception plus souple de l'anonymisation, qu'elle obéit à un gradient, deux techniques pourraient être mises en œuvre afin

3 Loi n° 2004-801 du 6 août 2004, article 5, relative à la protection des personnes physiques à l'égard des traitements de données à caractère personnel et modifiant la loi n° 78-17 du 6 janvier 1978 relative à l'informatique, aux fichiers et aux libertés.

4 Groupe des 29 / CNIL : « Pour rendre des données anonymes, il faut en retirer suffisamment d'éléments pour que la personne concernée ne puisse plus être identifiée. Plus précisément, les données doivent être traitées de façon à ne plus pouvoir être utilisées pour identifier une personne physique en recourant à "l'ensemble des moyens susceptibles d'être raisonnablement mis en œuvre", soit par le responsable du traitement, soit par un tiers. Un facteur important est que le traitement doit être irréversible ».



de préserver – de manière relative – l’anonymat. La première est une technique assez simple et banale de pseudonymisation que l’on peut réaliser en modifiant les prénoms ou en ne gardant que les initiales. Néanmoins, cette technique n’empêche en rien de reconnaître les individus en croisant d’autres données individuelles (domicile, emploi, âge, nombre d’enfants).

La seconde, issue des techniques de « randomisation » (altération des données individuelles), consiste à introduire sur certaines données clés « un brouillage » : taire le nom de l’employeur et le métier en ne retenant que la PCS, donner une fourchette d’âge plutôt que l’âge, changer le nom de la commune de résidence, etc. Ce brouillage des données pose néanmoins un problème considérable d’analyse pour les chercheurs travaillant sur la dimension spatiale des pratiques puisqu’un entretien portant sur les déplacements et les lieux de vie, dont on gommerait les localisations ou bien les toponymes, perdrait toute sa valeur. Le brouillage ne peut donc être réalisé que très à la marge (remplacer une adresse par un nom de quartier) ou porter sur des données jugées « secondaires ».

Dans les deux cas, si ces techniques abaissent le risque d’identification personnelle de la personne interrogée, elles n’excluent en aucun cas une possible identification par un tiers qui pourrait reconnaître un individu par recoupement d’informations. Elles n’apportent donc que des garanties relatives. Cette anonymisation partielle ne peut donc constituer une réponse à elle seule en cas d’absence de consentement ; néanmoins, celle-ci nous semble devoir participer d’une déontologie de la circulation des données qualitatives.

Contrat de circulation des données

En effet, au nom de la protection des données personnelles, l’objectif de mutualisation – et donc de circulation – des matériaux qualitatifs impose de définir des règles d’usages visant à garantir la continuité d’une éthique de la confidentialité. Cette dernière doit passer par l’anonymisation partielle des données par le chercheur qui a opéré la collecte et qui en contrôle la diffusion. Mais, plus globalement, la circulation du matériau devrait être accompagnée d’un « contrat de circulation des données ». Cette contractualisation liant le chercheur-producteur des données et le chercheur-utilisateur semble être une pratique largement partagée dans les différentes expériences de mutualisation-réutilisation : beQuali a par exemple élaboré un « contrat de réutilisation », signé par la banque de données et l’utilisateur⁵. C’est également une pratique adoptée dans les recherches utilisant des bases de données quantitatives : A. Decuyper (2017) évoque les « *non-disclosure (or confidentiality) agreements* » qui sont passés avant tout partage de données.

Ce contrat – qui pourrait se présenter sous la forme de conditions d’utilisation engageant moralement l’utilisateur – rappellerait aux futurs utilisateurs les principes essentiels de confidentialité : non-transmission à un tiers, sécurisation du stockage, confidentialité dans le traitement et l’analyse des entretiens (réunions de travail), anonymisation renforcée lors de citation, etc. Il aurait finalement pour vocation d’assurer la transmission et l’extension du contrat de confiance très personnel qui s’établit initialement entre l’enquêté et l’enquêteur, dont on sait par ailleurs qu’il est

5 Ce contrat est disponible dans le chapitre consacré à l’expérience beQuali.

précisément une condition de réalisation d'un entretien⁶. Il viserait par là même à prévenir les risques engendrés par l'absence de lien direct entre l'enquêté et l'utilisateur de seconde main qui, sans mauvaise intention, pourrait être moins vigilant en matière de confidentialité.

Si ce contrat de confidentialité doit être associé à toutes les opérations de circulation du matériau, il pourrait jouer un rôle plus important encore dans le cas où l'enquêté n'a pas pu être recontacté et le consentement obtenu. En effet, dans cette situation de « zone grise », en application de l'exceptionnalité définie par la CNIL, ce contrat de confidentialité pourrait devenir un dispositif à même de garantir le « droit moral » de l'enquêté (confidentialité, anonymat, non-divulgateion) et autoriserait « moralement » l'enquêteur à transmettre ce matériau à un autre chercheur – sur demande – au titre de l'intérêt scientifique.

