



HAL
open science

Archiver, documenter, enquêter sur l'enquête qualitative. Le travail de l'ombre de beQuali

Sarah Cadorel, Guillaume Garcia, Emilie Groshens, Emeline Juillard, Jeremie Vandebunder, Stefan Le Courant, Anthony Pecqueux, Camille Noûs

► To cite this version:

Sarah Cadorel, Guillaume Garcia, Emilie Groshens, Emeline Juillard, Jeremie Vandebunder, et al.. Archiver, documenter, enquêter sur l'enquête qualitative. Le travail de l'ombre de beQuali. Tracés : Revue de Sciences Humaines, 2020, HS 19, pp.183-198. 10.4000/traces.10968 . hal-03036951

HAL Id: hal-03036951

<https://hal.science/hal-03036951v1>

Submitted on 2 Dec 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.



Tracés. Revue de Sciences humaines

#19 | 2019

Les sciences humaines et sociales au travail (ii): Que faire des données de la recherche ?

Archiver, documenter, enquêter sur l'enquête qualitative. Le travail de l'ombre de beQuali

Archiving, documenting and investigating qualitative surveys. BeQuali's shadow work

Sarah Cadorel, Guillaume Garcia, Émilie Groshens, Émeline Juillard, Jérémie Vandebunder, Stefan Le Courant, Anthony Pecqueux et Camille Noûs



Édition électronique

URL : <http://journals.openedition.org/traces/10968>
ISSN : 1963-1812

Éditeur

ENS Éditions

Édition imprimée

Date de publication : 31 décembre 2019
Pagination : 183-198
ISBN : 979-10-362-0227-8
ISSN : 1763-0061

Ce document vous est offert par Centre national de la recherche scientifique (CNRS)



Référence électronique

Sarah Cadorel, Guillaume Garcia, Émilie Groshens, Émeline Juillard, Jérémie Vandebunder, Stefan Le Courant, Anthony Pecqueux et Camille Noûs, « Archiver, documenter, enquêter sur l'enquête qualitative. Le travail de l'ombre de beQuali », *Tracés. Revue de Sciences humaines* [En ligne], #19 | 2019, mis en ligne le 22 juillet 2020, consulté le 24 août 2020. URL : <http://journals.openedition.org/traces/10968>



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

Archiver, documenter, enquêter sur l'enquête qualitative Le travail de l'ombre de beQuali

ENTRETIEN AVEC SARAH CADOREL, GUILLAUME GARCIA,
ÉMILIE GROSHENS, ÉMELINE JUILLARD
ET JÉRÉMIE VANDENBUNDER
PROPOS RECUEILLIS ET PRÉSENTÉS PAR STEFAN LE COURANT,
ANTHONY PECQUEUX ET CAMILLE NOÛS

BeQuali¹ est à ce jour la première et seule banque d'enquêtes qualitatives en sciences humaines et sociales en France, au sein du Centre de données socio-politiques (CDSP, UMS Sciences Po-CNRS)². Dans l'esprit du cycle de hors-séries de *Tracés* consacré aux sciences humaines et sociales au travail, qui souhaite notamment explorer la diversité des métiers qui concourent à faire les sciences humaines et sociales aujourd'hui et la façon dont ils interagissent, la rencontre avec l'équipe de beQuali est particulièrement riche puisqu'elle met autour de la table une archiviste (Sarah Cadorel), deux documentalistes (Émilie Groshens et Émeline Juillard) et deux docteurs en sciences sociales occupant un poste d'ingénieur de recherche (Guillaume Garcia et Jérémie Vandebunder). Leur rôle est en quelque sorte d'« apporte[r] de la donnée aux données existantes », comme le dit Émilie Groshens ci-dessous : c'est-à-dire de se saisir d'enquêtes déjà réalisées, de les documenter, les archiver, enquêter sur elles et les rendre accessibles aux personnes porteuses d'un projet d'enseignement ou de recherche qui souhaitent les consulter.

Une des particularités de ce projet est de prendre pour unité l'enquête dans sa globalité, c'est-à-dire selon une échelle bien plus large que lorsque l'on s'intéresse aux données en prenant pour seule focale d'entrée l'article ou la publication scientifique. Ainsi, les éléments de l'enquête que prend en compte beQuali ne se limitent pas aux entretiens et aux observations dans un carnet : par exemple, un document annoté par un chercheur ou une chercheuse est tout aussi pertinent. Chaque enquête fait l'objet d'un travail minutieux pendant une année entière, de la part d'un binôme composé d'un part d'un-e documentaliste ou d'un-e archiviste, et d'autre part d'un-e ingénieur-e de

1 Deux autres membres de l'équipe n'ont pu être présentes lors de l'entretien : Selma Bendjaballah et Émilie Fromont.

2 En décembre 2019, 14 enquêtes qualitatives en sociologie et science politique ont déjà été mises à disposition sur le site bequali.fr. Les thèmes principaux couvrent les comportements politiques, l'éducation, les élu-e-s, la famille, le genre, l'immigration, les institutions politiques, les professions, la sexualité, le territoire. Pour chaque enquête, le volume d'archives représente un corpus de 70 à 500 documents, comportant des transcriptions d'entretiens, des vidéos de *focus groups*, des photographies, des notes de terrain, de la documentation, des rédactions intermédiaires, des documents méthodologiques et de préparation du terrain, etc.

recherche. Au bout de ce processus, l'enquête est ouverte à toutes les réutilisations possibles (avec cette seule condition d'avoir un projet de recherche ou d'enseignement). Le projet scientifique de beQuali s'est affiné au fil du temps et s'inscrit précisément dans cette diversité : alimenter un panorama des méthodologies et épistémologies des enquêtes qualitatives en sciences humaines et sociales.

