GUILLAUME GARCIA

La fabrication des normes d'anonymisation des données qualitatives

Pratiques, régulations et arbitrages professionnels

The manufacture of standards for anonymizing qualitative data
Pratices, regulations, and professional conflicts

Regards Sociologiques, n°67, pp.11-26

GUILLAUME GARCIA

La fabrication des normes d'anonymisation des données qualitatives

Pratiques, régulations et arbitrages professionnels

The manufacture of standards for anonymizing qualitative data
Pratices, regulations, and professional conflicts

Mots-clés :
anonymisation
données
entretiens
bonnes pratiques
régulation

Cet article propose une réflexion sur l'anonymisation des données d'enquêtes qualitatives, à partir d'un retour d'expérience sur le projet beQuali. Son ambition est de contribuer à rendre plus intelligibles les problèmes concrets, les dilemmes pratiques et les enjeux scientifiques de l'anonymisation de ce type de données, en partant des situations de travail effectives dans lesquelles elle se fabrique, se discute et se négocie. Plutôt que de juger l'anonymisation à l'aune de standards abstraits de « bonnes pratiques », l'article montre que ses tensions, ses coûts et ses impasses ne peuvent être compris qu'en replaçant cette opération au cœur des conditions sociales, institutionnelles et professionnelles dans lesquelles le partage des données qualitatives devient aujourd'hui possible — ou impossible.

Keywords:
anonymization
data
interview
best practices
regulation

This article offers a reflection on the anonymization of qualitative research data, based on feedback from the beQuali project. Its aim is to contribute to a clearer understanding of the concrete problems, practical dilemmas, and scientific challenges of anonymizing this type of data, starting from the actual work situations in which it is implemented, discussed, and negotiated. Rather than judging anonymization according to abstract standards of "best practices," the article shows that its tensions, costs, and dead ends can only be understood by placing this operation at the heart of the social, institutional, and professional conditions in which the sharing of qualitative data is now possible –or impossible.

Cet article propose une réflexion sur la fabrication des normes d'anonymisation des données qualitatives à partir d'une expérience située au sein d'un entrepôt de données de la recherche (projet beQuali, hébergé au Centre de données sociopolitiques)[1]. La problématique de l'anonymisation des données qualitatives cristallise une part importante des tensions liées à l'ouverture des données de la recherche en sciences sociales. L'obligation d'anonymiser les données personnelles et sensibles — qui constituent le sel de disciplines comme la sociologie ou la science politique — est devenue une condition quasi incontournable de leur partage élargi dans le cadre des entrepôts de données[2], lesquels se sont fortement développés récemment[3]. Cet enjeu est désormais bien identifié et fait l'objet de discussions croissantes, auxquelles ce dossier contribue. Cette problématique est d'ailleurs perçue par certains comme participant d'un mouvement plus large de montée des contraintes bureaucratico-éthiques pesant sur le travail d'enquête et sur le traitement des

1 Je remercie ici Catherine Leclerq et Yves Mirman pour leurs conseils dans la préparation de cet article, ainsi que les relecteurs anonymes pour leur contribution à l'amélioration de son propos.

2 Cf. Rebouillat Violaine & Schöpfel Joachim, « Le dispositif d'entrepôt de données de recherche », *in* Joachim Schöpfel, Violaine Rebouillat (dir.), *Partage et valorisation des données de recherche*, ISTE Group, 2023, pp.7-38.

3 Cf. Schöpfel Joachim, « Le paysage des entrepôts de données de recherche en France », *in* Joachim Schöpfel, Violaine Rebouillat (dir.), *Partage et valorisation des données de recherche*, *op. cit.*, pp.39-56.

données, un mouvement qui suscite inquiétudes et critiques dans une partie des sciences sociales françaises comme ailleurs[4]. La fabrication des normes d'anonymisation constitue pourtant encore aujourd'hui une boîte noire, malgré l'accroissement de l'attention portée à cette question depuis le fort développement du mouvement de la science ouverte en sciences sociales, il y a une dizaine d'années[5]. Le cadrage dominant de cette question en termes de « bonnes pratiques » tend à masquer un manque de retours d'expérience détaillés. Ce constat est d'autant plus paradoxal que l'anonymisation des publications constitue une pratique ordinaire du travail d'écriture fondée sur des entretiens ou des observations, tout en étant rarement discutée dans ses implications concrètes, notamment dans la littérature francophone. De nombreux discours sur l'anonymisation des données continuent de privilégier des approches techniques, juridiques ou procédurales, formulées en termes de principes généraux, souvent détachés des contraintes situées de l'enquête ; par contraste, les dimensions scientifiques, disciplinaires et pratiques de l'anonymisation restent peu abordées.

L'article défend la nécessité de dépasser le prisme des « bonnes pratiques », dans un contexte où un nombre croissant de chercheur·es sont incité·es à se sensibiliser, voire à participer au mouvement d'ouverture des données. Il prolonge une réflexion déjà engagée à partir des pratiques d'anonymisation mises en œuvre dans le cadre du projet beQuali[6], en approfondissant l'analyse du contexte et des configurations d'action dans lesquels ces pratiques prennent sens. Il défend la nécessité d'une réflexion davantage collective sur ces questions, au-delà des seuls retours réflexifs individuels, en proposant d'en sociologiser l'analyse.

Dans cette perspective, la première partie s'appuie sur l'expérience du projet beQuali pour analyser la manière dont les pratiques d'anonymisation peuvent se construire dans un contexte d'action situé, au croisement de contraintes techniques, juridiques et scientifiques. La deuxième partie propose un déplacement du regard en élargissant l'analyse à d'autres contextes et acteurs de l'écologie professionnelle[7] de la science ouverte, afin de mettre en évidence des tensions et des problèmes qui, dépassant le seul cas de beQuali, invitent à se demander jusqu'où il est possible de formaliser ce qui relève avant tout d'une casuistique. À titre programmatique, il croise et discute l'apport de plusieurs cadres théoriques susceptibles d'éclairer ce qui se joue en matière d'anonymisation des données, à différentes échelles — complémentaires — de la construction sociale des normes et dispositifs professionnels : celle des régulations collectives[8] au sein de l'écologie professionnelle de la science ouverte ; celle de la circulation, entre groupes professionnels mobilisés, d'objets-frontières[9] tels que les données, les normes ou les outils d'anonymisation ; celle des luttes de juridiction entre ces mêmes groupes professionnels pour définir qui a le droit de définir et traiter l'anonymisation[10] ; et à une échelle plus individuelle, celle des épreuves[11] concrètes auxquelles sont confronté·es les praticien·nes de l'anonymisation. Une dernière piste consiste à interroger l'influence des normes disciplinaires, à l'œuvre dans l'anonymisation des publications scientifiques, sur celle des données.

ENCADRÉ MÉTHODOLOGIQUE

Mon propos s'appuie sur trois sources d'information. D'abord, et à titre principal, sur une quinzaine d'années d'expérience au sein d'un centre de données, le CDSP, à travailler sur le projet beQuali, dédié à la diffusion numérique d'archives d'enquêtes qualitatives en sociologie et en science politique, principalement fondées sur des entretiens, et plus marginalement sur des observations. J'ai été impliqué dans ce dispositif à la fois dans la réflexion sur ses principes fondateurs et dans sa mise en œuvre pratique, en tant que responsable d'équipe, animateur de groupes de travail, en mobilisant un comité scientifique et technique, et à travers des activités de formation et de publication[12]. Au fil de l'existence du projet beQuali, une vingtaine d'enquêtes ont été traitées, représentant plus de 1500 transcriptions d'entretiens et plusieurs centaines de notes d'observation, dont la plupart ont dû être

4 CF. Rambaud Elsa, « (Faire) consentir, de la règle à la lettre. À propos du nouvel esprit de l'éthique », *Sociologie*, n°1, vol.16, 2025, pp.73-94.

5 Sur la carrière récente des discours autour de l'ouverture des données de la recherche, voir par exemple : Catone Maria Carmela, « The Role of Open Data in Digital Society : The Analysis of Scientific Trending Topics Through a Bibliometric Approach », *Frontiers in Sociology*, vol.8, 2023. Sur les difficultés de définition du mouvement même de la Science ouverte, voir : Gomez-Diaz Teresa, Recio Tomás, « Towards an Open Science Definition as a Political and Legal Framework : Sharing and Dissemination af Research Outputs », *Polis*, vol.19, 2020, pp.36-56.

