



HAL
open science

Le travail des données. Entretien autour du service des enquêtes de l'Ined

Cris Beauchemin, Gwennaëlle Brilhault, Amandine Morisset, Patrick Simon, Juliette Galonnier, Stefan Le Courant, Camille Noûs

► To cite this version:

Cris Beauchemin, Gwennaëlle Brilhault, Amandine Morisset, Patrick Simon, Juliette Galonnier, et al.. Le travail des données. Entretien autour du service des enquêtes de l'Ined. *Tracés: Revue de Sciences Humaines*, 2019, Les sciences humaines et sociales au travail (II) : que faire des données de recherche, 19, pp.163-181. hal-02924168

HAL Id: hal-02924168

<https://hal.science/hal-02924168>

Submitted on 27 Aug 2020

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

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



Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0 International License

Entretien

Le travail des données. Entretien autour du service des enquêtes de l'INED

Data at work. An interview with INED's surveys unit

Cris Beauchemin, Gwennaëlle Brilhault, Amandine Morisset, Patrick Simon, Juliette Galonnier, Stefan Le Courant et Camille Noûs

p. 163-181

[Résumé](#) | [Index](#) | [Plan](#) | [Texte](#) | [Notes](#) | [Citation](#) | [Auteurs](#)

Résumés

[Français](#) [English](#)

L'Institut national d'études démographiques (INED) se distingue dans le paysage de la recherche français par l'attention qu'il porte à la mise à disposition des données issues des enquêtes de ses chercheurs, avec un service dédié depuis les années 1980. Dans cet entretien accordé à la revue *Tracés*, Gwennaëlle Brilhault, cheffe du service des enquêtes à l'INED, Amandine Morisset, chargée de la documentation des enquêtes, Cris Beauchemin et Patrick Simon, chercheurs à l'INED, mettent en lumière l'ampleur de ce travail de mise à disposition des données et la complexité des questions qu'il pose. Au cours de la discussion, elles et ils abordent tour à tour les défis concrets, en termes de coût et de temps, que pose ce travail ; son manque de reconnaissance ; les enjeux d'utilisation et de propriété intellectuelle qu'il soulève ainsi que le flou qui entoure la question des données dites *sensibles*. En détaillant la multiplicité des opérations de documentation et de présentation des données, cet entretien révèle que les « données » ne se suffisent pas à elles-mêmes et doivent être accompagnées d'un appareillage conséquent pour pouvoir être mises à disposition.

Entrées d'index

Mots clés :

[propriété intellectuelle](#), [archives](#), [données sensibles](#), [INED](#), [mise à disposition des données](#)

Keywords :

[intellectual property](#), [archives](#), [INED](#), [access to data](#), [sensitive data](#)
[Haut de page](#)

Plan

[Le travail de mise à disposition des données](#)

[Le travail des chercheurs et la propriété intellectuelle des données](#)

[La réutilisation des données](#)

[La question des données sensibles](#)

Créé en 1945, l'Institut national d'études démographiques (INED) produit des enquêtes quantitatives et qualitatives, dont les thématiques varient en fonction des projets de ses chercheurs. Le service des enquêtes de l'INED, qui existe depuis les années 1980, accompagne les chercheurs dans leurs enquêtes, de la conception à la mise à disposition, en passant par le processus de collecte. Il réunit actuellement 18 personnes, qui exercent des métiers différents (statisticiens, méthodologues, informaticiens et logisticiens d'enquête, chargés de traitement et spécialistes de la mise à disposition de données d'enquêtes). Dans cet entretien accordé à la revue *Tracés*, Gwennaëlle Brilhault, cheffe du service des enquêtes et Amandine Morisset, chargée de la documentation et de la mise à disposition des enquêtes, accompagnées de deux chercheurs de l'INED, Patrick Simon et Cris Beauchemin, reviennent ensemble sur l'effort de mise à disposition des données d'enquêtes fourni par l'Institut. En réfléchissant collectivement et à voix haute sur les coûts, les dilemmes et les interrogations que ce travail implique, elles et ils soulèvent des questionnements importants sur ce qui constitue une « donnée » en sciences sociales et une donnée « sensible » en particulier, ainsi que sur les enjeux scientifiques, techniques, politiques et déontologiques que pose la mise à disposition.

[Le travail de mise à disposition des données](#)

TRACES : Sur combien d'enquêtes le service des enquêtes de l'INED travaille-t-il en ce moment ?

GWENNAËLLE BRILHAULT : Actuellement, nous travaillons sur une bonne vingtaine d'enquêtes en même temps. Certaines se font entièrement en dehors du service ou avec l'appui du service uniquement sur certains moments de l'enquête ; dans beaucoup de cas néanmoins, nous intervenons sur l'ensemble de la vie des données, de la construction à la mise à disposition. Parfois, l'INED sous-traite la collecte à des instituts de sondage privés. L'enquête « Virage DOM »¹ a par exemple été conduite avec les enquêteurs téléphoniques d'Ipsos Outre-mer, avec un suivi très régulier de l'équipe de l'INED. Pour l'enquête « Trajectoires et Origines 2 » (TeO 2)², c'est encore une autre façon de faire avec une maîtrise d'ouvrage INED-Insee et des enquêteurs Insee qui vont faire le travail de collecte. En somme, nous travaillons toujours en parallèle sur des enquêtes passées qu'il faut finir de documenter, mettre en forme, mettre à disposition, et une quinzaine d'enquêtes qui sont à différents stades de préparation : certaines sont très préliminaires, pour d'autres nous sommes en train de finaliser le dossier à soumettre à la CNIL (Commission nationale de l'informatique et des libertés), et d'autres seront bientôt sur le terrain.

TRACES : Concrètement en quoi consiste le travail de mise à disposition des données ?