TRACÉS : *Pour commencer, pourriez-vous présenter comment est né beQuali ?*

GUILLAUME GARCIA : C'est un projet qui est développé depuis 2010 au CDSJ, avec plusieurs soutiens financiers : la FNSP (Fondation nationale des sciences politiques), le CNRS, le ministère de l'Enseignement supérieur, de la Recherche et de l'Innovation (Mesri), et le soutien de l'ANR à 2 moments : le projet blanc Réanalyse en 2010 qui a permis de mettre en place le prototype de ce qu'est devenu beQuali, et l'attribution de l'Equipex (projet DIME-SHS, financé par l'ANR) qui a permis de monter la partie « équipement » (Duchesne et Garcia, 2014). C'est à ce moment-là que le projet a vraiment démarré en permettant de recruter des personnes, en parallèle de l'attribution de postes pérennes par le CNRS et le Mesri. En 2013, la production du site Internet de mise à disposition des enquêtes a constitué un autre moment charnière. L'année 2019 a été un peu faste : on est 7 mais avec la fin de l'Equipex fin 2019, mécaniquement on va redescendre à 4 postes³. Toutes ces ressources nous ont permis de dégager plus de temps pour la valorisation, des communications, des formations, de l'enseignement, des productions d'outils (Bendjaballah *et al.*, 2017).

TRACÉS : *Pourriez-vous préciser le contexte dans lequel s'inscrit un tel projet ?*

G. GARCIA : Au début le projet beQuali a été inspiré par la banque anglaise Qualidata⁴, sauf qu'on intervient dans un contexte très différent. Contrairement à la Grande-Bretagne, en France il n'y a pas d'obligation de dépôt. Il n'y a pas non plus de politique d'inventaire national des enquêtes qui ont été réalisées ou archivées, comme cela peut exister en Suisse par exemple. Il a fallu créer tout ce système d'offre, il a fallu aller à la pêche aux chercheurs qui voudraient bien partager leurs enquêtes et identifier les enquêtes qui

3 Outre les membres de son comité scientifique et technique, ainsi que des développeurs informatiques (du CDSJ, du Médialab ou de la DRIS de Sciences Po), d'autres membres de l'équipe ont par le passé contribué au développement de beQuali, notamment Anne Both, Virginie Descoutures, Sophie Duchesne, Thibaut Rioufrey et Karim Souanef.

4 Depuis 2012, United Kingdom Data Archive est devenu UK Data Service ; la banque Qualidata y a été intégrée. Seul subsiste un catalogue spécifique aux enquêtes qualitatives, *Qualibank*, qui correspond uniquement aux enquêtes patrimoniales mises à disposition à l'origine par Qualidata : [URL : <https://discover.ukdataservice.ac.uk/QualiBank#close>], consulté le 22 novembre 2019.

étaient disponibles. Au Royaume-Uni, avec l'obligation de dépôt, Quali-data doit faire face à une masse d'informations et traite les enquêtes beaucoup moins finement que ce que l'on fait. Le deuxième aspect propre au Royaume-Uni c'est qu'il y a eu très tôt une incitation institutionnelle forte à la réutilisation de données. L'ESRC (Economic and Social Research Council) a mis en place des financements assez massifs pour inciter les chercheurs à mener des analyses secondaires ou réanalyses. Des financements importants ont permis que les chercheurs s'emparent des enquêtes du catalogue⁵. Ce qui a créé toute une littérature et un débat méthodologique sur l'analyse secondaire. Ce n'est pas du tout le cas en France. Nous, on est intervenu à un moment où il fallait faire un peu la quadrature du cercle, c'est-à-dire créer et faire tourner l'équipement et en même temps essayer de créer une offre et inciter à la demande. Le fait que les communautés disciplinaires ne soient pas très impliquées, que les institutions financent peu ce type d'initiatives sont des limites du contexte dans lequel on intervient. On construit un bien public qui doit aussi être approprié par les communautés.

TRACÉS : Dans la mesure où vous êtes pionniers en la matière, vous avez aussi développé une offre de formation. Comment vous situez-vous aujourd'hui à l'heure où de nombreux établissements de recherche mettent en place leurs propres plateformes d'archivage ?

JÉRÉMIE VANDENBUNDER : Notre communication est à destination des laboratoires de sciences humaines et sociales et des institutions. Donc, plutôt pour des chercheurs, des enseignants-chercheurs et des étudiants. Ensuite, on réalise aussi des formations à destination des ingénieurs et techniciens et plus récemment beQuali est assez présent dans les plateformes universitaires de données (PUD), qui proposent des services et un accompagnement pour tout ce qui concerne les données d'un établissement. Ce sont des choses qui sont en train d'être créées.