6 Voir : Bendjaballah Selma, Cadorel Sarah, Fromont Émilie, Garcia Guillaume, Groshens Émilie & Juillard Émeline, « Anonymat et confidentialité des données. L'expérience de beQuali », *in* Ginouvès Véronique & Gras Isabelle (dir.), *La Diffusion numérique des données en SHS. Guide des bonnes pratiques éthiques et juridiques*, Aix-en-Provence, Presses universitaires de Provence, 2018, pp.207-221 ; Bendjaballah Selma, Garcia Guillaume & Sauger Nicolas, « Protéger les enquêtes, mais à quelles conditions ? Anonymiser des données d'enquêtes en sociologie et en science politique », *Terrains & travaux*, n°43, 2023, pp.257-279.

7 Cf. Abbott Andrew, « Linked Ecologies: States and Universities as Environments for Professions », *Sociological Theory*, n°3, vol.23, 2005, pp.245-274.

8 Cf. Reynaud Jean-Daniel, « Les régulations dans les organisations : régulation de contrôle et régulation autonome », *Revue française de sociologie*, n°1, vol.29, 1988, pp.5-18.

9 Cf. Star Susan Leigh & Griesemer James R., « Institutional ecology, "Translations", and Boundary objects : amateurs and professionals on Berkeley's museum of vertrebate zoologie », *Social Studies of Science*, n°3, vol.19, 1989, pp.387-420.

10 Cf. Abbott Andrew, *The System of Professions*, Chicago, University of Chicago Press, 1988.

11 Voir : Goffman Erving, *Stigmate. Les usages sociaux des handicaps*, Paris, Minuit, 1975 ; Becker Howard, *Outsiders : Études de sociologie de la déviance*, Paris, Métailié, 2020.

anonymisées, à des degrés divers, pour être diffusées, malgré un cadre d'accès contrôlé. Mon propos s'appuie également, à titre secondaire, sur les premiers résultats d'une enquête plus large sur les pratiques d'accompagnement au dépôt de données d'enquête, menée à partir d'un échantillon d'une vingtaine d'entrepôts de données de la recherche en sciences sociales (en Europe et en Amérique du Nord)[13]. Il s'appuie également sur de nombreux échanges avec des chercheurs, des ingénieurs de données, des juristes, des DPO, des archivistes et des homologues étrangers.

Le statut du propos qui sera développé est donc hybride, entre auto–analyse, observation participante et enquête dédiée. Les limites d'un tel exercice doivent être soulignées. Celles–ci tiennent d'abord à ma position d'acteur directement impliqué dans les configurations décrites, ce qui rend difficile toute prétention à une objectivation pleinement distanciée. Elles tiennent aussi au caractère en grande partie rétrospectif du retour d'expérience proposé sur le projet beQuali, qui ne repose pas sur une enquête construite en amont, mais sur une reconstruction a posteriori, à partir de notes de travail, de comptes rendus de réunions et de souvenirs nécessairement partiels. Les analyses développées ne sauraient de plus être assimilées ni à une position collective de l'équipe de beQuali, ni à celle du CDSP. Elles ne sont cependant pas entièrement dissociables de l'ensemble du travail collectif qui les a précédées. BeQuali a mobilisé, au fil des années, une dizaine de personnes issues de métiers différents (recherche en sciences sociales, archives et documentation), à laquelle se sont ajoutées des contributions plus ponctuelles dans le cadre de postdoctorats, d'un comité scientifique et technique ou de groupes de travail. Une partie de ces personnes est citée sur le site de beQuali : https://bequali.fr/fr/propos/.

DES PRATIQUES D'ANONYMISATION CONSTRUITES DANS UN CONTEXTE D'ACTION SITUÉ

Pour donner à comprendre la fabrication de principes d'anonymisation dont nous avons été les acteurs dans le cadre du projet beQuali, il faut d'abord revenir sur la manière dont nos positions et nos pratiques se sont construites en interaction étroite avec un contexte d'action situé et des cas concrets d'enquêtes, plutôt que comme des choix abstraits.

Mon expérience de l'anonymisation des données qualitatives débute en 2009, lorsque, à l'issue d'un postdoctorat, je suis recruté à Sciences Po pour participer à la mise en place d'une banque d'enquêtes qualitatives à destination de la communauté scientifique et académique, étudiants inclus. Très tôt, l'un des enjeux centraux consiste à concilier diffusion élargie des matériaux d'enquête et respect de la confidentialité des propos recueillis auprès des enquêtés, ainsi que la protection de leurs données personnelles[14].

Deux interrogations structurent alors la réflexion. La première concerne la possibilité d'obtenir a posteriori le consentement des enquêtés, ce qui aurait permis, dans certains cas, d'éviter le travail d'anonymisation. Cette option s'avère rapidement difficilement praticable à grande échelle, en dehors de situations très spécifiques que nous avons parfois réussi à traiter. La seconde interrogation porte sur les modalités concrètes de l'anonymisation elle–même. À cette époque, les boîtes à outils disponibles sont particulièrement limitées. On ne parle pas encore vraiment de Science ouverte, ni même d'ouverture des données de la recherche, mais plutôt de diffusion et de réutilisation des données d'enquêtes, dans une logique infrastructurelle ou archivistique. L'écologie professionnelle de la diffusion des données qualitatives est alors marquée par une configuration[15] mettant en présence un nombre restreint d'acteurs — notamment des historiens et des archivistes[16] — et des modalités de consultation souvent sur place, sans diffusion de copies numériques, ce qui limite fortement la nécessité d'anonymisation des matériaux. Il existe bien un

12 Sur les ambitions initiales du projet, au début des années 2010, voir : Duchesne Sophie, Garcia Guillaume, « BeQuali : une archive qualitative au service des sciences sociales », *in* Marie Cornu & Jérôme Fromageau (dir), *Les archives de la recherche. Problèmes et enjeux de la construction du savoir scientifique*, Paris, L'Harmattan, pp.35-56. Sur les contours du dispositif beQuali, voir : Bendjaballah Selma, Cadorel Sarah, Garcia Guillaume, Groshens Émilie, Fromont Émilie, Juillard Émeline, « Valoriser les données d'enquêtes qualitatives en sciences sociales : le cas français de la banque d'enquête beQuali », *Documentation et bibliothèques*, n°4, vol.63, 2017, pp.73-85. Voir également : Cadorel Sarah, Garcia Guillaume, Groshens Emilie, Juillard Emeline, Vandenbunder Jérémie, Le Courant Stéphane, Pecqueux Anthony & Noûs Camille, « Archiver, documenter, enquêter sur l'enquête qualitative : le travail de l'ombre de beQuali », *Tracés. Revue de sciences humaines*, n°19, 2019, pp.183-198.

13 Projet UPMET (Upscaling Metadata for Increasing Reuse in the Social Sciences), financé dans le cadre de l'appel Flash Science ouverte 2019 de l'ANR (ANR-19 DATA-0009-01). Les premiers résultats de cette enquête seront bientôt publiés dans : Garcia Guillaume, Colin Paul & Danciu Alina, « Between quality standards and organizational constraints: regulating research data deposit support in research data repositories », *Journal of Documentation*, à paraître.

14 Cf. Duchesne Sophie & Garcia Guillaume, « beQuali : une archive qualitative au service des sciences sociales », *art. cit.*

15 Il s'agit ici de désigner des ensembles d'acteurs, d'institutions et de problèmes en interaction, marqués par des interdépendances, où chaque groupe professionnel occupe, défend ou perd des juridictions (cf. infra) ; en ce sens, « l'anonymisation » peut être vue comme un site écologique où interagissent chercheurs et enseignants, juristes et DPO, ingénieurs, documentalistes, archivistes, et institutions de régulation.