AMANDINE MORISSET : Cela dépend, car chaque enquête est différente au niveau du procédé de collecte, de l'enquête elle-même, de celles et ceux qui la font. Les fichiers de données qui nous arrivent – quand ils nous arrivent – sont dans des états totalement différents, plus ou moins finalisés. Le service de mise à disposition a véritablement commencé son travail

autour des années 2000³. L'INED par contre existe depuis 1945, ce qui veut dire qu'il y a tout un historique à rattraper et qui est en même temps irrattrapable : par faute de temps, par défaut de fichiers... Il y a parfois des fichiers que j'ai sauvés de disquettes. On rencontre aussi souvent des formats Excel que l'on doit repasser dans des formats plus récents pour pouvoir les conserver. Le pire des cas, ce sont les fichiers qu'on ne peut plus lire car trop vieux. C'est dommage, c'est une perte. Maintenant on documente au minimum en trois formats (SAS, SPSS, qui sont des formats de logiciels statistiques et CSV, un format texte) donc on peut espérer que pour l'avenir on soit un peu couvert. Notre travail va vraiment s'adapter à l'équipe de recherche, quand elle est encore « vivante », à ce qui reste dans nos archives, ou à ce qu'on a réussi à récupérer avec le temps. Ce dont on a idéalement besoin pour la mise à disposition, c'est d'un fichier de données finalisé et plus ou moins anonymisé (pas au sens strict du RGPD⁴, mais plutôt qui vise à limiter autant que possible les risques de réidentification des personnes concernées), accompagné de sa documentation, et si possible de contacts avec l'équipe de recherche. Mais on a aussi des enquêtes pour lesquelles on n'a pas de fichier, ou pour lesquelles on a le fichier mais sans la documentation, avec des variables dont on ne sait pas exactement à quoi elles renvoient. Parfois il faut aussi apurer davantage car l'apurement n'a pas été réalisé dans une optique d'anonymisation, voire n'a pas été réalisé du tout. *Apurer* signifie séparer les variables utiles de celles qui sont inutiles, dans le cadre de la mise à disposition. Ensuite, dans la variable elle-même, il faut s'assurer que les modalités (c'est-à-dire les différentes réponses) correspondent à la documentation et au questionnaire, qu'il n'y a pas d'erreurs ou encore de modalités qui n'existent pas... Il m'arrive aussi de procéder à des regroupements. Par exemple, lorsqu'une variable est un peu trop détaillée, on la met à disposition sous forme agrégée. En somme, le travail de mise à disposition, c'est beaucoup de documentation, pêcher les informations à droite et à gauche dans les archives, auprès des personnes, et documenter comment ont été faits à la fois l'enquête et le fichier de données. Pour nous, c'est très important de laisser une trace précise de comment on a fait l'enquête, pourquoi on l'a faite, avec quels moyens et de lui donner un vrai contour pour qu'elle entre dans la postérité et puisse être consultée des dizaines d'années après par les nouveaux chercheurs, à l'INED ou ailleurs. Le fichier de données doit également être propre et documenté pour qu'il soit lui aussi réutilisable des années et des années après. On a des enquêtes qui sont consultées aujourd'hui qui datent des années 1970. Et j'espère que ça ne va pas s'arrêter ! Les gens ont soif d'historique. Quand ils se lancent dans une thématique, ils veulent savoir ce qui a été fait sur le sujet. On a cette richesse-là à l'INED. En tout cas on essaie, au fur et à mesure du temps d'enrichir cette possibilité de consulter tout ce qui a été fait dans le passé et à l'heure actuelle.

CRIS BEAUCHEMIN : Il faut garder à l'esprit que les fichiers sont des codes, des numéros. On ne peut rien faire des fichiers tant qu'on n'a pas la documentation qui renseigne la signification des modalités (1 ça veut dire ça, 2 ça veut dire ça, 3 ça veut dire ça) pour chaque variable. Un fichier sans cette information, c'est comme si tu avais perdu tes données. Et puis parfois tu as des variables qui sont recodées entre le questionnaire et le fichier. Le coup classique ce sont les variables où tu réponds par *oui* ou par *non*. Tu as des questionnaires dans lesquels c'est codé 1-2. Et en pratique dans les fichiers, on préfère quand c'est 0-1. Or, si tu passes de 1-2 à 0-1 mais que ça n'a pas été documenté, le 1 a changé de sens...

PATRICK SIMON : Sur les variables de sexe par exemple, il y a une convention : le 1 c'est l'homme et le 2 c'est la femme. Or, plusieurs équipes de recherche changent cette

convention, pour des raisons politiques notamment. Et donc tu peux te retrouver à prendre les hommes pour des femmes et les femmes pour des hommes.

A. MORISSET : Ce 1-2, c'est horrible. À toutes les enquêtes ça change. Le service de mise à disposition est la dernière roue du carrosse qui doit s'assurer que tout correspond. Des fois ça prend des mois, car les enquêtes deviennent de plus en plus lourdes. Il y a de plus en plus de questions, de plus en plus longues, de plus en plus complexes parfois, avec plusieurs questionnaires. Ça peut se révéler être un casse-tête.

TRACES : Mettre à disposition des données implique donc de faire un ensemble de choix, de regrouper les variables, de les nettoyer, etc. Les décisions se prennent-elles en coordination avec les chercheurs ou devez-vous parfois les prendre seules ? Quelles sont les négociations entre chercheurs et service des enquêtes sur la mise à disposition ?

A. MORISSET : Quand on peut, on en discute : qu'est-ce qu'on fait avec ci, avec ça ? Malheureusement parfois c'est compliqué, même avec des chercheurs qui sont encore en activité, mais qui sont débordés. Donc parfois on est obligé de faire des choix. On essaie de trancher entre nous. Le but c'est de prendre des décisions raisonnables sur ce qu'on met à disposition ou pas. Notre intérêt n'est pas d'être en conflit avec les chercheurs. Là où c'est un peu compliqué et où on se bat encore souvent, c'est plutôt avec les chercheurs qui restent réticents à mettre à disposition leurs données et qui font traîner le processus.

TRACES : Est-ce que vous conservez une trace des décisions de mise à disposition que vous prenez ?

A. MORISSET : Tout est documenté ! Chaque variable qui est supprimée est documentée. A posteriori, on peut voir que telle décision a été prise, pour des raisons d'anonymisation, de supprimer ça, ça, ou ça. En règle générale, je préfère la solution la plus sécuritaire à la moins sécuritaire. Ensuite, les variables en question sont censées être conservées aux archives et retrouvables. La liste des variables supprimées est mise en ligne sur Nesstar (Networked Social Science Tools And Resources)[5](#) dans la documentation qui est accessible à tous.

G. BRILHAULT : En tout cas, le travail d'Amandine c'est de parvenir à un ensemble de documentation qui se suffise à lui-même : n'importe qui dans dix ou vingt ans doit pouvoir s'y repérer entièrement. L'idée est de pouvoir répondre par avance à toutes les questions que l'utilisateur se poserait en utilisant ces données.