SARAH CADOREL : C'est sûr que le marché est devenu un peu concurrentiel. Mais nous, on a un parti pris scientifique qui est de reconstituer un objet qui est l'enquête. Pour cela on va aller chercher des documents qui relèvent de plusieurs fonds, par exemple si on prend une enquête qui a été menée par un chercheur de Sciences Po et un autre de l'EHESS, on va prendre les archives dans les deux fonds, on va créer un artefact qu'on met en ligne

5 Dès 1996, la politique de financement sur projet de l'ESRC incitait fortement les chercheurs à réutiliser des données préexistantes. Depuis 2012, un vaste programme de bourses vise spécifiquement des projets d'analyse secondaire.

et ensuite les archives sont rendues à chaque institution. L'objet que l'on crée n'est pas exactement celui qui va être dans les PUD. On peut aussi aller chercher des archives de laboratoire. On va faire tout un travail de recontextualisation assez approfondi, qui n'est pas du tout ce que va faire un entrepôt, lequel va plutôt se concentrer sur l'archivage et le stockage assez systématique des données et des métadonnées dans l'état. Nous faisons un gros travail d'enrichissement et d'archivage pérenne. Cela limite aussi le volume de données que l'on peut traiter. Mais on apporte une solution différente, car pour le moment ces PUD sont essentiellement des endroits où des membres d'un même projet de recherche vont pouvoir partager des documents, il n'y a pas de véritable mise à disposition. Nous, on ne fournit pas un serveur partagé, ce n'est pas juste une Dropbox. Aujourd'hui, le contexte est plus favorable car les gens ont déjà une idée de ce dont on va leur parler quand on vient présenter beQuali, ce qui n'était pas le cas au début. On se souvient des premières visites de labo qu'on faisait : « Ah ! non, mais moi je ne donne pas mes données à Sciences Po ! »

TRACÉS : Justement, comment se passe, pour un chercheur ou une chercheuse, une procédure de dépôt ?

ÉMELINE JUILLARD : C'est à leur initiative : ils répondent aux appels à enquête que l'on diffuse, jusqu'à présent une fois par an, sur la plupart des listes professionnelles. Le formulaire décrit le corpus concerné, les méthodes utilisées, les possibilités de réutilisation de l'enquête à des fins d'enseignement, de recherche, etc. Les propositions de dépôts sont évaluées par un comité scientifique et technique composé d'une dizaine de membres représentant les métiers que l'on cherche à faire dialoguer depuis le début dans le cadre du projet beQuali : des documentalistes, des archivistes, des professionnels des humanités numériques et des (enseignants-)chercheurs et (enseignantes-)chercheuses en sciences sociales. Nous, on donne juste notre avis sur la faisabilité technique.

TRACÉS : Comment se construit la discussion entre les différents métiers sur ce qui vaut la peine d'être archivé ?

S. CADOREL : Il ne s'agit pas d'évaluer la qualité scientifique intrinsèque de la recherche mais plutôt de voir si ça crée un pôle thématique avec d'autres enquêtes de beQuali.

ÉMILIE GROSHENS : Par contre, un critère indispensable, c'est qu'il y ait déjà des publications scientifiques qui attestent de la qualité scientifique. Donc

le comité va plutôt se prononcer autour du potentiel de réutilisation : est-ce qu'il y a assez de matériaux ? est-ce que le corpus est complet ? On ne prend que des enquêtes qui sont terminées, on fait attention à ça dans nos critères de sélection. On vérifie que la personne a plus ou moins fini de travailler sur cette enquête. La question de savoir si l'exploitation des données a été faite par le chercheur avant qu'on ne la mette à disposition d'autres chercheurs est importante dans les discussions.

G. GARCIA : Un des critères importants aussi, c'est de représenter la diversité des méthodes d'enquêtes, ou des écoles théoriques, ou des postures d'enquête.

TRACÉS : *Pouvez-vous tracer une typologie des gens qui déposent chez vous ? Y a-t-il des profils particuliers, des moments particuliers dans la carrière ?*

É. GROSHENS : On a du succès auprès des retraités, indubitablement. Ils ont fini, ils se disent : « J'ai tout ça, je ne sais plus quoi en faire » et puis il y en a qui sont contents de léguer, de transmettre. Après on a aussi, dans les propositions, de jeunes doctorants. Ils n'ont pas forcément de publications mais ils sont dans la dynamique générale de : « On nous dit qu'il faut mettre à disposition des données ». On n'a pas besoin de les convaincre que c'est bien de mutualiser et de partager ses données. Ils sont déjà dans cet état d'esprit.

J. VANDENBUNDER : Je rajouterai un profil de déposant possible : les collectifs de chercheurs parce que là l'enquête est beaucoup plus facilement identifiable, elle est financée, bien bornée. Il y a des entretiens, des données, des comptes rendus de réunion, etc. C'est aussi un profil qui peut être intéressé par le dépôt.

G. GARCIA : Je pense qu'il y a une diversité de cas de figure possibles : ceux qui sont à la retraite et qui veulent léguer, d'autres qui sont dans le partage, d'autres qui sont un peu dans la valorisation du travail. La défense de son travail est un enjeu professionnel très saillant actuellement et les retraités sont ceux qui ont le moins de risques à prendre, le moins de choses à perdre, qui ont le temps, qui ont été cobayés aussi. Pour amorcer le processus, il faut se lancer et ce n'est pas évident.

TRACÉS : *Si on reprend le fil du dépôt, une fois que l'enquête est sélectionnée, que se passe-t-il ?*

É. GROSHENS : On fait une première visite sur le lieu où les archives sont conservées, on organise le transfert provisoire des documents ici au CDSP

(les archives sont ensuite restituées). Puis commence le protocole de traitement de l'enquête tout au long duquel le chercheur ou la chercheuse qui dépose son enquête est autant que possible sollicité-e pour définir et sélectionner le corpus que l'on va numériser et mettre en ligne ; et pour documenter l'enquête par ce qu'on appelle *l'enquête sur l'enquête*. Parfois il s'agit seulement de questions ponctuelles sur les métadonnées comme donner un titre à un document, car on peut trouver une feuille volante sur laquelle le chercheur n'a pas forcément inscrit un titre, une date, un auteur...