16 Par exemple, le Réseau Archives des mondes contemporains à Dijon, devenu le centre Marcel Maget (https://pandor.u-bourgogne.fr/fr/node/7), le réseau Archives des ethnologues à Nanterre (https://www.archivesethnologues.fr/) ou la phonothèque de la MMSH (https://phonotheque.hypotheses.org/)

guide « Informatique et libertés » à l'usage des chercheurs, produit par la Direction des affaires juridiques du CNRS, mais qui fournit surtout un cadre juridique général[17]. Le « guide des bonnes pratiques pour les corpus oraux »[18] constitue une ressource plus opérationnelle, avec des exemples issus de la sociolinguistique, mais sa transposition aux enquêtes de sociologie ou de science politique reste partielle. Quelques travaux en sociologie abordent également l'anonymisation dans les publications scientifiques (recours aux pseudonymes, brouillage ou généralisation des informations identifiantes), sans proposer de cadre stabilisé pour la diffusion de données d'enquête[19]. Ces travaux restent rares en France, et peu nombreux en anglais[20]. À ces ressources s'ajoutent des pratiques formalisées dans certains centres de données, à l'étranger comme en France, diffusant des données quantitatives (dont le réseau Quetelet, dont le CDSP était membre) ou qualitatives (à l'époque essentiellement Qualidata au Royaume–Uni, et le Finnish Social Science Data Archive — FSD — en Finlande), qui proposent des « recettes » similaires : attention aux données sensibles, prise en compte des identifications indirectes *via* la combinaison de multiples sources d'information non identifiantes prises isolément, ou encore généralisation de variables sociographiques ou géographiques. Enfin, moi et certains de mes collègues pouvons-nous appuyer sur notre propre expérience de recherche, marquée par des pratiques ordinaires d'anonymisation dans les publications, issues d'une socialisation professionnelle diffuse plutôt que d'un apprentissage formalisé. Dans ce contexte, il a fallu improviser, expérimenter et ajuster progressivement les pratiques.

Dans un nombre non négligeable de cas, l'anonymisation s'est révélée relativement aisée — bien que (parfois très) chronophage — et (assez) peu pénalisante pour la richesse des données ainsi remises en circulation. Des enquêtes comme *La formation des couples*, *Le sens de l'hospitalité* d'Anne Gotman (1997–1999), ou encore *Les Français et la politique* d'Étienne Schweisguth (1983) et *Représentation du champ social, attitudes politiques et changements socio–économiques* (1978) de Jean–Marie Donégani, Guy Michelat et Michel Simon, relèvent de cette configuration. Ces enquêtes portent le plus souvent sur des individus à faible surface sociale, aux trajectoires peu visibles, et sur des sujets de sensibilité variable, mais globalement modérée. Pour ces enquêtes, assez anciennes, la probabilité de retrouver aujourd'hui des informations complémentaires permettant l'identification des personnes apparaît réduite, ou du moins atténuée par le temps[21]. Ces enquêtes n'ont pas de dimension ethnographique forte, c'est–à–dire de contexte local très précis difficile à neutraliser, et reposent sur des échantillons d'individus pour lesquels l'absence d'interconnaissance permet de traiter chaque entretien de manière relativement indépendante, sans stratégie globale de prévention des recoupements possibles. Ces configurations limitaient mécaniquement les risques liés à la diffusion élargie des matériaux, sans les annuler totalement. Les données ont alors été anonymisées de manière relativement « chirurgicale » : suppression des identifiants directs (noms, adresses précises) et atténuation limitée des identifiants indirects, par exemple par généralisation parcimonieuse d'informations socioprofessionnelles ou géographiques, ou floutage de certaines trajectoires biographiques très spécifiques.

Nous aurions pu en rester là, en restreignant notre périmètre d'action à des enquêtes « faciles » à anonymiser. Mais les conditions mêmes de développement du projet beQuali nous ont conduits à élargir le spectre des cas traités. Il faut rappeler ici que le projet s'est construit de manière contingente, au gré des opportunités de financement : d'abord via des microfinancements (Adonis — ancien IR Corpus, aujourd'hui Humanum ; puis le Scientific Advisory Board de Sciences Po), puis à travers le projet ANR Réanalyse, dans une logique d'expérimentation. Le projet a ensuite pris une dimension nationale lorsqu'il a été hébergé au CDSP, puis intégré à l'Equipex DIME-SHS (ANR–10–EQPX–19–01).

La diversification progressive des enquêtes déposées à partir de 2015, notamment via des appels à dépôt nationaux, a joué un rôle décisif à cet égard. À la nécessité d'élargir la base disciplinaire et méthodologique des enquêtes ont correspondu des propositions effectivement reçues et sélectionnées par le comité scientifique. L'arrivée de corpus plus ethnographiques a fait émerger de nouveaux problèmes d'anonymisation. Ceux–ci étaient liés au fait qu'une partie des nouvelles enquêtes déposées se situait dans une zone grise : le partage de leurs matériaux restait envisageable, mais au

17 Cf. CNRS, Direction des affaires juridiques, *Informatique et libertés — Guide à l'usage des chercheurs*, 2006 [en ligne, consulté le 30 janvier 2026].

18 Cf. Baude Olivier, Blanche-Benveniste Claire, Calas Marie-France, Cappeau Paul, Cordereix Pascal, Goury Laurence, Jacobson Michel, de Lamberterie Isabelle, Marchello-Nizia Christiane & Mondada Lorenza, *Corpus oraux, guide des bonnes pratiques*, Orléans, Presses Universitaires d'Orléans/CNRS, 2006, [en ligne, consulté le 30 janvier 2026].

19 Sur ce point : Béliard Aude & Eideliman Jean-Sébastien, « Au-delà de la déontologie. Anonymat et confidentialité dans le travail ethnographique », *in* Fassin Didier & Bensa Alban (dir.), *Les Politiques de l'enquête*, Paris, La Découverte, 2008, pp.123-141 ; Weber Florence, « Publier des cas ethnographiques : analyse sociologique, réputation et image de soi des enquêtés », *Genèses*, n°1, vol.70, 2008, pp.140-150 ; Zolesio Emmanuelle, « Anonymiser les enquêtés », *¿Interrogations ?, Revue pluridisciplinaire de sciences humaines et sociales*, n°12, 2011, pp.174-183.

20 Voir : Bendjaballah Selma, Cadorel Sarah, Fromont Émilie, Garcia Guillaume, Groshens Émilie & Juillard Émeline, « Anonymat et confidentialité des données. L'expérience de beQuali », *art. cit.*

21 Weber Florence, « Questions de déontologie », *in* Stéphane Beaud & Florence Weber, *Guide de l'enquête de terrain*, Paris, La Découverte, 2010, p.269.

prix d'un travail d'anonymisation plus complexe, plus coûteux et générateur d'incertitudes plus marquées.

Des enquêtes comme *Le choix de l'école*, dirigée par Agnès Van Zanten, ou *Dilapidation et prodigalité*, réalisée par Anne Gotman, illustrent ces configurations ; celles–ci se caractérisent par des sujets plus sensibles, des terrains localisés, une interconnaissance parfois marquée entre certains enquêtés, des références croisées entre entretiens et notes d'observation, une densité élevée d'informations biographiques, une hétérogénéité interne des corpus, ou encore le caractère un peu plus récent de ces recherches, qui augmentent la probabilité de recoupement avec des sources externes. L'enquête *Avec ceux du FN* de Daniel Bizeul constitue sans doute un cas limite particulièrement éclairant[22]. Elle repose sur un matériau ethnographique dense, combinant entretiens et observations auprès de militants du Front national dans différents contextes : association caritative satellite du parti, siège national, Front national des jeunes, section d'arrondissement parisien. Les précautions à prendre pour protéger les enquêtés étaient d'autant plus nécessaires que l'appartenance au FN constitue, encore aujourd'hui, un objet d'opprobre social, susceptible d'entraîner des conséquences professionnelles, familiales ou relationnelles importantes. Même pour des militants occupant à l'époque des positions élevées au sein du FN, les risques demeuraient : certains avaient quitté le parti, d'autres continuaient leur engagement dans d'autres conditions, et tous pouvaient être exposés à des formes de préjudice liées à la réactivation de propos tenus en entretien sur des éléments stratégiques internes au parti ou sur des positions idéologiques radicales.