Pour être validée, cette procédure inédite et située dans une zone grise ouverte par le droit devrait toutefois être soumise à des juristes. Dans tous les cas, ces conditions éthiques et déontologiques doivent être examinées dans le cadre des trois scénarios d'ouverture des données que nous souhaitons tester dans cet article.

Trois scénarios d'ouverture des données

Enjeux éthiques des différents scénarios

Les enjeux éthiques de la transmission des données se posent différemment suivant que l'ouverture des données s'opère au sein d'un groupe de chercheurs travaillant sur un même programme de recherche (scénario 1), concerne d'autres chercheurs (scénario 2) ou est envisagée en *open access* (scénario 3).

La première situation est très banale. Le chercheur qui a collecté le matériau partage celui-ci avec ses collègues impliqués dans un même programme. L'interconnaissance et la confiance qu'impose le travail collectif permettent d'élargir tacitement le contrat de confidentialité à l'ensemble des chercheurs, sans qu'une protection en interne des données individuelles soit nécessaire. La garantie sur la confidentialité des données adopte implicitement le périmètre du collectif.

Cette garantie (qui relève d'une confiance interne quant au respect des règles de confidentialité) est pourtant déterminante, car souvent les recrutements des enquêtés sont locaux et s'effectuent de proche en proche. Les possibilités d'identification précise des enquêtés par une autre personne que l'enquêteur sont donc fréquentes. Dans ce contexte précis, la question de la protection de l'intimité n'est alors pas complètement mutualisée et il revient, à chaque chercheur, suivant une éthique qui lui est propre et la relation qu'il entretient avec l'enquêté, de savoir gommer – au besoin – certaines données. C'est ainsi qu'au sein du programme Marg'in⁷ sur les migrations roms, une chercheuse a pris la responsabilité de brouiller certaines informations sur le parcours d'un individu poursuivi par les autorités.

6 En situation d'entretien, en effet, la parole de l'enquêté est rendue possible par la confiance de l'enquêté dans l'usage qu'en fera le chercheur.

7 ANR, programme Défi 8 « Sociétés innovantes, intégrantes et adaptatives », 2015-2017 – Coordinateur principal : O. Legros (UMR CITERES, université de Tours).



Dans le scénario 2, défini par l'ouverture des données à d'autres chercheurs, l'interconnaissance et l'engagement des chercheurs au sein d'un projet collectif ne peuvent plus garantir la confiance dans le respect des règles. Nous considérons que le « méta-projet » sur lequel nous travaillons nous place dans ce scénario, quand bien même la mutualisation de données issues de différents programmes de recherche se fait entre chercheurs du même laboratoire.

Nous pourrions faire l'hypothèse que le titre académique et la posture éthique que devrait partager la communauté des chercheurs constituent, en soi, une garantie.

Cette extension du contrat de confiance ne nous semble pas aller de soi. Comme énoncé ci-dessus, nous croyons à l'inverse que l'éloignement de l'opération de collecte est propice à l'affranchissement de certains garde-fous. La mise en œuvre du contrat de confidentialité présenté précédemment et rédigé par le chercheur primaire, qui rappelle et réengage les autres chercheurs sur une éthique (des règles) d'utilisation des entretiens, nous semble être le dispositif le mieux à même de garantir et de faire circuler ce contrat de confiance. Pour l'enquêteur qui en est à la source, cette charte de bonne utilisation qu'il soumet aux futurs utilisateurs peut alors apparaître comme l'expression et la réaffirmation de l'engagement moral qu'il a pris vis-à-vis de son enquêté. Dans le cadre du scénario 2, des modalités de diffusion des données devraient être définies, visant à garantir un certain contrôle aux chercheurs-auteurs.

Concernant l'accès aux données, nous prenons exemple sur le réseau Quetelet, qui permet un « accès gratuit à des fins de recherche » à condition de justifier de leur utilité pour un projet de recherche⁸. Dans le cas où cette justification ne convaincrat pas les chercheurs, ceux-ci restent évidemment libres de refuser la diffusion de leurs données. Dans un certain nombre de recherches, une autre solution est mise en œuvre pour contrôler l'accès aux données, qui consiste à rendre accessibles les données uniquement sur les ordinateurs de l'institution des chercheurs ayant mené la recherche ; cette solution limite toutefois l'accès aux données et nous semble relativement contradictoire avec un objectif de diffusion (même contrôlée) des données de la recherche.