J. VANDENBUNDER : « L'enquête sur l'enquête » est un document produit pour chaque enquête qui part de l'étude du corpus, des documents, des publications mais aussi des entretiens réalisés avec les chercheurs. On revient sur la genèse de l'enquête, comment elle s'inscrit dans leur carrière, leur cadre théorique, méthodologique, le contexte scientifique, la réalisation du terrain. On présente le corpus, le mode d'anonymisation, ce que le chercheur a perdu, etc. Et en dernière partie, on présente les pistes d'utilisation possibles identifiées par le chercheur et par nous. Mais le centre de l'enquête sur l'enquête, c'est l'entretien approfondi avec les chercheurs.

É. GROSHENS : Pour nous le rêve, ce serait de collecter du numérique anonymisé par le chercheur, ou avec les consentements des enquêtés. Mais anonymiser ce n'est pas juste changer les prénoms ; le risque d'identification des individus se cache partout : dans les métadonnées d'un document Word, dans les notes de bas de page, etc. On fait la différence entre les identifiants directs (noms, coordonnées, tout ce qui est lié à l'état civil), et les identifiants indirects qu'il faut aller fouiller dans les textes. On est obligé de tout lire, ça peut être plein de choses : l'entreprise où le beau-père a travaillé, etc... Et pour y arriver on travaille avec le chercheur car l'intérêt de beQuali est que l'enquête soit réutilisable et réutilisée donc si on masque tout, on perd aussi la richesse des données. On essaye donc de trouver un équilibre entre protéger les enquêtés et donner de la matière à réutiliser. On identifie avec le chercheur le niveau de risque de son enquête (si les données sont plus ou moins sensibles) et ensuite on établit avec lui un protocole. Il faut aussi protéger le chercheur ou la chercheuse car il met parfois des notes personnelles. Quand on ne met pas à disposition un document, on signale qu'on enlève quelque chose. C'est le même principe pour l'anonymisation : de manière générale, on ne remplace pas par des pseudonymes, on enlève l'information et on signale qu'on a enlevé l'information (Bendjaballah *et al.*, 2018).

TRACÉS : *Cela nous fait accéder à une dimension qui intéresse beaucoup le cycle « Les sciences humaines et sociales au travail » de notre revue : celle de votre travail concret. Pour traiter chaque enquête, vous formez donc un binôme composé d'une personne plutôt praticienne de l'enquête et d'une autre plutôt axée sur la documentation et l'archivistique : quelles questions se posent et comment se règlent-elles au sein du binôme ?*

É. GROSHENS : Je dirais que nous, les archivistes et documentalistes, on a tendance à vouloir garder des documents pour le côté patrimonial et peut-être que vous, les sociologues, vous avez tendance à dire : « Mais qu'est-ce que tu veux faire de ça ? » Vous avez plus tendance à vous mettre dans la peau du réutilisateur et dire qu'il faut laisser tomber.

S. CADOREL : Après il y a aussi le fait que nous, on ne fait pas de la collecte d'archives pour créer des fonds d'archives. On sélectionne des parties de fonds d'archive pour créer des enquêtes. Ce que l'on ne prend pas, cela ne veut pas dire qu'on le détruit – de toute façon, on n'a pas le droit de détruire parce qu'on n'est pas propriétaires des fonds –, c'est juste que l'on ne le prend pas pour beQuali.

G. GARCIA : À mon avis, il y a eu des moments forts dans l'histoire du projet où il y a eu un certain nombre de questions qui ont été réglées de manière plus ou moins collective, plus ou moins entre les personnes du métier, en particulier sur le plan de classement par exemple (Both et Garcia, 2014 ; Both et Cadorel, 2015). La logique archivistique revêt une importance indéniable pour beQuali, et chacun est plus ou moins compétent pour gérer ces choses-là. Mais j'ai l'impression que cet aspect a été amélioré, ajusté au fil de l'eau. Au sujet de l'enquête sur l'enquête, plusieurs personnes se sont succédé sur des postes à durée déterminée et ont rempli cette fonction d'ingénieur en sciences sociales : là aussi on a eu beaucoup de questions, enquêtes après enquêtes, pour savoir quelles sont les limites du modèle que l'on a mis en place au début. Et finalement on a été amené à le modifier et l'élargir, à ajuster la posture. Il y a beaucoup de choses qui se décident aussi dans les binômes. Quand on traite une enquête, les questions qui ne sont pas réglées en binôme, sont tranchées en réunion d'équipe.

É. GROSHENS : Les controverses qu'on a eues se sont aussi atténuées avec la diversification des usages. Au départ on était axé sur un type, la réanalyse, donc il y avait une idée très précise de ce que l'on devait faire avec les matériaux. Petit à petit on a élargi les usages et on s'est dit : « Tous ces matériaux, ça peut servir et ce sont les utilisateurs qui verront pour quoi ». Les

documents sont liés entre eux donc on les met, peu importe ce que les gens veulent ou vont faire avec.

G. GARCIA : Et quand on ne sait pas décider au sein d'un binôme, on va voir le chercheur. Et si le chercheur ne sait pas non plus, la décision finit par s'imposer en ayant un peu réfléchi⁶. Il y a souvent beaucoup d'échanges. Pour l'enquête « Dilapidation et prodigalité », d'Anne Gotman⁷, on a eu plein d'échanges où on se torturait l'esprit.