C. BEAUCHEMIN : Je voudrais insister sur le fait que c'est un travail immense, que très peu d'institutions font à l'heure actuelle. En France, il n'y a pas d'obligations à le faire, tandis qu'en Grande-Bretagne un financement en ESRC (Economic and Social Research Council) exige que les données soient mises à disposition. C'est pour cela que cet effort que fait l'INED de mettre à disposition les données d'enquête, certes pas systématiquement mais tout de même très souvent, est assez extraordinaire. Mais c'est un travail énorme, porté par une institution qui a fait le choix de se doter d'un service pour cela et qui repose aussi sur les équipes de chercheurs qui doivent préparer les données pour qu'elles soient utilisables par d'autres. Pour l'enquête « MAFE » (Migrations entre l'Afrique et l'Europe)[6](#), cela a été possible car on avait pendant toute la durée de ce projet européen et même après, un *data manager* qui a mis en forme les données. Mais en réalité très peu de projets prévoient cette

dépense dans leur budget. Dans la plupart des cas, les données ne sont pas dans un état tel qu'elles peuvent être transmises à Amandine. Le travail de préparation, en amont, est très coûteux en temps et difficile à évaluer. En général, il n'est pas budgété et donc n'est pas pris en charge. Une autre difficulté tient aussi à la temporalité des projets sur contrat. Quand on a seulement trois, voire quatre ans, pour concevoir, réaliser et analyser une enquête, on n'a pas le temps de nettoyer et documenter les données pour les autres. Dans des durées si courtes, tu es déjà content d'avoir les tout premiers résultats à la fin ! Compte tenu des modes de financement des projets d'enquête, compte tenu surtout de la durée de ces projets-là, il est quasiment impossible d'intégrer tout le travail de préparation de la mise à disposition. Des moyens supplémentaires sont nécessaires.

G. BRILHAULT : Oui, la mise à disposition, c'est souvent la dernière roue du carrosse. C'est le truc qu'on n'a pas prévu et qu'on laisse tomber. Pour le prévoir correctement, il faut le faire en amont, et c'est moins douloureux si on prévoit de mettre de côté les choses progressivement, avec quelqu'un dans le budget pour le faire. Et pour rebondir sur ce que tu disais, l'ANR commence tout de même à exiger que les chercheurs expliquent dans leur demande de financement ce qu'ils vont mettre à disposition et quand. Et c'est vrai qu'en exigeant que cette réflexion se fasse dès le départ, ça peut aider tout le monde à anticiper les difficultés et faire en sorte que ce soit moins douloureux. Un chercheur qui ne l'a jamais fait ne va pas voir la difficulté. Un chercheur dont c'est la deuxième enquête va forcément essayer d'anticiper les choses !

A. MORISSET : Il y a des équipes de recherche qui n'anticipent pas et se rendent compte du travail une fois qu'ils se posent la question : « Comment va-t-on faire pour mettre à disposition ? Qu'est-ce qu'il vous faut ? » Des fois on n'est pas accueilli qu'avec des sourires ! Parce qu'ils se rendent compte de la charge de travail qui va en découler. Certes, ils ont leur base de données documentée, mais pas toujours de la façon aussi précise que nous voulons. C'est sûr que c'est une culture à mettre en place.

Le travail des chercheurs et la propriété intellectuelle des données

TRACES : *Est-ce que vous avez des exemples où la projection de la mise à disposition a transformé la manière même de concevoir et de mener l'enquête ?*

C. BEAUCHEMIN : Nos enquêtes sont souvent très larges, et elles permettent de répondre à beaucoup de questions. C'est justement ce qui justifie leur mise à disposition : on peut en faire beaucoup d'exploitations, bien au-delà de ce que font les équipes qui les ont conçues. Sur les enquêtes « MAFE » ou « TeO » par exemple, c'est frappant. Les données sont utilisées maintenant par de nombreux autres chercheurs qui font plein de choses auxquelles on n'avait pas forcément pensé ou que, de fait, on n'a pas faites. On peut en prendre la mesure en regardant les listes des publications sur les sites de ces enquêtes. Par exemple, un article vient d'être publié sur le mariage à Kinshasa, qui montre que le mariage recule du fait des difficultés économiques des jeunes hommes qui ne parviennent pas à une sécurité économique leur permettant de s'engager⁷. Le projet était conçu pour étudier les migrations internationales, mais d'autres usages sont possibles.

G. BRILHAULT : Ce qui se passe dans la mise à disposition, c'est qu'on est obligé de faire un appauvrissement des données pour que justement on puisse les mettre à disposition sans mettre les enquêtés en difficulté parce que leur anonymat serait remis en question. C'est plutôt la mise à disposition qui va conduire à une version dégradée du fichier. Je ne pense pas que ça influe dans l'autre sens.

P. SIMON : Sur cette problématique de mise à disposition, je pense qu'il faut réfléchir en termes de taille et de conception des enquêtes. Certaines enquêtes sont des équipements de recherche : même s'il y a eu beaucoup de réflexion de la part des chercheurs qui y contribuent en amont, elles ont vocation à servir à des réexploitations. On est donc d'emblée dans de la diffusion. Il y en a d'autres, souvent sur des échelles beaucoup plus petites, qui ne sont pas très éloignées d'enquêtes qualitatives en réalité, bien qu'elles soient quantitatives. Elles sont conçues comme étant un appareil de recherche individuel et beaucoup moins dans l'idée de la diffusion. De ce point de vue-là, les problématiques sont les mêmes que pour quelqu'un qui va faire des entretiens sur le terrain et qui se demande comment faire pour donner accès à ces entretiens. Il y a nécessairement de la cuisine dedans qui n'est pas toujours communicable sans un travail de présentation. Est-ce que quelqu'un va relire mon entretien et s'apercevoir que j'ai posé les questions d'une façon un peu différente de la grille ou que je dis « euh euh euh » toutes les trois phrases ? Il y a l'idée que tu es évalué sur ta compétence à toi et que tu t'exposes. Je ne sais pas si c'est une bonne raison mais cela peut expliquer que certains chercheurs aient des susceptibilités à ce sujet. Ce n'est pas évident de laisser des tiers entrer dans ton système méthodologique. L'idée que l'enquête soit un appareil de recherche communicable, réutilisable et appartienne de fait à la collectivité n'est pas complètement évidente.