La diffusion des données, même au sein de la sphère universitaire, implique une perte de contrôle de leur devenir : en particulier, on ne peut s'assurer que la personne à qui les données auront été confiées ne va pas à son tour les diffuser. Il est dès lors impossible pour le chercheur-auteur de garantir la continuité du contrat de confidentialité. Au mieux, on peut comme le réseau Quetelet engager les demandeurs à « ne pas céder les fichiers à un tiers ». Cette perte de contrôle est à notre sens un risque à prendre, dont on ne peut se prémunir si l'on accepte le principe de diffusion des données ; une nouvelle fois, on ne peut que compter sur l'éthique des membres de la sphère universitaire.

Enfin, une dernière modalité devrait être définie : les chercheurs conservant leurs droits d'auteurs sur les données même après leur diffusion, ils doivent être informés de leur réutilisation, et cités en cas de publication d'articles. Si les données de seconde main constituent le principal matériau de l'article, nous considérons même que le chercheur ayant produit les données devrait être considéré comme coauteur.

8 http://www.reseau-quetelet.cnrs.fr/spip/article.php3?id_article=154.

Dans le troisième scénario, celui du partage en libre accès, le respect du principe de confidentialité est beaucoup plus problématique : en diffusant des enquêtes en libre accès en dehors de la communauté scientifique, plus aucun garde-fou ne peut protéger les données personnelles. Les opérations de pseudonymisation et de brouillage de certaines données ne constituent pas en soi des remparts suffisants ; ils ne peuvent exclure qu'un utilisateur retrouve l'identité de l'enquêté, voire s'emploie à restaurer et à diffuser une version désanonymée d'un entretien.

L'impossibilité de garantir la confidentialité et les risques afférents nous semblent compromettre sérieusement la possibilité de soutenir d'un point de vue éthique ce scénario.

Que faire des données « à risque » ?

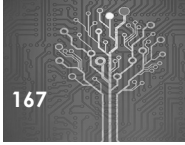
De fait, les questions éthiques concernent en particulier les données « à risque⁹ », qui pourraient nuire aux enquêtés en cas de diffusion des entretiens. Par exemple, des données qui pourraient être utilisées contre l'enquêté à des fins judiciaires, ou des données sur la situation médicale des enquêtés qui pourraient être utilisées par des sociétés d'assurance.

Pour limiter ces risques, nous préconisons de ne pas diffuser en libre accès (dans le cadre du scénario 3) les données à caractère personnel ; *a fortiori*, de ne pas diffuser en libre accès les retranscriptions ou les enregistrements des entretiens passés avec les enquêtés.

Par ce positionnement, nous rejoignons les préconisations émises par le CNRS (COMETS 2015), qui indique que « le libre accès aux données scientifiques fait désormais partie des objectifs de la recherche publique », exclusion faite des données « communiquées aux chercheurs sous réserve de confidentialité ». Nous considérons que les données à caractère personnel que nous analysons, en particulier les données les plus sensibles, constituent des données « confidentielles » qu'il n'est dès lors pas souhaitable de diffuser : nous pensons donc que les entretiens semi-directifs, quand bien même ils constituent la majeure partie des données analysées en sciences humaines et sociales et représentent une base de données extrêmement riche, ne devraient pas être diffusés en *open access*. En revanche, ces restrictions ne s'appliquent pas aux métadonnées relatives aux enquêtés (sexe, tranche d'âge, catégorie socio-professionnelle, etc.), qui nous semblent pouvoir être diffusées sans risque.

Par ailleurs, en cas de diffusion des entretiens dans des cercles restreints, dans le cadre des scénarios 1 ou 2, on ne peut exclure que certaines données soient récupérées par des services policiers ou judiciaires et utilisées contre les enquêtés. Dès qu'il y a diffusion, même limitée, des entretiens, il nous semble donc que les données les plus sensibles devraient impérativement être supprimées par le chercheur ayant recueilli les discours, puisqu'il n'est pas du ressort du travail du chercheur de « servir » la machine judiciaire ou quelque structure à fin commerciale.

9 La CNIL définit quant à elle les données « sensibles » ou « à risque » comme « toute information concernant l'origine raciale ou ethnique, les opinions politiques, philosophiques ou religieuses, l'appartenance syndicale, la santé ou la vie sexuelle » (cnil.fr).



Comment faire respecter le sens du discours des enquêtés ?

Enfin, au-delà de cette question de la confidentialité, l'archivage et la mutualisation des données posent à nouveau, de manière amplifiée, une question déontologique centrale en sciences humaines, celle du respect du sens du discours des enquêtés. Si la situation d'entretien est toujours une co-construction de sens entre l'enquêteur et l'enquêté, l'analyse et la mise en forme du matériau récolté sont le plus souvent assurées par le chercheur seul, même si certains d'entre eux assurent une restitution de leurs résultats aux enquêtés afin que ceux-ci puissent contrôler l'interprétation de leurs discours et pratiques (Bizeul 2008). L'assurance de voir cette règle respectée est très différente selon les scénarios envisagés, et dépend notamment des risques liés à la décontextualisation des données.