É. JUILLARD : En effet, pour l'anonymisation, Anne Gotman proposait de modifier les informations, comme ça peut se faire dans les publications. Elle était partie dans une logique de remplacer une ville par une autre qui lui ressemblait, un métier par quelque chose qui lui ressemblait. Nous on voulait remplacer par une définition un peu générale : « ville moyenne de l'est de la France ».

G. GARCIA : On doit trouver un équilibre entre l'anonymisation et la protection de la richesse des données et à chaque fois c'est à remettre en place, à actualiser en situation. Et je sais que pour cette enquête-là, plein de fois, on basculait d'un côté ou de l'autre. Moi je disais : « Ça, il faut l'anonymiser », puis en discutant dans le binôme ou l'équipe, on se disait : « Si on fait ça c'est absurde : personne ne pourra jamais retrouver de qui on parle parce que c'est vieux mais par contre si on enlève cette information-là, c'est dommage ». Il y a des inventions, par exemple les noms des catégories du plan de classement, qu'il faut discuter en collectif sinon, pour le reste, ça peut se faire en binôme. Suivant nos expertises on se sent plus ou moins compétent pour trancher tel ou tel point.

S. CADOREL : Au sein des binômes on a un peu une majeure et une mineure ; même si ça se chevauche en partie en pratique, l'un prend plutôt en charge un aspect métier et l'autre membre du binôme un autre aspect métier.

É. JUILLARD : Quand on demande à notre prestataire de numériser un document, on peut lui demander de nous fournir un document « .odt », comme ça, on a un document modifiable et éditable. Ce qui peut faciliter des choses pour anonymiser mais ce qui peut rendre plus difficile la vision

6 Les membres du comité scientifique et technique de beQuali ont également apporté leur contribution au règlement d'un certain nombre de dilemmes professionnels. Sur la composition actuelle du comité, voir [URL : <https://bequali.fr/fr/propos/>], consulté le 20 novembre 2019.

7 Voir [URL : https://bequali.fr/fr/enquetes/lenquete-en-bref/cdsp_bq_s6/], consulté le 6 novembre 2019.

d'un document où il y avait partout en marge des annotations, des traces de ce qu'a fait le chercheur pour analyser son document. Ça, c'est un sujet de discussion qui se pose fréquemment. Est-ce qu'on choisit de mettre en ligne un document qui rend compte de ces annotations en marge ou est-ce qu'on préfère un document modifiable quitte à faire disparaître quelques annotations quand elles sont plus rares ?

S. CADOREL : Je me suis fait la réflexion suivante quand on construisait les fonctionnalités du site : Émilie et Émeline sont documentalistes tandis que moi, je suis archiviste. On n'a pas la même manière de voir les choses. Je vois les choses de manière plus figée, arborescente – des boîtes dans des boîtes – et elles vont plutôt avoir des réflexes de transversalité dans l'exploration des documents. Il s'agit davantage de manières d'approcher les choses que d'avis différents.

É. GROSHENS : On n'en a pas parlé tout à l'heure quand on discutait des controverses mais ça me fait penser que lorsque je suis arrivée dans l'équipe, il fallait anonymiser l'enquête d'Agnès van Zanten⁸, un truc assez monstrueux avec pleins de documents, plusieurs enquêteurs, etc. On avait des post-doctorant-e-s à l'époque dans le projet qui travaillaient sur l'analyse secondaire. Moi j'étais là à mettre mes Post-it sur les documents pour dissimuler des informations et une post-doc est entrée et a dit : « Mais qu'est-ce que tu fais Émilie ? » C'était un sacrilège, ça inquiétait que je masque des informations. Cette question de l'anonymisation était un peu sacralisée. Elle devait se dire : « Tu es documentaliste, tu viens d'arriver, est-ce que tu sais bien ce que tu fais ? C'est compliqué, c'est une question scientifique » [rires].

J. VANDENBUNDER : Là où il y a une discussion intéressante dans le binôme, qui pour moi fonctionne bien, c'est quand on cherche à quel moment arrêter le travail, notamment dans la production de l'enquête sur l'enquête. On ne doit pas se substituer aux chercheurs qui vont analyser les données. On n'est pas là pour faire une socio-analyse. L'enquête sur l'enquête contextualise les données, elle n'est pas là pour analyser l'enquête. Déjà on n'est pas là pour faire un travail d'analyse et on n'est pas là non plus pour faire une histoire de la sociologie. J'ai fait de la recherche par ailleurs et j'associe ce que je fais aujourd'hui à une activité de recherche mais je pourrais m'enflammer et aller beaucoup plus loin que ce qu'on fait dans l'enquête sur l'enquête !

8 Voir [URL : https://bequali.fr/fr/enquetes/lenquete-en-bref/cdsp_bequali_sl/], consulté le 6 novembre 2019.

É. GROSHENS : Moi je regarde un document, je me dis : « On le garde ou on le jette ? » et vous, les sociologues, vous savez plus que nous, les documentalistes-archivistes, la pertinence de le conserver ou pas. Après il y a le chercheur ou la chercheuse qui a mené l'enquête qui va donner son avis, et ensuite il y a l'utilisateur et là on sait moins si tout ce travail que l'on fait lui sert ou s'il va juste télécharger les fichiers, prendre les entretiens et les mettre dans des logiciels.

TRACÉS : Toujours du point de vue de votre travail concret, avec quels outils travaillez-vous au quotidien ?