La forte dimension ethnographique de l'enquête renforçait ces risques. Les citations croisées entre entretiens et notes d'observation augmentaient mécaniquement les possibilités de reconnaissance par recoupement, même à partir d'éléments isolément anodins. Il fallait également intégrer le risque, auquel le chercheur à l'origine de l'enquête, Daniel Bizeul, était particulièrement sensible, qu'une partie du public académique — notamment étudiant — puisse disposer de ressources militantes ou relationnelles spécifiques ; il n'était pas irréaliste d'envisager que certains utilisateurs puissent mobiliser ces ressources pour tenter d'identifier des militants évoqués dans les matériaux, ce qui justifiait une vigilance accrue.

Face à ces contraintes, et en concertation étroite avec Daniel Bizeul, nous avons mis en place des stratégies différenciées selon les profils d'enquêtés et la nature des informations. Pour les cadres du parti, nous avons conservé l'essentiel des trajectoires militantes et électorales, en réduisant toutefois la précision de certains éléments géographiques ou organisationnels. Pour les militants moins visibles, l'anonymisation a été plus poussée. Nous avons également supprimé quelques propos susceptibles de porter préjudice à certains militants, notamment lorsque ces propos étaient déjà connus ou publiés ailleurs, à l'image de propos racistes ou négationnistes. Ce choix se justifiait du point de vue de l'économie générale de l'enquête : comme le souligne Daniel Bizeul, l'existence de tels propos était déjà largement documentée et ne constituait pas en soi l'objet central de son travail, dont l'originalité résidait plutôt dans l'analyse des logiques d'engagement et des pratiques militantes. Avec cet exemple, on atteint une forme de point de butée des possibilités d'anonymisation[23].

Pour ces enquêtes qui se situent dans cette « zone grise », voire qui en sont à la limite, nous avons été amenés à travailler au plus près des limites acceptables de l'anonymisation, en assumant des choix parfois coûteux du point de vue de la fluidité de lecture — puisque la multiplication des marqueurs de l'anonymisation dans une transcription d'entretien en perturbe d'autant la lecture — ou de la richesse informationnelle des matériaux, mais pensés comme des compromis mesurés permettant la diffusion de l'enquête dans un cadre contraint. Cette expérience soulève la question de savoir jusqu'où il est possible de formaliser ce qui semble relever, principalement, d'une casuistique.

JUSQU'OÙ PEUT-ON FORMALISER CE QUI RELÈVE PRINCIPALEMENT D'UNE CASUISTIQUE ?

Une activité d'arbitrage sous contrainte, fondée sur l'expérience

Ces stratégies d'anonymisation ont été « inventées » et déployées dans un cadre et un contexte d'action en partie spécifiques, mais qui renvoient sans doute à des configurations relativement communes dans ce secteur. Il s'agit dès lors d'en dégager une expérience centrale plus générale.

Globalement, nous avons procédé à une pseudonymisation renforcée au sens du RGPD, c'est–à–dire à des transformations destinées à réduire les risques d'identification sans prétendre supprimer toute possibilité théorique de réidentification[24]. À notre échelle, l'anonymisation n'a jamais été appliquée de manière standardisée ou mécanique. Nous avons dû, enquête après enquête, nous ajuster aux difficultés spécifiques que chacune posait, et qui variaient fortement selon plusieurs paramètres interconnectés, comme le montre le

22 Je reprends ici des éléments exposés dans le rapport de l'enquête, établi avec Daniel Bizeul : Garcia Guillaume & Cadorel Sarah, *Enquête sur l'enquête «Avec ceux du FN (1995-2003)» de Daniel Bizeul*, Rapport de recherche, Centre de données socio-politiques, Fondation nationale des sciences politiques, 2020, pp.31-34 [en ligne, consulté le 30 janvier 2026].

23 Il m'est impossible, par manque de place, d'aborder ici les problèmes spécifiques posés par la pseudonymisation des prénoms.

24 Cf. Bendjaballah Selma, Garcia Guillaume & Sauger Nicolas, « Protéger les enquêtes, mais à quelles conditions ? », *art. cit.*

tableau n°1 : la sensibilité des sujets abordés, la dimension ethnographique des terrains, la localisation et la surface sociale des populations ciblées, l'ancienneté relative de l'enquête, l'existence de citations croisées entre enquêtés dans les entretiens (ou entre observations et entretiens), ou encore la présence de bases de données supplémentaires susceptibles de favoriser l'identification des personnes.

Tableau n°1 : Panorama des problèmes d'anonymisation selon les propriétés des enquêtes

nom de l'enquête	sensibilité de l'objet	dimension ethnographique	ancienneté relative du terrain	présence de BDD annexes	citations croisées entre entretiens	entretien et notes d'observation	Bilan problèmes d'anonymisation
La formation des couples	moyenne	faible	assez ancien	opaque	très faible	non	- - -
Le sens de l'hospitalité	faible	faible	relativement récent	faible	très faible	non	-
Les Français et la politique	moyenne	faible	assez ancien	faible	faible	non	-
Représentations du champs....	moyenne	très faible	encore plus ancien	faible	très faible	non	-
Le choix de l'école	moyenne	forte	relativement récent	probable	forte	non	+
Dilapidation et prodigalité	forte	moyenne	relativement ancien	probable	moyenne	non	+
Avec ceux du FN	très forte	très forte	plus récent	très probable	très forte	oui	+ +

Anonymiser, en tant qu'activité, ne s'est jamais limité à la suppression ou à la généralisation mécanique d'identifiants directs ou indirects ; cette activité nécessite une réflexion pleinement intégrée au raisonnement sociologique[25], afin d'opérer un tri raisonné entre les informations indispensables à l'interprétation et celles qui peuvent être euphémisées ou obfusquées. Cela a impliqué un travail d'interprétation attentif aux détails, aux allusions, aux implicites, aux informations disséminées et embarquées, parfois de manière subtile, dans les descriptions et les scènes rapportées par les enquêtés, et a nécessité d'être particulièrement attentif aux possibles effets de cascade, notamment lorsqu'on modifie certaines informations dans des récits biographiques, ou lorsque les entretiens d'un corpus donné s'inscrivent dans un même contexte ethnographique. Mais contrairement à une publication, où le chercheur choisit une perspective démonstrative explicite et anonymise en conséquence, anonymiser des données en vue de leur partage élargi impose d'arbitrer a priori, sans intention analytique prédéfinie. Cette expérience indique que chercher un hyperonyme ajusté, adapter le niveau de généralité d'une information, préserver les propriétés les plus pertinentes d'une situation ou d'une personne au regard d'un contexte spécifique et d'une enquête orientée vers un objectif de recherche ciblé, tout en réduisant les risques d'identification, le tout sans savoir précisément à qui ni à quoi les matériaux vont servir, s'apparente à certains égards à une forme de pari — dont le coût est surtout élevé, sans doute, pour les enquêtes situées dans la « zone grise » de l'anonymisation précédemment évoquée. Les choix d'anonymisation que nous avons faits ont cependant toujours été orientés vers certains usages jugés plus plausibles que d'autres. Ils ont consisté à privilégier la cohérence disciplinaire des enquêtes et les réutilisations les plus probables, au prix de compromis assumés, par exemple entre précision géographique et finesse sociologique[26].

Dans cette expérience, l'anonymisation a ainsi relevé moins d'un savoir–faire procédural acquis d'emblée par la lecture et l'application de guides, textes ou préconisations déjà disponibles — bien que ces principes aient été fort utiles — que d'une compétence acquise progressivement par l'expérience, fondée notamment sur l'accumulation de cas.

Pour répondre à ces besoins d'expertise, nous avons progressivement constitué une équipe plurimétier

25 Cf. Béliard Aude & Eideliman Jean-Sébastien, « Au-delà de la déontologie », *art. cit.*

26 Cf. Bendjaballah Selma, Garcia Guillaume & Sauger Nicolas, « Protéger les enquêtes, mais à quelles conditions ? », *art. cit.* Cette situation explique pourquoi certains centres de données historiques dans la diffusion de matériaux qualitatifs, comme Qualidata ou le Finnish Social Science Data Archive, ont longtemps mobilisé des groupes d'étudiants en sociologie ou en science politique pour anonymiser les corpus déposés, reconnaissant de fait l'importance d'une expertise disciplinaire.

capable de couvrir l'ensemble de la chaîne de traitement des enquêtes, dont l'anonymisation.[27] Cette organisation reposait, pour chaque enquête, sur un binôme associant une personne familière des pratiques d'enquête en sciences sociales (titulaire d'un doctorat en sociologie ou en science politique qualitatives) et une autre disposant d'une formation et de compétences en documentation ou en archivistique, ce qui permettait d'hybrider les compétences nécessaires pour gérer ces contraintes.