C. BEAUCHEMIN : Il peut effectivement y avoir des susceptibilités. Il y a aussi une part de protectionnisme : un certain nombre de chercheurs ne souhaitent pas que d'autres qu'eux utilisent les données. Concevoir et réaliser une enquête prend beaucoup de temps. On peut avoir envie de récolter les fruits de cet effort avant les autres. On peut concevoir qu'un chercheur qui a investi dans la préparation d'une enquête veuille être le premier à sortir des publications avec ces données et qu'il n'ait pas envie d'être doublé par un collègue qui n'a pas fait cet effort de production des données. C'est le nerf de la guerre. On peut aussi concevoir qu'un chercheur, qui a déjà investi dans la conception et la réalisation d'une enquête, n'ait pas envie d'investir encore davantage pour partager les données. Parce que, comme on l'a dit, préparer les données pour la mise à disposition est coûteux. En fait, il faudrait que tu fournisses un effort supplémentaire pour que d'autres exploitent les données que tu as produites !

P. SIMON : Oui, il y a un problème de balance entre la contribution à la collectivité à travers l'appareil que le chercheur construit et le temps qu'il y consacre. Un chercheur n'est pas un ingénieur de recherche : il n'a pas pour mission de produire des données mais d'apporter des savoirs sur la société. La production de données fait partie de ces savoirs, mais elle incorpore des compétences qui ne sont pas reconnues. Le temps investi, disons les quatre ans mis à produire des données, est du temps qui n'est pas dédié à publier. Or, le paysage international de la recherche est de l'ordre du pousse-bouton : aujourd'hui l'exploitation des données est très rapide et très simple, en raison des systèmes informatisés dont on dispose, ce qui n'était pas le cas dans les années 1970. À l'époque, l'exploitation était un travail long

et compliqué, il fallait bien connaître les données, il y avait un coût d'entrée très lourd, n'importe qui ne se mettait pas à réutiliser une enquête comme ça. Aujourd'hui, en quelques heures, n'importe quel chercheur peut faire une analyse de données sur une enquête qui est disponible ! Il y a des enquêtes comme l'« Enquête sociale européenne » (ESS) qui sont conçues pour être un appareil de recherche : un collectif travaille à faire les questionnaires, mais c'est un engagement reconnu. Par contre, pour les enquêtes conçues spécifiquement pour répondre à des questions que l'on se pose, c'est vrai que les mettre à disposition peut constituer un *trade-off* (compromis) douloureux. Cela aboutit à solidifier deux strates de chercheurs dans le monde : celles et ceux qui vont produire les données et qui vont y passer beaucoup de temps, et celles et ceux qui vont exploiter les données disponibles et sortir en très peu de temps des publications sur le sujet. Le risque est de voir disparaître les chercheurs consacrant du temps à monter des opérations de collecte chronophages s'il n'y a pas de reconnaissance spécifique de cet engagement.

C. BEAUCHEMIN : Oui et non. Pour moi, un bon exemple qui montre que ce peut être productif, c'est celui de Douglas Massey (sociologue américain, université de Princeton). Il est une des grandes pointures dans le domaine des migrations internationales. Or il met systématiquement ses données à disposition. Et il dit que sa renommée vient aussi de là. C'est-à-dire qu'à partir du moment où tu fais connaître tes données par ton travail, que tes données sont mises à disposition et que tu es correctement cité, alors tes données participent à ta publicité.

P. SIMON : Oui mais il ne fait pas des enquêtes qui s'appellent « l'enquête de Princeton ». Il fait « l'enquête Doug Massey ». Alors que nous faisons des enquêtes qui sont estampillées Insee-INED. Ce n'est l'enquête ni de Cris Beauchemin ni de Patrick Simon.

C. BEAUCHEMIN : Tu as raison, la question, c'est celle du *trade-off*. Moi je pense plutôt qu'il y a un bénéfice à mettre ses données à disposition. Le peu de renommée que j'ai dans les migrations internationales, ça vient largement des données en fait, notamment de l'enquête « MAFE ». Mais effectivement ça demande un effort préalable important. Cela pose aussi la question du calendrier de mise à disposition de l'enquête. Selon le dispositif institutionnel dans lequel tu t'inscris, tu as l'obligation ou pas de mettre à disposition et de le faire plus ou moins vite. Quand on fait des enquêtes avec l'Insee, il y a une mise à disposition qui est très rapide et qui vient avant même que nous ayons eu le temps de publier. Dans ce genre de cas, la paternité de l'enquête peut échapper aux chercheurs-concepteurs. Ce n'est pas très plaisant de voir sortir des publications qui exploitent les données que l'on a produites sans être proprement cité.

TRACES : *Comment s'assure-t-on de la reconnaissance de ce travail ? Qui a la propriété intellectuelle des données ? Est-ce qu'on ne pourrait pas imaginer un système de citation qui mettrait plus en valeur le travail du service des enquêtes et des chercheurs ?*

A. MORISSET : Il y aurait beaucoup trop de gens à citer [rires] !

G. BRILHAULT : Effectivement les chercheurs sont auteurs de leurs publications donc c'est là qu'ils peuvent être cités nommément, mais par contre pour les données... Du point de vue de la propriété, les données appartiennent à celui qui a financé la collecte, en l'occurrence l'INED-Insee pour « TeO ».

A. MORISSET : Par ailleurs, sur le réseau Quételet Progedo Diffusion⁸, on a une base de questions. Pas pour toutes les enquêtes mais pour un certain nombre d'enquêtes qui ont été documentées, tu peux taper un sujet de recherche et obtenir plein de questions. Ça s'appelle *La Base de Questions*. Ce n'est pas une *data bank*, mais une *question bank*, en l'occurrence.

P. SIMON : S'ajoute donc aussi le fait que beaucoup de questionnaires sont repris pour d'autres enquêtes, mais ces reprises ne sont jamais créditées. La circulation des questions est très positive, d'ailleurs je participe au projet « Ethmigsurvey » qui construit une banque de questions qui permettront aux chercheurs travaillant sur les migrants et les minorités ethniques de s'en inspirer. Mais la mise à disposition des questionnaires pose la question de la propriété scientifique. Il n'y a pas de propriété du questionnaire.

C. BEAUCHEMIN : C'est vrai en sciences sociales mais en santé par exemple, les questionnaires appartiennent aux gens qui les ont conçus et ils sont vendus. C'est très standardisé. Au fond dans la conception des questionnaires et dans la fabrique des données elles-mêmes, c'est peut-être ce manque de reconnaissance qui est le principal frein à la mise à disposition dans nos disciplines.