J. VANDENBUNDER : Globalement, on s'est doté d'outils qui sont assez commodes et assez souples, mais il faut quand même au minimum un an pour traiter entièrement une enquête. Par exemple on a le plan de classement : c'est une arborescence qui permet de classer les différents documents, mais il peut s'adapter pour chaque enquête, ce n'est pas un moule. En fait, on essaye de choisir ce qui permet la meilleure lecture pour l'utilisateur.

É. GROSHENS : Il faudrait mentionner un métier qui reste relativement invisible, ce sont les développeurs qui ont fait le site. Ils ne participent plus aux réunions. Il y a eu plein de profils divers qui ont travaillé avec nous ; à l'origine, sur la banque d'enquête, c'était les collègues du Médialab : Comment coconstruire ensemble ? Que peuvent apporter les métiers du numérique à beQuali pour ce qui concerne l'exploration et la visualisation des données ? Ça a été très utile et très riche pour construire la base ; mais aujourd'hui il s'agit davantage de maintenance.

TRACÉS : Vous évoquez la question des usages, connaissez-vous les usages qui sont faits des enquêtes déposées ? Qui les consulte ?

É. GROSHENS : Pour pouvoir y accéder, il faut avoir un projet de recherche ou d'enseignement. On veut être sûr que derrière c'est sérieux. On peut avoir des étudiant-e-s, doctorant-e-s et post-doctorant-e-s, des chercheurs et chercheuses et des enseignants-chercheurs et enseignantes-chercheuses. Côté enseignement, c'est surtout des cours de méthode qualitative en licence et master de sociologie du droit, de la famille, en sciences politiques, en sciences de l'éducation et j'en oublie.

J. VANDENBUNDER : Pour parler des usages, c'est vrai que l'enseignement permet aux étudiant-e-s de travailler directement avec les données : de pouvoir traiter les entretiens, faire des fiches à partir des entretiens. Dans ce

domaine, je crois que l'enquête de Bozon-Héran⁹ marche bien parce qu'elle a une valeur patrimoniale : c'est une enquête très connue et un sujet très intéressant pour les étudiants. On a eu un retour de Christophe Giraud qui est professeur de sociologie de la famille à Paris 5 et on l'a publié sur le site de beQuali¹⁰ : pour un TD, les étudiant-e-s ont eu des fiches à faire à partir d'un entretien qu'ils avaient lu entièrement et ensuite ils en discutaient collectivement. Les étudiant-e-s peuvent manipuler les données et en plus ça leur permet de découvrir un autre contexte : il y a un décalage certain entre la formation des couples dans les années 1980 et la formation des couples aujourd'hui. L'année dernière on a mis au point un filtre d'exploration spécialement pour les enseignant-e-s. On a croisé toutes nos enquêtes et des thématiques d'enseignement : il suffit de choisir la thématique, par exemple « faire des entretiens semi-directifs ou directifs » ou « guide d'entretien », et les résultats tombent : dans telle enquête il y a un guide d'entretien, dans une autre des discussions autour des guides d'entretien, etc.

TRACÉS : *Est-ce qu'il y a eu des usages d'enquête qui vous ont totalement surpris ?*

G. GARCIA : Il y a un peu plus d'un an, un chercheur a pris une des premières enquêtes que l'on a traitées, une enquête d'Etienne Schweisguth intitulée « Les Français et la politique »¹¹, sur la prégnance de l'idéologie chez les citoyens ordinaires (notamment sur comment ils se situent par rapport au clivage gauche-droite). Ce chercheur, Dominique Labbé, politiste avec une orientation plutôt linguistique s'est servi des entretiens dans une perspective des sciences du langage. Il travaille beaucoup sur les outils lexicaux de quantification du langage. Il était à la recherche de matériaux qu'il avait du mal à trouver par ailleurs. Il expliquait que ce n'était pas facile pour lui de trouver du français parlé dans un cadre conversationnel. Ce qui veut dire que les usages ne sont pas seulement disciplinaires : les matériaux peuvent parler de d'autres disciplines.

É. GROSHENS : Ce que l'on peut dire aussi, c'est qu'au départ on était parti sur un usage très « réanalyse » du fait du montage de projet et que petit à petit on a essayé d'être à l'écoute de nos utilisateurs et on a vu que ce n'était pas forcément l'usage qui prédomine et que les gens ont été intéressés

9 Voir [URL : https://bequali.fr/fr/enquetes/lenquete-en-bref/cdsp_bq_s2/], consulté le 6 novembre 2019.

10 Voir [URL : https://bequali.fr/media/ckeditor/uploads/2019/02/13/colloque_dime_pres_cgiraud_gaj68zk.pdf], consulté le 15 novembre 2019.

11 Voir [URL : https://bequali.fr/fr/enquetes/lenquete-en-bref/cdsp_bequali_sp2/], consulté le 6 novembre 2019.

pour compléter des corpus, pour voir une méthode en pratique, pour faire d'autres choses. On a eu des premières demandes de doctorant-e-s qui voulaient avoir accès à des enquêtes, dans une démarche d'apprentissage. Petit à petit on s'est rendu compte que c'était aussi utilisé pour l'enseignement.

TRACÉS : *Il n'y a pas encore d'enquête au sens de vérification que l'enquêteur était un bon enquêteur ?*

J. VANDENBUNDER : Non et ça peut être un intérêt de beQuali. Les enquêtes sont parfois enjolivées dans les encarts méthodologiques mais là il n'y a pas la possibilité de bluffer. Justement ça permet de montrer que, dans une enquête, il y a parfois des échecs. Il n'y a pas une manière de faire une enquête. Au niveau pédagogique c'est superintéressant de montrer que parfois les enquêteurs se retrouvent devant des portes fermées...