Pendant une longue période, à l'instar de nombreux centres de données, beQuali a pris en charge directement le travail d'anonymisation pour le compte des chercheurs déposants. L'externalisation — du point de vue des déposants — de ce travail permettait de consacrer du temps à cette activité chronophage et de rendre ce dispositif attractif à une époque où il fallait, plus encore qu'aujourd'hui, légitimer l'existence d'une « banque » d'enquêtes qualitatives. Mais ce travail a, dans la mesure du possible, reposé également sur une collaboration étroite avec les chercheurs et chercheuses déposants, faite de réunions, d'échanges itératifs et d'ajustements successifs. Dans le cas de *Dilapidation et prodigalité*, par exemple, plusieurs réunions approfondies avec Anne Gotman ont permis de définir les principes généraux de l'anonymisation, puis d'en affiner l'application au fil d'un examen détaillé des entretiens, en discutant cas par cas des informations à conserver ou à atténuer. De même, pour *Avec ceux du FN*, les stratégies ont été élaborées à l'issue de nombreux échanges approfondis avec Daniel Bizeul, combinant définition de principes transversaux et ajustements document par document, à partir d'un premier travail d'anonymisation déjà engagé par le chercheur. Ce modèle s'est progressivement trouvé mis sous tension par la diminution de nos ressources humaines, à la suite de la fin de l'Équipex à partir de 2020. Dans ce contexte, une transformation de l'économie de la curation des données a nécessité le passage progressif à un modèle d'accompagnement au dépôt, impliquant une délégation partielle de certaines tâches — dont l'anonymisation — vers les chercheurs eux–mêmes, avec une vérification en fin de processus[28]. Ce basculement, initialement contraint par des impératifs pragmatiques, a également constitué un moyen de réimpliquer davantage les chercheurs déposants, en leur redonnant une place plus importante dans cette activité qui demeure structurellement délicate et parfois source de multiples problèmes d'alignement (de calendriers, de stratégies d'anonymisation, etc.), mais qui les concerne au premier chef en tant que responsables de la protection des enquêtés.

Dans le prolongement de ces évolutions, nous avons engagé un travail de formalisation de l'expertise accumulée en la matière, sous forme de guides, transmis aux déposants. Cette démarche soulève toutefois une question centrale : les « recettes » mises en place sont-elles susceptibles d'être généralisées, objectivées et formalisées dans des guides, ou bien relèvent-elles d'une casuistique difficilement transmissible autrement que par des principes généraux et une mise à l'épreuve pratique ? Autrement dit, peut-on s'engager dans la valorisation de « bonnes pratiques » à destination des déposants, et quelles « bonnes pratiques » précisément ?

Quand les « bonnes pratiques » précèdent l'analyse des pratiques

Avant d'apporter quelques éléments de réflexion à cette question, il faut rappeler que le registre des bonnes pratiques n'est évidemment pas nouveau en ce domaine. Sa saillance remonte au moins au milieu des années 2000, visible dans des guides méthodologiques ou juridiques destinés aux chercheurs, puis dans des initiatives collectives visant à encadrer juridiquement et éthiquement la diffusion des données en SHS. À partir de la fin des années 2010, toutefois, le débat sur l'ouverture des données de la recherche a changé d'échelle[29]. Sur les quinze dernières années, on peut observer comment les discours et les incitations institutionnelles en faveur du partage des données ont progressivement mis à l'agenda la question de l'anonymisation, en particulier pour les données qualitatives, dans le cadre de l'institutionnalisation du discours sur la Science ouverte, autour de 2016-2018. Ce qui relevait auparavant d'un régime de volontariat, caractérisant un problème confiné à des arènes localisées, restreintes et spécialisées (par exemple au sein du réseau Quetelet ou de circuits de relations privilégiées avec certaines universités et laboratoires), tend à devenir un enjeu plus public, inscrit dans des dispositifs institutionnels plus visibles et de plus en plus mobilisateurs : plans nationaux pour la science ouverte depuis 2018, politiques de l'ANR incitant les projets financés à déposer leurs données ou obligeant à fournir des plans de gestion des données, création de nouvelles instances (notamment le COSO/FNSO), multiplication des dispositifs ou réseaux d'accompagnement ou d'encadrement (Ateliers de la donnée, RDG, réseau interdonnées, etc.), actions de sensibilisation et de formation autour du RGPD, guides pour la recherche fournis par l'INSHS[30], et enfin multiplication des événements, séminaires et formations autour de l'ouver-

27 Cf. Cadorel Sarah, Garcia Guillaume, Groshens Émilie, Juillard Émeline, Vandenbunder Jérémie, Le Courant Stefan, Pecqueux Anthony & Noûs Camille, « Archiver, documenter, enquêter sur l'enquête qualitative », *art. cit.*

28 BeQuali s'est ainsi retrouvé dans une situation comparable à celle de nombreux entrepôts de données. Ce problème concerne, dès leur installation, la plupart des entrepôts institutionnels (par exemple, en France, RDG).

29 Sur ce point, voir l'étude exploratoire (dans une perspective d'analyse de la construction des problèmes publics) de Hamzaoui Oussama, Marrel Guillaume : « Traitements de données et procéduralisation de la recherche », *in* Anaïs Théviot (dir.), *Gouverner par les données ? Pour une sociologie politique du numérique*, Lyon, ENS, 2023, pp.151-171.

30 Cf. CNRS, *Les sciences humaines & sociales et la protection des données à caractère personnel dans le contexte de la science ouverte*, V2, février 2021, CNRS.

ture des données de la recherche. Ces arènes semblent de plus en plus investies par de nouveaux acteurs qui sont censés se mobiliser aussi sur ces questions (comme les ingénieurs, quels que soient leurs métiers spécifiques, qui peuplent le réseau MATE-SHS) ou qui se spécialisent sur ces enjeux (comme les DPO), sans avoir toujours de formation ou d'expertise adaptées à la résolution de cas pratiques. Le nombre de chercheurs qui intègrent ces arènes pour s'informer, voire se préparer à un dépôt et à l'anonymisation, demeure relativement limité — en tout cas nettement inférieur au potentiel[31]. Si l'anonymisation des données devient un enjeu objet de nouvelles appropriations, elle reste, en tant qu'activité concrète, relativement peu discutée dans sa complexité et ses ambivalences. Nombre d'événements prennent la forme de formations ou de séminaires à dominante éthico-juridique, sans produire de véritables espaces de débat disciplinaire ni donner lieu à des retours d'expérience largement diffusés — les publications sur le sujet restent peu nombreuses, malgré ce changement d'échelle[32].

Dans ce contexte, on observe une situation paradoxale : alors même que les difficultés de l'anonymisation sont relativement bien perçues par les acteurs concernés, et que les retours d'expérience empiriques restent rares, partiels et dispersés, la question tend à être de plus en plus fortement cadrée par le registre des « bonnes pratiques ».

Cette situation oblige à clarifier ce que l'on entend par « bonnes pratiques », souvent mobilisées comme une évidence, alors même que ce registre recouvre en réalité deux visions distinctes. Dans un certain nombre de cas est mobilisée une acception que l'on peut qualifier de « souple », reposant sur la promotion de grands principes — comme la recherche d'un équilibre entre protection des enquêtés et richesse des données, l'évitement du recours aux cas fictifs, ou encore un usage parcimonieux des hyperonymes. Ces bonnes pratiques, formulées de manière très générale, proposent avant tout des orientations, d'autant plus facilement acceptables qu'elles renvoient à des stratégies éprouvées dans les publications de recherche ou à des procédures appelées à être adaptées en pratique — autrement dit, elles n'imposent pas de solutions toutes faites. C'est ce type de « bonnes pratiques » que l'expérience de beQuali est susceptible, éventuellement, de nourrir et d'étayer. Les choses sont différentes lorsqu'on considère une conception plus exigeante des bonnes pratiques, parfois exprimée de manière implicite, mais qui embarque un projet de standardisation — par exemple obfusquer systématiquement tous les éléments sociographiques relatifs aux professions, hyperonymiser systématiquement tous les noms de lieux, changer systématiquement tous les prénoms, ou encore faire disparaître systématiquement l'année de naissance ou l'âge précis d'une personne (pour les remplacer, par exemple, par une fourchette).