La réutilisation des données

TRACES : *Y a-t-il des droits d'utilisation des données ? Tout le monde peut-il les utiliser ?*

A. MORISSET : Non, nos données sont réservées au monde de la recherche. Il faut faire une demande sur le réseau Quételet Progedo Diffusion pour y accéder. Les utilisateurs doivent s'authentifier, créer un compte, en détaillant qui ils sont, de quel institut ils viennent, etc. Puis, on étudie leur demande en fonction de leur projet de recherche. Nos données ne sont pas accessibles aux particuliers ni aux journalistes. Il faut prouver qu'on vient bien d'une institution de recherche et qu'on est lié à la communauté scientifique, ce qui peut inclure les étudiants de master. Jusque-là ça s'est passé comme ça. Maintenant il va y avoir des changements.

G. BRILHAULT : En effet, il peut y avoir des changements en raison de la loi pour une République numérique de 2016, qui dit à l'administration d'ouvrir davantage ses données quand elles ne sont pas protégées par un autre secret et qui dit aussi que ce qu'on met à disposition d'une catégorie d'utilisateur, il faut l'ouvrir à tous. Si bien que, dans ce cadre-là, une réflexion a été engagée autour des fichiers de production et de recherche (FPR) qui sont sur le réseau Quételet Progedo Diffusion. Si on n'avait rien fait, on aurait pu considérer qu'à partir du moment où ces fichiers étaient mis à disposition de la communauté scientifique, la loi pour une République numérique impliquait de les mettre à disposition de tous. Or ces fichiers n'ont pas du tout été conçus pour être à disposition de tous sur Internet. Ils ne sont pas documentés pour. Il faut les pondérer pour rétablir leur représentativité. Ils portent sur des sujets compliqués. Ce sont des données individuelles. Ils sont anonymisés, entre guillemets, et certes on les appauvrit, mais le moins possible pour qu'ils restent intéressants pour l'activité de recherche. Donc la réflexion qui a été faite d'abord à l'Insee et plus largement au sein du Comité du secret statistique, c'est de considérer que ces fichiers-là sont soumis au secret statistique. Au lieu de simplifier, on rajoute un détour, qu'on va

essayer de rendre le plus léger possible, et qui consiste, pour ces fichiers-là, à faire homologuer les chercheurs auprès du Comité du secret statistique. Pour ne pas être obligé de les mettre en ligne. Parce que ce serait dangereux de les mettre en ligne tels quels. On n'a pas du tout le temps de les retraiter, et je ne sais pas si on y arriverait, de manière à en faire des fichiers grand public. Certes, le principe de la loi pour une République numérique, c'est d'ouvrir davantage. Nous ne voulons en aucun cas ouvrir moins. Nous voulons continuer à pouvoir donner ce niveau de détail à un public privilégié et qualifié, qui est la communauté scientifique.

P. SIMON : Vous dites : « C'est dangereux ». Quel est le danger en fait ?

G. BRILHAULT : Que quelqu'un qui ne connaisse pas les sujets ou qui ne sache pas forcément manipuler les données ou qui n'ait pas de background scientifique ou statistique se retrouve avec ces données en main et en dise n'importe quoi. Ce sont quand même des données individuelles ! Et puis on est quand même sur des sujets souvent sensibles. Quel est votre parcours migratoire ? Quels sont vos revenus ?

A. MORISSET : Certes on les anonymise, mais entre tout enlever comme information et tout laisser, on essaie de maintenir une balance pour ne pas trop perdre d'informations tout en garantissant un certain anonymat. Maintenant en croisant 15 variables et à condition de savoir qu'une personne a répondu à l'enquête, on pourrait peut-être retrouver quelqu'un. On ne peut jamais être sûr à 100 %.

TRACES : *Ce qui ressort ici, c'est donc un risque de mésusages des données. Mais est-ce que cette question ne se pose pas aussi avec des chercheurs ?*

P. SIMON : Bien sûr que si ! Je vais donner un exemple, qui renvoie à cette question du niveau de données à rendre accessible. À la suite du papier sur les transmissions des prénoms que nous avons publié avec Baptiste Coulmont⁹, un chercheur a eu envie de reproduire ce qu'on a fait parce qu'il n'était pas convaincu par les résultats. Il a voulu faire une contre-analyse. Assez hostile. Les prénoms eux-mêmes étaient protégés par le Comité du secret statistique, mais ce chercheur a fait une demande au Comité du secret pour les obtenir. Puis il m'a contacté pour me demander notre système de classification des prénoms. Il m'a donc demandé un outil de recherche et a justifié sa demande par un objectif de répliquabilité de l'étude, comme s'il s'agissait d'une obligation de permettre cette répliquabilité. C'est une question déontologique et scientifique passionnante à laquelle je n'ai pas de réponse, mais l'INED l'a fait étudier sur le plan juridique et déontologique. La classification que nous avons fait faire est considérée comme un brevet et doit faire l'objet d'une licence d'exploitation. On voit qu'on sort progressivement de l'ère du bricolage pour entrer dans des systèmes juridiques et déontologiques de production, d'exploitation et de mise à disposition des données plus normés. C'est encore une autre étape de réflexion : dans les strates successives de contribution, il y a ce qui doit être diffusé et ce qui relève de ta création scientifique, qui t'est personnelle.

C. BEAUCHEMIN : Par ailleurs, il y a un certain nombre de revues dans notre domaine qui exigent maintenant que tu mettes à disposition, quand ils te publient, tes données et tes programmes, à des fins de répliquabilité. C'est le cas à *Demographic Research* par exemple. Et ça demande un travail supplémentaire, parce qu'effectivement quand tu travailles tu peux

bien faire ton travail mais de façon néanmoins un peu brouillonne. Dès lors que tu dois rendre ton fichier de données et ton programme disponibles, il faut faire le ménage ! C'est comme si tu devais ranger ton bureau à chaque fois. Mais sur le fond je trouve ça bien de pouvoir garantir la répliquabilité, de même que, sur le principe, je trouve ça bien qu'on mette nos données à disposition parce qu'elles sont produites avec de l'argent public. Cependant l'application de ces principes-là est plus difficile. En fait, on est d'accord sur le principe mais ça a un coût énorme qui est souvent non estimé ou sous-évalué et dont la récompense peut être parfois douteuse : tu n'es pas forcément cité, tu peux être démolé, etc. Il y a une mise en danger.

G. BRILHAULT : En fait la question, c'est à quel point tu as le sentiment qu'il faut que tu proposes des instruments d'analyse de tes données parce qu'elles sont complexes. Ce peut être une façon de normer l'utilisation des données, parce que tu penses que ça va être trop négatif si tu ne le fais pas et que des personnes qui connaîtraient moins bien les données risquent de ne pas les utiliser à bon escient.