G. GARCIA : Dans le cas anglais, qui est la référence, les cas de réutilisation des enquêtes pour vérification sont vraiment marginaux.

J. VANDENBUNDER : C'est vrai que pour un chercheur ça peut représenter un risque de mettre à disposition les données. Il peut se dire qu'il prête le flanc à des critiques nouvelles qui n'existaient pas avant parce que ça restait dans des boîtes noires. Mais ça peut être une motivation de dépôt de justement pouvoir appuyer ses publications sur les données, en disant : « Voilà sur quoi s'appuient mes analyses ». Comme une manière de dire : « Voilà je montre mes données, je suis transparent ». Peut-être qu'un jour on demandera aux doctorant-e-s de fournir leurs données. Ce qui ne pose aucune question dans les sciences dures deviendra aussi obligatoire dans les sciences sociales. Andrew Abbott a sorti un article en 2007 qui dit qu'un jour ou l'autre pour publier dans *American Journal of Sociology*, il faudra fournir les données avec l'article. Ça fera moins d'articles à traiter : tout le monde ne le fera pas et ça donnera des articles qui mettront beaucoup plus de temps à être évalués mais qui seront meilleurs.

TRACÉS : *Cette injonction à fournir ses données pose la question de la normalisation de la science, notamment de la normalisation des méthodes privilégiant les entretiens et les consultations d'archives au détriment des observations. Quel peut être le statut des observations dans un tel processus de publication ?*

S. CADOREL : Je trouve que ça pose une vraie question mais c'est toute la politique ministérielle de la science ouverte qui pose cette question. Plutôt que de savoir comment les données sont produites se pose la ques-

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

J. VANDENBUNDER : La normalisation, c'est un risque mais ce qu'essaie de montrer beQuali c'est qu'il n'y a pas une seule façon de faire de la recherche et qu'il y a une diversité complète des méthodes, des points de vue et des cadres théoriques.

TRACÉS : *Traiter une enquête vous prend une année, mais quand les enquêtes sont utilisées, comment sont-elles citées? Comment est-ce que votre travail est reconnu?*

É. GROSHENS : Il y a un modèle, un format de citation de l'enquête en fait. Mais c'est difficile à tracer, c'est pour ça qu'on fait des enquêtes auprès des utilisateurs de manière récurrente en leur rappelant : « N'oubliez pas de nous faire part de ce que vous faites des données ». Quand c'est de l'enseignement c'est facile de recontacter les professeurs mais parfois quand c'est pour un projet de recherche et que c'est pris dans une masse de données quantitatives... On ne sait pas si ça va être utilisé dans la publication et donc on ne peut pas savoir ce qui a été publié avec ces données-là; malheureusement c'est aussi un indicateur que nous devons fournir à l'ANR, pour justifier que l'outil sert.

TRACÉS : *On imagine que l'on va citer l'enquête mais pas l'année que vous avez passée, vous, à la rendre accessible et utilisable par d'autres.*

É. GROSHENS : On est les travailleurs de l'ombre!

É. JUILLARD : Normalement si vous utilisez les données, le nom de la banque d'enquête et du CDSP figurent dans la citation. On propose un modèle de citation qui mentionne les producteurs, le titre de l'enquête, le diffuseur (beQuali-CDSP), l'année de publication du jeu de données et le DOI qui est l'identifiant unique et pérenne d'accès au jeu de données.

G. GARCIA : Je voudrais insister sur ce point important. Dans des centres plus importants comme le UK Data Archive, ils insistent sur la nécessité de citer à la fois le chercheur premier et les centres de données, les médiateurs. Cela montre bien qu'ils ont du mal à faire accepter cette pratique comme normale, c'est-à-dire la citation du jeu de données comme quelque chose d'important lorsqu'on produit une nouvelle publication. Ils insistent beaucoup, ça veut dire qu'il y a des gens qui ne le font pas. UK Data Service a du mal à recenser le nombre exact de publications qui sont faites à partir des données qu'ils diffusent parce qu'il y a toujours un volet de gens qui ne vont pas y penser.

TRACÉS : *La discussion de tout à l'heure sur la forme du document numérisé pose la question de savoir ce qu'est une donnée. Comment distinguer un griffonnage sur le bord d'un document de ce qui vaut la peine d'être conservé?*

S. CADOREL : La définition des données est un vaste débat, il y a de nombreuses définitions. L'OCDE les définit comme un enregistrement lié à une activité de recherche. Mais c'est compliqué, c'est trop polysémique. Une note de synthèse est-ce que c'est une donnée, sachant que c'est aussi l'enregistrement d'une activité de recherche? Pour nous il y a un noyau dur des données dans une enquête qui sont les observations, les entretiens, les matériaux de terrain écrits et après, ce qui est autour et qu'on a du mal à nommer : des métadonnées, des données périphériques? On rentre dans des questions de vocabulaire qui sont compliquées. Pour certaines personnes, les données de l'enquête ce sont les transcriptions d'entretiens ou les carnets de terrains. Nous, on va prendre tous les documents qui attestent de la préparation du terrain et en aval, tout ce qui relève de l'analyse et a mené à des publications ou des valorisations, quelles qu'elles soient, de l'enquête.