Une exploration rapide de la littérature sur le sujet tend d'ailleurs à montrer la faible assise empirique des discours relevant de ce registre plus prescriptif. Les sciences de gestion offrent ici des éclairages intéressants : certains travaux montrent que les « bonnes pratiques » dans le domaine du changement organisationnel reposent largement sur des anecdotes[33]. D'autres travaux en *library and information science* soulignent également la difficulté d'apporter la preuve que les « bonnes pratiques » dans ce domaine ont effectivement démontré leur efficacité[34]. Il serait utile d'approfondir cette question, notamment pour déterminer qui, précisément, parle de quelles bonnes pratiques.

Dans le domaine des centres de données, les rares publications existantes s'appuient soit sur des considérations générales, technico-juridiques, souvent produites par des entrepôts de données ou des institutions, soit sur des retours d'expérience encore limités de chercheurs confrontés au dépôt de leurs données[35]. La rareté de ces textes, bien qu'importants, favorise peu l'accumulation de connaissances, la comparaison systématique entre stratégies d'anonymisation ou l'évaluation des risques concrets d'identification. Les retours d'expérience discutant précisément les arbitrages d'anonymisation effectivement mis en œuvre restent eux aussi rares[36]. Une comparaison avec d'autres centres de données du même univers disciplinaire — UK Data Archive (hébergeant feu Qualidata), Finnish Social Science Data Archive, AUSSDA, Qualitative Data Repository, etc. — montre que, si les consignes d'anonymisation varient dans leur degré de détail ou leur formalisme, elles reposent généralement sur des principes similaires et demeurent souvent peu explicitées. Une « bonne » anonymisation n'y est presque jamais illustrée par des cas concrets servant

31 Cf. Hamzaoui Oussama & Marrel Guillaume, « Traitements de données et procéduralisation de la recherche », *art. cit.*

32 Voir : récemment Riou Camille, « Les enjeux de l'anonymisation dans une enquête auprès d'expert·e·s de la question des violences sexuelles », *Les Enjeux de l'information et de la communication*, n°3, vol.25, 2025, pp.91-105.

33 Cf. Hallencreutz Jacob & Turner Dawn-Marie, « Exploring organizational change best practice: Are there any clear-cut models and definitions? », *International Journal of Quality and Service Sciences*, n°1, vol.3, 2011, pp.60-68.

34 Cf. Druery Jackie, McCormack Nancy & Murphy Sharon, « Are Best Practices Really Best ? A Review of the Best Practices Literature in Library and Information Studies », *Evidence Based Library and Information Practice*, n°4, vol.8, 2013, pp.110-128.

35 Dont : Stewart Emma & Shaffer Madeleine, « The Ethical and Practical Challenges of Archiving Refugee Accounts », *Bulletin of Sociological Methodology / Bulletin de méthodologie sociologique*, n°150, 2021, pp.51-69 ; Bizeul Daniel, « Faut-il tout dévoiler d'une enquête au Front national ? Réflexions sur le partage des données et le devoir éthique en sociologie », *Bulletin of Sociological Methodology / Bulletin de méthodologie sociologique*, n°150, 2021, pp.70-105 ; Bouchet Célia « Comment j'ai déposé les données de ma recherche (sans savoir ce qui m'attendait) », *Genèses*, n°133, 2023, pp.113-128.

36 À quelques exceptions près, comme dans : Saunders Benjamin, Kitzinger Jenny & Kitzinger Celia, « Anonymising Interview Data : Challenges and Compromise in Practice », *Qualitative Research*, n°5, vol.15, 2015, pp.616-632.

de modèles, permettant de comprendre comment arbitrer, dans la pratique, entre protection des enquêtés et maintien de la valeur analytique des données.

Cette absence d'explicitation des arbitrages renvoie plus largement à une difficulté structurelle : dans l'expérience de beQuali — comme probablement ailleurs — aucune des solutions mises en œuvre n'a été véritablement mise à l'épreuve de manière contrôlée. Il est en effet extrêmement difficile de démontrer empiriquement qu'une stratégie d'anonymisation donnée constitue une « bonne » pratique au sens fort, faute de dispositifs permettant de comparer systématiquement les effets de versions anonymisées et non anonymisées dans des conditions réalistes d'usage et de circulation des données. Et ce même si l'on peut raisonnablement distinguer les effets de l'anonymisation selon les contextes de réutilisation — probablement plus pénalisante pour la réanalyse qualitative fine que pour l'enseignement, avec toute une gamme de situations intermédiaires (analyse secondaire, socio-histoire des sciences, préparation d'une nouvelle enquête, réutilisation méthodologique, etc.).

Cette tension entre exigences de formalisation et caractère situé des arbitrages s'est manifestée de différentes manières dans la trajectoire du projet beQuali. À un niveau institutionnel, le développement du projet dans le cadre de l'Équipex DIME-SHS devait s'inscrire, dans sa première phase, comme une preuve de concept avant une éventuelle stabilisation des procédures. Le destin de beQuali n'a pas été exactement celui–ci en matière d'anonymisation, en raison notamment de la nature du projet, qui pouvait difficilement produire un modèle standardisé applicable à des enquêtes déjà réalisées et soumises au dépôt. Il fallait au contraire s'adapter aux spécificités de chacune et collaborer étroitement avec les chercheurs déposants, ce qui plaidait pour le maintien d'une marge de manœuvre. La fin de l'Équipex et le passage à la formule du dépôt accompagné ont renforcé cette orientation. Cette tension s'est également exprimée dans les interactions plus concrètes autour du projet. À plusieurs reprises, l'expérience accumulée à beQuali a suscité des sollicitations pour « partager des bonnes pratiques ». Or, les solutions que nous pouvions restituer reposaient sur une expérience finalement limitée au regard de la diversité des configurations possibles. Une vingtaine d'enquêtes, certes diverses, mais très loin de couvrir l'ensemble des configurations possibles — sans aucune prétention à la saturation, pour reprendre les termes de la théorisation ancrée. L'expérience accumulée permettait surtout de proposer des récits réflexifs sur des arrangements situés, construits au fil des cas, souvent dans l'incertitude et la discussion collective. Ces récits avaient surtout vocation à être partagés et discutés à partir de cas pratiques, au même titre que les discussions méthodologiques sur l'entretien ou l'observation. Ce décalage était également perceptible lors de notre participation au projet de *Guide de bonnes pratiques éthiques et juridiques pour la diffusion des données en SHS*. Notre contribution, motivée par le désir de contribuer à ce débat plutôt que d'en rester éloigné, y figurait dans une partie dédiée aux « réalisations et retours d'expérience ». Ce positionnement était cohérent avec notre approche[37].

SOCIOLOGISER L'ANALYSE DES PRATIQUES D'ANONYMISATION

Les arguments qui semblent circuler dans les débats auxquels j'ai pu assister dans des arènes souvent très localisées — séminaires spécialisés, réunions professionnelles, échanges informels entre pairs — ou que l'on retrouve dans certaines prises de position écrites, donnent l'impression, fréquemment, de s'appuyer sur des positions de principe ou des micro-exemples rarement pleinement contextualisés. Ils prennent parfois la forme de dénonciations d'injonctions bureaucratico-techniques ou d'une vision excessivement procédurale de l'anonymisation, ou à l'inverse de discours proches de « rappels à la loi ». Si ces prises de position peuvent en situation parfois être voilées, ou exprimées de manière mesurée, elles sont de nature à nourrir des postures défensives, au détriment d'un travail collectif d'élucidation des contraintes situées dans lesquelles les attentes et les pratiques réelles des parties prenantes s'inscrivent. Il me paraît nécessaire d'appliquer à ces débats les outils de la sociologie, encore insuffisamment mobilisés. Je voudrais ici mettre en perspective le tableau que je viens de décrire à l'aide de plusieurs cadres théoriques situés à des niveaux ou des échelles d'analyse différents. Soulignons qu'il s'agit essentiellement d'une réflexion sur leur potentiel heuristique, et que ces pistes devraient être approfondies dans le cadre d'un programme de recherches à part entière.