C. BEAUCHEMIN : Pour l'enquête « MAFE », qui est une enquête très complexe du point de vue de la structure des données, on a mis au point des petits manuels, accessibles en ligne, qui donnent des exemples de programmation pour pouvoir exploiter les données. Cela a été possible grâce à l'appui du service des méthodes statistiques de l'INED, et notamment l'intervention de Marc Thévenin. Mettre les données à disposition, ça peut ne servir à rien si les données sont si complexes que les usagers ne savent pas s'en servir.

TRACES : *En fait tout cela pose la question : Qu'est-ce qu'une donnée ? Au début de l'entretien, vous avez dit que les fichiers de données mis à disposition doivent être « des éléments qui se suffisent à eux-mêmes ». Mais en fin de compte, on s'aperçoit qu'il faut tout un appareillage pour qu'une donnée se suffise à elle-même. La donnée dite brute reste très fragile et sujette à mésinterprétation.*

A. MORISSET : Ça dépend de qui l'utilise et comment. Le problème est qu'on ne peut pas avoir un contrôle individuel sur chaque personne.

P. SIMON : Moi j'assumerai complètement le risque du mésusage : si quelqu'un veut utiliser les données d'une façon critiquable, eh bien ! qu'il publie et on verra très clairement le biais dans l'utilisation. Cette personne prend des risques, fatalement, du point de vue de sa notoriété et de sa crédibilité. Donc je pense qu'il y a une régulation des mésusages par le débat entre pairs. Cela suppose néanmoins de se situer dans l'univers du dialogue scientifique ! Si ce sont des *fake news* qui sont échangées par médias interposés, cela rend les choses plus compliquées. Par exemple les adeptes du « grand remplacement », qui sont assez imaginatifs en matière d'utilisation des données, peuvent s'en saisir d'une façon problématique.

C. BEAUCHEMIN : Non car en fait, ils mobilisent plutôt des données secondaires et des résultats déjà publiés. Ce n'est pas si facile d'exploiter des données d'enquêtes, de transformer des réponses individuelles en résultats synthétiques. Sans compétences en analyse statistique, sans logiciel spécialisé et sans le temps nécessaire, il n'est pas possible d'entrer dans les données et d'en faire quelque chose. C'est tellement plus facile de manipuler l'opinion en disant n'importe quoi à partir de données secondaires plutôt que d'exploiter des données

micro ! Je ne crois pas non plus tellement au mauvais usage des données individuelles d'enquêtes.

P. SIMON : Sur le fond, il est possible de limiter les risques en construisant un cadrage serré des analyses. Pour « TeO », je suis quasiment certain qu'il est très difficile d'y trouver des scoops négatifs. Je pense qu'au contraire on a plutôt intérêt à être dans la transparence. La question de la transparence signifie que même quand des personnes mal intentionnées veulent utiliser des données, elles ne peuvent pas leur faire dire n'importe quoi.

G. BRILHAULT : Parce que tu les as suffisamment retravaillées et explorées toi-même. Mais prenons l'exemple d'une enquête qui porte sur les violences subies à la fois par les hommes et les femmes : lorsqu'on fait des prévalences à plat, on trouve que les hommes déclarent subir autant de violences que les femmes. Ce qui paraît étonnant. En fait il faut rentrer dans des choses beaucoup plus fines, voire du qualitatif, pour comprendre ce que les hommes mettent derrière cette question. Donc il faut que l'équipe arrive à une maturité de la base de données qui permette d'éviter que ce genre de mésinterprétations rapides soit fait par n'importe qui.

C. BEAUCHEMIN : L'exemple sur les violences est assez intéressant de ce point de vue-là. Tu te dis : « Comment la question a-t-elle été tournée ? » Dès lors, il te faut le libellé de la question exacte. Et puis peut-être que, pendant la formation des enquêteurs, l'équipe a donné des indications complémentaires pour qu'ils puissent guider les enquêtés en cas d'hésitation : « Si les enquêtés hésitent, vous ne dites rien de plus, c'est à eux d'interpréter » ou alors : « Si les enquêtés hésitent, vous leur donnez des exemples ». Dans ce cas, il faut savoir quels étaient ces exemples. Donc, idéalement, il faut aussi mettre les manuels de formation des enquêteurs à disposition. Souvent, il y a des foires aux questions qui sont mises en place pendant le travail de terrain : les enquêteurs se posent des questions, et les concepteurs y répondent. Est-ce qu'il faut aussi mettre ça à disposition parce que potentiellement ça influence la façon dont les gens répondent ? Effectivement, il y a toute une cascade depuis le niveau zéro de l'information qui est le nom de la variable, son libellé, jusqu'à la façon dont ça a pu être compris par les enquêtés. Quand tu es face à des données avec des modalités simples de réponse, en réalité, tu as un ensemble de cercles concentriques d'informations qui donne du sens aux réponses des enquêtés.

La question des données sensibles

TRACES : *Qui est-ce qui qualifie une donnée de sensible ?*

G. BRILHAULT : Il y a une définition des données sensibles au sens du RGPD. Il s'agit des données d'origine raciale ou ethnique, les opinions politiques, les convictions religieuses ou philosophiques, l'appartenance syndicale, les données génétiques et biométriques, les données concernant la santé ou les données concernant la vie sexuelle ou l'orientation sexuelle d'une personne physique. Si on souhaite collecter ces données, il faut dans le dossier à la CNIL justifier que le projet de recherche nécessite de le faire et détailler comment ces données seront ensuite protégées.

C. BEAUCHEMIN : À ma connaissance, il n'y a eu que peu de cas à ce jour où la CNIL a dit préalablement : « Non, vous ne poserez pas cette question ». En revanche, on peut s'autocensurer, soit individuellement comme chercheur, soit au niveau de l'institution qui n'a pas envie de mettre en jeu sa réputation. Quand tu fais ton dossier CNIL, tu indiques de façon préalable comment les données seront rendues disponibles. Tu peux avoir différentes approches : tu peux dire que ce ne sera jamais disponible, ou que ce sera rendu disponible mais avec des garanties d'anonymat. Dans l'enquête « MAFE » par exemple, on a des questions sur la trajectoire légale des migrants. Ces données-là sont accessibles avec l'ensemble des autres données du projet : on l'avait déclaré d'emblée et ça n'a pas posé de problème, mais uniquement parce que ces données sont réputées anonymes. En réalité, le champ des interdictions est très restreint, pour peu qu'on respecte le principe de l'anonymat.