TRACÉS : *En fait, votre unité, c'est l'enquête.*

É. GROSHENS : Sachant que ce n'est pas si facile que ça, que les contours varient d'une enquête à l'autre.

J. VANDENBUNDER : L'idée c'est que l'enquête ne se limite pas aux données.

G. GARCIA : Tout dépend aussi de la réutilisation. Si on est dans la réanalyse ou l'analyse secondaire on va avoir une définition plus restrictive – autour des matériaux et des documents d'analyse – de ce qui constitue les données du point de vue de ceux qui vont réutiliser l'enquête que si on fait de l'his-

toire des sciences ou de l'enseignement, où là finalement tout peut devenir une donnée (donc, aussi, les élaborations théoriques du projet de recherche, la correspondance du ou des chercheurs, etc.). Sachant que dans le plan de classement, on a évité cette question piège. Nulle part il n'y a le terme *données*. Par exemple, pour la transcription d'entretien, on applique plusieurs catégories dans le plan de classement. C'est une transcription d'entretien, ça, c'est sûr, mais pour la question de savoir si c'est une donnée, c'est le réutilisateur qui va trancher. Bien sûr, il ne faut pas se leurrer : quand on va voir le chercheur premier et qu'il nous explique comment il a travaillé, quels étaient les matériaux qu'il a collectés sur le terrain, quelque part il y a une notion assez évidente de ce que sont les données au sens où la plupart des gens pourraient l'entendre. Mais on a un peu réglé cette difficulté en évacuant le terme!

É. GROSHENS : Mais je trouve que c'est intéressant parce qu'on pourrait dire qu'à beQuali, on apporte de la donnée aux données existantes. On va ajouter au corpus des choses qu'on produit là maintenant. Par exemple : des documents qui vont aider à établir les liens entre les pseudonymes de la publication et les pseudonymes du corpus des données. Des choses que le chercheur ou la chercheuse n'a pas produites à l'époque mais qui nous semblent nécessaires pour utiliser les données. Et au-delà de ça, quand on anonymise, si on remplace par un hyperonyme (par exemple : *grande ville de l'ouest de la France* à la place de *Rennes*), on peut dire qu'on supprime de la donnée et qu'on en ajoute...

S. CADOREL : C'est aussi pour ça qu'on dit *banque d'enquêtes* – et pas de *données* – qualitatives.

TRACÉS : *Finalement, c'est l'utilisateur qui définit la donnée.*

G. GARCIA : Oui, même si à la base il y a une collecte sur le terrain qui correspond à la définition assez commune qu'on a d'une donnée : on fait du terrain et on collecte des choses et c'est ce sur quoi on va se baser pour faire des analyses secondaires. Mais c'est souvent plus compliqué.

J. VANDENBUNDER : Pour l'enquête sur laquelle je suis en train de travailler en ce moment, les chercheuses¹² ont mis des codes couleur sur les entretiens. Donc il y a un mélange sur le document qui est à la fois de la donnée parce que c'est de la transcription d'entretiens mais qui fait déjà partie de

12 Danièle Chabaud et Dominique Fougeyrollas pour leur enquête sur le temps de travail domestique.

l'analyse parce que si on n'a pas le code couleur, on ne saisit pas toute la démarche de recherche.

S. CADOREL : Dans tous les cas, nous archivons de manière vraiment pérenne l'objet que l'on a créé : les données traitées et toute la documentation qu'on a produite. Ce que nous conservons dans nos archives propres n'est pas tout à fait la même version que celle qu'on peut trouver sur le site : il peut y avoir des données non anonymisées mais elles sont dans un coffre-fort.

Bibliographie

- ABBOTT Andrew, 2007, « Notes on replication », *Sociological Methods and Research*, vol. 36, n° 2, p. 210-219.
- BENDJABALLAH Selma, GARCIA Guillaume, CADOREL Sophie, GROSHENS Émilie, FROMONT Émilie et JUILLARD Émilie, 2017, « Valoriser les données d'enquêtes qualitatives en sciences sociales : le cas français de la banque d'enquête beQuali » *Documentation et bibliothèques*, vol. 63, n° 4, p. 73-85.
- BENDJABALLAH Selma, GARCIA Guillaume, CADOREL Sarah, GROSHENS Émilie, FROMONT Émilie et JUILLARD Émilie, 2018, « Anonymat et confidentialité des données : l'expérience de beQuali », *La diffusion numérique des données en SHS*, V. Ginouvès et I. Gras éd, Aix-en-Provence, Presses universitaires de Provence, p. 207-221.
- BOTH Anne et CADOREL Sarah, 2015, « Pour en finir avec l'original? », *Frontières d'archives. Recherches, mémoires, savoirs*, J.-F. Bert et M. J. Ratcliff éd., Paris, Éditions des archives contemporaines, p. 157-164.
- BOTH Anne et GARCIA Guillaume, 2014, « Le chercheur, l'archiviste et le webmaster : la polyphonie patrimoniale? Le cas de beQuali, banque d'enquêtes qualitatives en sciences sociales », *Heritage and Digital Humanities. How should training practices evolve?*, B. Dufréne éd., Berlin, Lit Verlag, p. 353-364.
- DUCHESNE Sophie et GARCIA Guillaume, 2014, « beQuali : une archive qualitative au service des sciences sociales », *Les archives de la recherche. Problèmes et enjeux de la construction du savoir scientifique*, M. Cornu, J. Fromageau et B. Müller éd., Paris, L'Harmattan, p. 49-70.