L'apport des approches de la construction sociale des normes et dispositifs professionnels

Une première piste, que je développerai davantage dans la mesure où elle fournit un cadre particulièrement structurant pour penser les tensions évoquées jusqu'ici, consiste à explorer les mécanismes de régulation des pratiques en matière d'anonymisation. L'anonymisation des données représente une activité éminemment normative, qui se déroule dans un cadre organisationnel complexe, mais rarement analysé sous cet angle. La théorie de la régulation sociale de Jean–Daniel Reynaud offre, de ce point de vue, un éclairage particulièrement fécond[38].

Développée pour étudier la production des règles dans les organisations, elle invite à concevoir la régulation non

37 Notons que cet ouvrage, et le projet plus large qui l'anime, pourraient sans doute être rapprochés de la version plutôt souple des « bonnes pratiques ».

38 Cf. Reynaud Jean-Daniel, « Les régulations dans les organisations », *art. cit*.

comme l'application descendante de normes formelles — qu'il s'agisse du RGPD, de principes déontologiques non codifiés ou de « bonnes pratiques » — mais comme un processus dynamique et négocié par lequel les acteurs définissent collectivement ce qui est faisable et acceptable en situation. Cette théorie paraît bien adaptée à l'analyse des activités d'anonymisation, qui confrontent durablement des groupes d'acteurs hétérogènes — chercheurs, curateurs, juristes, DPO, etc. — à des arbitrages permanents entre protection des enquêtés, qualité scientifique des matériaux, soutenabilité du travail de traitement et exigences de diffusion. Elle permet notamment d'envisager comment les règles d'anonymisation sont négociées, contournées, stabilisées dans la pratique, au sein d'une « anarchie organisée »[39], au sens où ces configurations se caractérisent par l'absence de relations hiérarchiques claires et de dépendances fonctionnelles nettement établies entre parties prenantes, contrairement à d'autres organisations analysées par Reynaud. Au cœur du cadre théorique de Reynaud se trouve la distinction entre régulation de contrôle, régulation autonome et régulation conjointe, qui éclaire la fabrique concrète des compromis et des règles locales.

Selon cette perspective, on peut d'abord s'interroger sur la réalité des formes de régulation de contrôle — qui renvoient aux règles formelles, procédures et normes définies, reprises ou appliquées par les organisations pour orienter et encadrer l'action. Il n'existe (quasiment) pas — ou du moins peut-on en douter sérieusement — de relation hiérarchique ou de dépendance fonctionnelle entre des autorités institutionnelles chargées de faire respecter des normes d'anonymisation précises et strictes et les acteurs chargés de les appliquer.[40]. Un pouvoir normatif peut toutefois exister dans quelques cas, notamment pour certaines enquêtes quantitatives[41], mais pour les données qualitatives, la diffusion reste largement[42] volontaire, négociée au cas par cas, ce qui laisse une marge de manœuvre importante aux déposants. Il nous est arrivé, à beQuali, de ne pas prendre en charge certaines enquêtes qui auraient nécessité une anonymisation extrêmement poussée, mais ce renoncement relevait toujours d'un accord mutuel plutôt que d'un refus unilatéral, pour éviter une charge de travail trop lourde, un risque d'identification trop élevé ou un appauvrissement jugé excessif des matériaux, notamment pour des analyses secondaires.

Il existe néanmoins des formes de contrôle exercées par les curateurs sur le travail d'anonymisation réalisé dans le cadre des autodépôts. Ce contrôle reste toutefois rarement exhaustif et porte plutôt sur des échantillons. Ce mécanisme est intrinsèquement lié à un enjeu de responsabilité. Bien que l'on puisse considérer que celle–ci repose principalement sur le chercheur déposant, principalement dépositaire du contrat de confiance initialement conclu avec les enquêtés, et qui s'engage à anonymiser les données comme obligation de résultat — c'est–à–dire à garantir une réduction suffisante du risque d'identification —, la responsabilité in fine repose malgré tout sur les entrepôts qui diffusent les données. On comprend ainsi que des entrepôts bien établis puissent imposer des règles strictes, allant parfois jusqu'à éviter les données trop sensibles ou les données qualitatives, afin de limiter le coût associé à ce contrôle.

La prégnance de cette régulation de contrôle dépend également de la force normative des cadres juridiques encadrant l'anonymisation. Or, les autorités institutionnelles se contentent le plus souvent de poser un cadre général, relativement abstrait — les règles juridiques elles–mêmes étant peu précises sur ce qui constitue une pseudonymisation réussie[43] —, tout en renvoyant aux communautés scientifiques la responsabilité de s'auto-organiser, sans que des normes réellement stabilisées n'émergent[44].

L'application de ce modèle invite ainsi à penser que la gestion concrète de l'anonymisation relève surtout de formes de régulation autonome — entendues comme l'ensemble des ajustements pratiques élaborés en situation par les acteurs, afin de rendre le travail possible dans des conditions concrètes. Notre expérience dans le projet beQuali, en écho à plusieurs témoignages d'autres centres de données à l'étranger[45], montre qu'il s'agit surtout de bricoler des arrangements avec des

39 Cf. Musselin Christine, « University Governance in Meso and Macro Perspectives », *Annual Review of Sociology*, vol.47, 2021, pp.305-325.

40 Les arbitrages formels en cas de manquement relèvent d'instances comme la CNIL ou les comités d'éthique, qui restent largement en marge des situations où se règlent au quotidien les problèmes d'anonymisation. La CNIL est souvent perçue comme distante de la gestion ordinaire de l'ouverture des données de la recherche. Les comités d'éthique interviennent surtout en amont des recherches ; leur rôle, bien que discret, est cependant susceptible d'orienter les pratiques en la matière, en instaurant une pression à la conformité aux standards de protection des données — mais cette influence reste encore, à ma connaissance, peu documentée en France.

41 Par exemple comme l'a montré l'enquête menée dans le cadre du projet UPMET, dans un certain nombre d'entrepôts étudiés en Europe ou ailleurs, la contrainte semble surtout effective lorsque la valeur ajoutée du dépôt est élevée pour les chercheurs déposants, comme dans le cas des grandes enquêtes quantitatives, notamment longitudinales, souvent soumises à des obligations institutionnelles de dépôt ou à de fortes attentes de la communauté scientifique.

42 La situation est différente dans un certain nombre de pays, comme le Royaume-Uni ou les Pays-Bas, qui ont mis en place des dispositifs d'obligation au dépôt des données pour tous les types d'enquêtes (y compris donc qualitatives) correspondant à des recherches financées sur projets à partir de fonds publics.

43 Cf. Bendjaballah Selma, Garcia Guillaume & Sauger Nicolas, « Protéger les enquêtes, mais à quelles conditions ? », *art. cit*.

44 Voir, par exemple : Marin Dacos, « Ouvrir les données de la recherche sur les sociétés contemporaines », intervention Atelier Digit_Hum 2022 [en ligne].

45 Les témoignages recueillis dans l'enquête UPMET indiquent plus largement que les curateurs prennent fréquemment des libertés avec les standards ou les bonnes pratiques officiellement reconnues, afin de résoudre des situations concrètes (cLee Dong Joon, Stvilia Besiki, « Practices of research data curation in institutional repositories : A qualitative view from repository staff », *PLOS ONE*, n°3, vol.12, 2017).

règles juridiques souvent trop générales pour fournir des solutions opératoires concernant le bon niveau de pseudonymisation. Ce qui se joue relève davantage de négociations fines entre chercheurs et curateurs, visant à arbitrer entre des risques hétérogènes, souvent non commensurables, au prix de l'acceptation d'un risque résiduel jugé faible — ce qui correspond à un « calcul de risque prenant en compte l'utilité »[46].