P. SIMON : En fait il y a un flou sur ce qui relève des prérogatives de la CNIL, ou en tout cas une marge d'interprétation sur ce qu'il faut soumettre à autorisation ou pas, et du côté de la CNIL la règle de décision est parfois imprévisible par manque de jurisprudence. Cela ouvre un champ d'incertitude qui n'est pas négligeable. Il y a des institutions qui se préoccupent peu de ces enjeux, comparativement à l'INED. Jusqu'à récemment, le fait de soumettre son enquête ou son traitement de données au correspondant informatique et libertés (CIL) de son établissement dépendait beaucoup de la bonne volonté des chercheurs.

G. BRILHAULT : La proximité de la statistique publique avec l'INED fait que forcément l'INED est plus précautionneux sur ces enjeux.

P. SIMON : Nous sommes dans un environnement très particulier et je ne sais pas si l'on peut tirer des lois générales de cet environnement particulier. Il y a des enquêtes faites par des chercheurs qui n'ont pas les contrôles que nous rencontrons et ne passent pas par les circuits que nous évoquons. En fait le dispositif est à géométrie variable. Il y a des publications dans des revues de sciences sociales de premier plan qui ne demandent pas s'il y a eu des garanties de la CNIL pour cette recherche et la CNIL ne va pas contrôler après coup.

G. BRILHAULT : Ils n'ont pas le temps ! Mais le RGPD renverse la responsabilité : désormais, c'est toi qui es responsable de faire les traitements qui conviennent. Tu seras comptable des problèmes. Il faudra pouvoir apporter la preuve que tu as fait tout ce qu'il fallait pour protéger ces données. C'est pour cela que le RGPD rend tout le monde très nerveux. Tout le monde est en train de surajuster.

P. SIMON : Le problème, c'est que c'est une évaluation ex-post. Quand tu regardes dans la pratique, est-ce que tu connais une seule condamnation faite par la CNIL sur des usages frauduleux de fichiers ? En fait la crédibilité de la menace de la CNIL est zéro. Par ailleurs, il y a une énorme loterie. Quand ils ont mis le RGPD en place, on faisait une enquête avec une autre université et avec un prestataire d'enquête privé. On avait donc deux CIL en dialogue. Et pour la première fois le CIL du prestataire privé, qui jusqu'ici négociait en début d'année un programme de travail annuel pour toutes les enquêtes et obtenait une autorisation générique, a décidé qu'il fallait une autorisation spécifique pour cette enquête, qui portait sur les discriminations – un sujet qui rend les gens nerveux. Ils ont dû négocier avec le CIL de l'autre université qui lui, par contre, était ignorant à 100 % ! Et là tu tombes dans une sorte

d'abysse où quelqu'un est censé valider les documents mais, d'une part, il n'ose pas prendre la responsabilité et, d'autre part, il ne connaît absolument pas quelles sont les conditions pour valider. Cette enquête sur les discriminations a fait tilter plein de voyants, qui se sont activés parce qu'on travaillait avec des données dites *sensibles*.

A. MORISSET : Dans les faits, lors de la mise à disposition, on évite de mettre les variables sensibles dans le fichier de production et de recherche. On les supprime ou on les agrège suffisamment pour qu'elles soient moins sensibles. Ensuite le chercheur qui est intéressé par ces données-là et qui a creusé le premier niveau de fichiers accessibles sur le réseau Quételet Progedo Diffusion peut s'adresser au Comité du secret statistique.

TRACES : *Les enquêtes de l'INED ont-elles vocation à être déposées aux Archives nationales ?*

G. BRILHAULT : Toute production de l'INED, que ce soit au niveau administratif ou recherche, fait l'objet d'un descriptif dans un fichier de gestion qui précise si ces données doivent être versées aux Archives nationales et dans quel délai. Il y a un débat actuellement sur la question des archives. Pour nous ça allait de soi qu'on gardait une copie des fichiers, mais pour la CNIL, ça ne va pas de soi. Parce qu'avec le RGPD, à l'issue du traitement (quand on a fini toutes les exploitations), on doit pouvoir dire aux enquêtés que leurs données ont été détruites et donc qu'elles sont bien protégées, puisqu'elles n'existent plus. Alors que du point de vue du chercheur et même du statisticien public, cette enquête, même si on ne l'exploite plus, a vocation à resservir et à alimenter les réflexions dans dix ou quinze ans. Il y a de toute façon une contradiction de base entre la CNIL qui veut protéger les données et en limiter l'accès au maximum et les Archives nationales qui veulent tout garder, dans le détail le plus fin, pour l'histoire et la postérité. Et au milieu de ça, il y a nous ! Par le passé, avant le RGPD, on avait la vision suivante : « Je fais une enquête, je l'anonymise, je mets un fichier réputé anonymisé à disposition de la communauté scientifique, et voilà ». À partir de là le fichier a sa vie et il est inoffensif, au sens où il a été traité correctement. J'ai l'impression que les développements techniques récents, et le RGPD y contribue aussi, donnent le sentiment que désormais on peut tout apparier dans tous les sens, que plus rien n'est anonyme et que donc tout est personnel. C'est comme si on n'arrivait plus à se rassurer et à se dire : « J'ai réussi à faire un fichier qui est inoffensif, j'ai confiance, il ne peut plus faire de mal aux enquêtés ».

C. BEAUCHEMIN : Il faut clarifier ce que sont les données identifiantes. Il y a souvent, dans le débat public, une confusion entre, d'une part, les fichiers administratifs et, d'autre part, les données d'enquêtes. Les premiers ont une vocation de gestion administrative : ils sont donc nominatifs et identifiants et ils comportent l'intégralité des personnes qui relèvent du service administratif concerné (par exemple, tous les bénéficiaires de la sécurité sociale, tous les titulaires d'une carte de séjour, etc.). Les données d'enquête, elles, ne sont pas exhaustives : on tire des gens au hasard. Et une fois les données collectées, elles sont anonymisées : on supprime le nom, l'adresse, etc. Les fichiers d'enquête ne sont donc pas *directement* identifiants. Ils peuvent, en revanche, être *indirectement* identifiants, c'est-à-dire qu'en croisant beaucoup d'informations différentes, on peut possiblement réidentifier des personnes. Par exemple, imaginons un avocat qui vit dans un village, sans confrère dans le même village. S'il a répondu à une enquête et que les données conservent la profession détaillée et le nom du village, on pourra indirectement le réidentifier et savoir ce qu'il a

répondu aux autres questions. Cela dit, au contraire des données administratives, on ne peut pas chercher à retrouver n'importe quel individu dans des données d'enquêtes. Et c'est vraiment important d'établir cette distinction parce que la confusion est courante et même entretenue. Je me souviens par exemple de la polémique autour de « TeO ». SOS Racisme avait lancé une pétition contre une partie des questions de l'enquête, notamment sur la couleur de la peau, et la pétition s'intitulait « Fiche pas mon pote »¹⁰. On retrouvait cette idée que les fichiers administratifs permettent de retrouver les personnes et, dans des contextes autoritaires, de les persécuter. Or les enquêtes de statistiques sociales ne font pas du tout ça.