Dans le premier scénario, la présence du chercheur qui a collecté le matériau lui permet de discuter directement avec ses collègues de l'analyse qu'ils font de celui-ci et de préciser le contexte de certains discours. Dans le second scénario, il s'agit pour le chercheur qui a produit les données de porter à la connaissance de la sphère universitaire toutes les métadonnées (relatives aux enquêtés et aux situations d'enquête) qui permettront de comprendre au mieux le matériau. Les chercheurs de beQuali, considérant que pour comprendre les données on ne peut les séparer de leur contexte de production (Duchesne 2010), réalisent et donnent accès à de véritables « enquêtes sur l'enquête » : celles-ci en expliquent « la genèse, l'univers théorique, la réalisation du terrain, le corpus constitué et conservé, l'analyse des matériaux, la postérité et les potentialités de l'enquête » (Duchesne et Garcia 2014). Bien qu'utiles, voire indispensables pour une bonne réappropriation des données, ces « enquêtes sur l'enquête » sont sûrement chronophages et compliquent la diffusion des données. Nous choisissons donc pour le « méta-projet » de limiter le nombre et la diversité des métadonnées qui seront renseignées : elles regrouperont notamment une présentation du contexte de production des données (objectifs du programme de recherche, méthodologie employée, principaux résultats) et une présentation de chaque entretien (principales données de présentation de l'enquêté, grille d'entretien, contexte de passation de l'entretien, ambiance de l'entretien), voire des notes personnelles du chercheur.

Apporter ces éléments de contexte n'empêche pas des désaccords d'interprétation entre le chercheur-source et les chercheurs extérieurs. Elles peuvent notamment donner lieu à des controverses, « qui ne seront pas toujours à l'avantage des chercheurs primaires » (Duchesne et Garcia 2014 : 7) ; on peut néanmoins considérer que ces controverses sont des processus positifs de production collective des connaissances, menant bien souvent à des avancées scientifiques (Latour 1989). Aucune contractualisation ne pourra garantir au chercheur un droit de réponse à l'interprétation de ses matériaux ; c'est pourquoi nous encourageons, comme dans le scénario 1, les discussions entre les différents chercheurs utilisant les données.

Enfin, dans le scénario de l'*open access*, le chercheur ne dispose d'aucun suivi sur l'interprétation, scientifique ou non, de ses matériaux. Il ne peut alors en aucune manière s'assurer du respect du sens du discours des enquêtés ; il ne peut en particulier pas empêcher que celui-ci soit « manipulé » pour aller dans le sens de la démonstration d'un auteur.

Conclusion : Mise(s) en œuvre

Le lancement d'un « méta-projet » impliquant la mutualisation de données à caractère personnel, dans un contexte global d'injonctions à l'*open data* et au *data sharing*, nous a conduits à nous interroger sur la question de la diffusion et de la réutilisation des données de la recherche. Plus spécifiquement, nous nous sommes intéressés à un contexte où cette question n'aurait pas été anticipée, en l'absence de plan de gestion des données. À la condition qu'une déontologie de la circulation des données soit respectée, il nous semble possible de diffuser et réutiliser les données : cela implique un travail important de resollicitation des enquêtés, d'anonymisation des entretiens et de respect d'un contrat de confidentialité qui assure au maximum le respect des données personnelles. Concrètement, cela implique d'accompagner la circulation des matériaux, d'une part d'une documentation prescriptive qui rappelle aux chercheurs leurs obligations juridiques et déontologiques ; d'autre part d'une documentation de présentation des métadonnées qui précise les contextes et les conditions de collecte des matériaux, qui permettra aux utilisateurs de faire un bon usage des données et garantira le respect des discours des enquêtés.

Plusieurs scénarios de diffusion ont été envisagés : dans un cercle restreint, entre chercheurs participant à un même programme de recherche ; dans un cercle étendu à l'ensemble de la communauté scientifique ; enfin, en libre accès total. Les deux premiers scénarios nous semblent juridiquement et éthiquement envisageables, à la condition de supprimer les données les plus sensibles (en particulier des données qui pourraient nuire aux enquêtés). En revanche, le troisième scénario nous semble éthiquement trop problématique : nous proposons donc de ne jamais diffuser en libre accès les entretiens contenant des données à caractère personnel.

Par ailleurs, au vu de l'intérêt de la capitalisation et de la circulation des données qualitatives, et au regard de la complexité de la mise œuvre *ex post*, il semble impératif qu'une réflexion d'ampleur portant sur la gestion des données se généralise dans les laboratoires de recherche. Il s'agit alors pour les laboratoires de définir des politiques de gestion spécifiques aux données produites par leurs chercheurs.