P. SIMON : Est-ce que le malentendu ne vient pas de ce qu'on définit comme étant des données ? J'ai pour ma part une position radicale sur la question des données sensibles : si on considère que les fichiers ont été anonymisés, et je pense qu'on a raison, alors il n'y a plus de données sensibles ! Après tout, le fait qu'il y ait certaines questions qu'on considère comme sensibles par rapport à d'autres est une question de définition relativement arbitraire.

C. BEAUCHEMIN : C'est une bonne façon de pointer les choses et effectivement ce n'est pas clair du tout. C'est-à-dire qu'il y a toujours cette confusion entre la question des données sensibles et la question des données anonymes. Je partage le point de vue de Patrick : dès lors que c'est anonyme, que tu n'as pas de quoi persécuter une personne, et que tu ne vises pas une personne en particulier, alors tu peux bien avoir des données dites *sensibles*, ce n'est pas un problème. En fait, la raison pour laquelle ça pose problème, alors même que c'est anonymisé, c'est qu'il y a cette idée que tu touches à quelque chose qui est trop « sensible », non pas au sens juridique, mais pour le débat public. Ce n'est pas une question de protection des données et des personnes. C'est un débat qui relève de la morale politique, sans toutefois être pointé comme tel.

P. SIMON : Prenons l'exemple des données sur les salaires ou les revenus : dans un contexte de conflictualité sociale assez forte, ces données peuvent être utilisées à charge contre les 1 % qui gagnent le plus d'argent en France et les condamner comme des profiteurs et exploités. C'est beaucoup plus sensible que la religion, et en tout cas plus politique ! Si tu veux produire un discours de critique et d'attaque contre un groupe de population, sur le revenu ou sur le patrimoine, ces données-là sont potentiellement très sensibles. Or, elles ne sont pas visées dans les listes de questions sensibles.

A. MORISSET : Il y a toutefois des seuils qui sont conseillés par le CNRS, comme le fait de ne pas mettre à disposition les modalités qui représenteraient moins de 5 % de réponses. Mais c'est vrai qu'il n'y a aucune interdiction sur le fait de diffuser ce genre de données. De toute façon, c'est un ajustement permanent. Il n'y a pas de règle miracle qui mettrait tout le monde à l'abri et qui conviendrait à tout le monde.

C. BEAUCHEMIN : C'est une somme de microdécisions dont le résultat dépend de rapports de force, qui varient d'une institution à l'autre.

Notes

1 Dans la continuité de l'enquête « Violences et rapports de genre » (Virage) menée dans l'Hexagone en 2015, « Virage dans les Outre-mer » a été réalisée par l'INED en 2018 dans les départements de la Réunion, la Guadeloupe et la Martinique.

2 En 2008-2009, l'INED a réalisé en collaboration avec l'Insee l'enquête « Trajectoires et origines » (TeO) auprès de 22 000 répondant-e-s. Elle visait à identifier l'impact des origines migratoires sur les conditions de vie et les trajectoires sociales. La seconde édition, « TeO 2 », est actuellement en cours de collecte, auprès de 26 500 répondant-e-s, de juillet 2019 à octobre 2020.

3 Sur l'histoire de la mise à disposition des données à l'INED, voir Arianna Caporali, Amandine Morisset et Stéphane Legleye, 2015, « La mise à disposition des enquêtes quantitatives en sciences sociales : l'exemple de l'INED », *Population*, vol. 70, n° 3, p. 567-597.

4 Règlement général sur la protection des données.

5 Nesstar est un logiciel créé en 2001 sur financement de la Commission européenne. Il est utilisé pour la documentation de données à destination du monde de la recherche.

6 Le projet « MAFE » est un projet de recherche collaboratif qui a débuté en 2005, avec pour objectif de fournir des données sur les migrations entre l'Afrique subsaharienne et l'Europe. Il porte sur trois pays européens et six pays africains.

7 Jocelyn Nappa, Bruno Schoumaker, Albert Phongi et Marie-Laurence Flahaux, 2019, « Difficultés économiques et transformation des unions à Kinshasa », *Population*, vol. 74, n° 3, p. 273-298.

8 Le réseau Quételet Progedo Diffusion regroupe divers partenaires : le service des enquêtes de l'INED, le CDSP (Centre de données socio-politiques, Sciences Po et CNRS), L'Adisp (Archives de données issues de la statistique publique) et le CASD (Centre d'accès sécurisé aux données).

9 Baptiste Coulmont et Patrick Simon, 2019, « Quels prénoms les immigrés donnent-ils à leurs enfants en France ? », *Population et sociétés*, n° 565, p. 1-4.

10 Voir [URL : https://www.liberation.fr/tribune/2007/11/15/enquete-sur-la-diversite-l-amalgame-pratique-par-sos-racisme_106289], consulté le 15 novembre 2019.

Pour citer cet article

Référence électronique

Cris Beauchemin, Gwennaëlle Brilhault, Amandine Morisset, Patrick Simon, Juliette Galonnier, Stefan Le Courant et Camille Noûs, « Le travail des données. Entretien autour du

service des enquêtes de l'INED », *Tracés. Revue de Sciences humaines* [En ligne], #19 | 2019, mis en ligne le 22 juillet 2020, consulté le 25 août 2020. URL : <http://journals.openedition.org/traces/10918>

Auteurs

[Cris Beauchemin](#)

démographe, chercheur à l'INED

[Gwennaëlle Brilhaut](#)

chefe du service des enquêtes à l'INED

[Amandine Morisset](#)

chargée de la documentation et de la mise à disposition des enquêtes de l'INED

[Patrick Simon](#)

socio-démographe chercheur à l'INED

[Juliette Galonnier](#)

sociologue, Sciences Po, CERI

Droits d'auteur



Tracés est mis à disposition selon les termes de la [Licence Creative Commons Attribution - Pas d'Utilisation Commerciale - Pas de Modification 4.0 International](#).