L'application de cette théorie invite enfin à réfléchir aux conditions d'émergence de formes de régulation conjointe — entendues comme des configurations issues de l'interaction entre régulation de contrôle et régulation autonome. La régulation conjointe caractérise des compromis relativement stabilisés par lesquels normes formelles et contraintes pratiques s'alignent, de manière progressive, partielle et réversible. D'après le modèle de Reynaud, c'est surtout cette forme de régulation qui est susceptible de donner lieu à des équilibres relativement satisfaisants pour l'ensemble des parties prenantes. La mise en œuvre de telles régulations conjointes suppose l'existence de dispositifs permettant de négocier et élaborer des règles communes entre parties prenantes. Or, les dispositifs institutionnels soutenant ces régulations — espaces de discussion sur les pratiques, instances de négociation ou de délibération professionnelle — restent encore très peu développés dans le champ académique. Ce sont donc les arènes scientifiques ordinaires qui devraient, en principe, jouer ce rôle, même si elles intègrent encore peu ces enjeux. Pour ne prendre que quelques exemples, les deux associations professionnelles centrales dans les disciplines couvertes par beQuali, l'AFS et l'AFSP ne disposent pas de groupes de travail dédiés à la science ouverte, et ces questions n'ont jamais véritablement été inscrites à l'agenda de leurs congrès, même comme thèmes transversaux émergents. Face à ce manque, nous avions tenté en 2015 de créer un espace de discussion au sein du comité scientifique et technique de beQuali, dans le cadre de l'Équipex DIME-SHS, via un groupe de travail ad hoc. Il s'agissait de dépasser le cadre interne de l'équipe où se discutaient ces questions à une échelle très micro[47]. Cette initiative est restée limitée — en raison du faible nombre de participants (une dizaine), du corpus restreint d'enquêtes mobilisées (trois), et du fait qu'il ne s'agissait pas d'un véritable dispositif d'analyse des pratiques — mais elle a néanmoins montré l'intérêt de confronter des points de vue situés sur la question.

Un second cadre analytique, complémentaire du premier, est fourni par l'approche des objets frontières, initialement développée pour étudier la coordination du travail et la délégation de l'activité scientifique[48]. Cette approche, qui se situe au niveau de l'interface socio-technique de ces configurations, permet de mieux saisir, au–delà des questions de régulation, le destin des objets en circulation, leurs appropriations, ainsi que les formes de dialogue entre groupes professionnels mobilisant des critères de qualité différents. Elle invite à considérer les matériaux, les protocoles et normes d'anonymisation, ainsi que les guides et tutoriels, comme des objets circulant entre des mondes professionnels hétérogènes — chercheurs, curateurs, archivistes, juristes, informaticiens — porteurs de finalités et de définitions de la qualité de l'anonymisation a priori non homogènes (défense du potentiel sociologique des données, intégrité du document, stricte application du droit ou de l'éthique, etc.) ; elle permet de penser comment « l'anonymisation » peut être comprise différemment par les parties prenantes, mais rester coordonnée entre ces dernières. Sous cet angle, les guides d'anonymisation peuvent être compris comme des dispositifs de traduction volontairement généraux et parfois ambigus, rendant possible une coordination minimale sans effacer les divergences entre parties prenantes. Dans cette perspective, l'ambiguïté des règles d'anonymisation — par exemple dans les guides destinés aux déposants — apparaît autant comme un défaut que comme une condition de possibilité de la coopération. Cette difficulté tient en fait à plusieurs facteurs : l'impossibilité de standardiser des protocoles face à la singularité des enquêtes ; la difficulté d'exemplifier sans exposer des informations sensibles (problème de rétro–ingénierie) ; le fait que la « boîte noire » des arbitrages concrets est le plus souvent traitée dans le face–à–face entre centres de données et déposants selon des logiques ad hoc ; et enfin sans doute une prudence compréhensible, l'anonymisation restant un terrain de débats éthiques et épistémologiques parfois vifs[49].

Ce cadre théorique invite ainsi à penser les tensions autour de l'anonymisation non pas seulement comme des conflits entre métiers supposés homogènes et étanches, mais comme des disputes internes à un espace professionnel hybride, où les acteurs mobilisent alternativement des registres scientifiques, juridiques, documentaires ou techniques selon les situations. Même si les formations et positions professionnelles

46 Lire : Silberman Roxane, « Partage des données et protection de la confidentialité : quels nouveaux enjeux pour les sciences sociales ? », *L'Année sociologique*, n°2, vol.74, 2024, pp.529-559.

47 Cf. Cadorel Sarah, Garcia Guillaume, Groshens Émilie, Juillard Émeline, Vandenbunder Jérémie, Le Courant Stefan, Pecqueux Anthony & Noûs Camille, « Archiver, documenter, enquêter sur l'enquête qualitative », *art. cit.*

48 Cf. Star Susan Leigh & Griesemer James R., « Institutional ecology, "Translations", and Boundary objects », *art. cit.* Voir aussi : Trompette Pascale, Vinck Dominique, « Retour sur la notion d'objet-frontière », *Revue d'anthropologie des connaissances*, n°1, vol.3, 2009, pp.5-27. Cette approche a déjà été mobilisée pour analyser les banques de données : Meyer Morgan, « Objet-frontière ou Projet-frontière ? : Construction, (non-)Utilisation et politique d'une banque de données », *Revue d'anthropologie des connaissances*, n°1, vol.3, 2009, pp.127-148.

49 Diaz Pablo, « Introduction : Archiving Qualitative Data in Practice. Ethical Feedback », *Bulletin of Sociological Methodology / Bulletin de méthodologie sociologique*, n°150, 2021, pp.7-27.

de ces groupes peuvent être éloignées, la mobilisation de ces registres est souvent plus pluraliste qu'on ne le supposes. On est peut–être en effet confronté à de nombreux discours tendant à réifier l'homogénéité de catégories professionnelles pourtant hétérogènes — chercheurs, DPO, ingénieurs, curateurs — alors qu'elles sont traversées par des tensions internes et des rapports de force rarement visibles de loin[50]. Par exemple, le fait que de nombreux curateurs (bien au–delà du seul cas de beQuali) aient une formation en sciences sociales, parfois un doctorat, brouille la frontière entre monde de la curation et monde de la recherche[51].

Enfin, située à un niveau macro-professionnel, la sociologie des professions d'Andrew Abbott[52] semble particulièrement utile pour penser les luttes de juridiction entre groupes professionnels (chercheurs, curateurs de données, juristes, DPO, etc.) cherchant à imposer leur définition légitime du problème (juridique, technique, éthique, scientifique). Cette piste ne sera ici qu'esquissée, mais elle mériterait à elle seule un développement spécifique. Cette approche éclaire les rapports de pouvoir entre groupes professionnels, qui irriguent les disputes autour de l'anonymisation des données d'enquête, au sujet de qui a le droit de définir et traiter ce problème. Ces disputes se déploient dans différentes arènes, où se stabilisent provisoirement des compromis entre professions concurrentes.

Une approche complémentaire : l'épreuve individuelle de l'anonymisation

À une échelle plus individuelle, une autre piste consiste à examiner les épreuves concrètes auxquelles sont confrontés les praticiens de l'anonymisation. Une des caractéristiques centrales de ces situations est que la responsabilité en matière d'anonymisation est souvent individualisée — elle repose sur les chercheurs ou les centres de données, voire sur les curateurs à titre individuel — tandis que le risque demeure collectif, diffus et mal défini. Il s'agit d'une activité exposant à des erreurs, dont les conséquences peuvent être potentiellement importantes pour les enquêtés, mais aussi par ricochet pour les chercheurs, ce qui fait peser une responsabilité lourde sur celles et ceux qui la réalisent. Dans ce contexte d'incertitude, la sur–anonymisation ou l'évitement des cas sensibles constituent des options répandues, même s'il est difficile d'objectiver avec précision leur saillance ; ces options apparaissent moins comme le résultat d'une dérive individuelle que comme une réponse rationnelle à un cadre normatif incertain. Cette situation est de nature à renforcer une culture de la science ouverte où l'anonymisation est vue a priori comme un impératif peu discutable.

Il me semble que cette culture est globalement influencée par une logique plus large de dramatisation des risques[53]. La possibilité — pourtant faible dans des régimes d'accès contrôlé et de pseudonymisation renforcée — de réidentification est ainsi souvent [54] érigée en menace centrale (autour du principe d'inférence « confirmatoire » ou de « recherche ciblée »)[55], sans que la sensibilité effective des données ni les dommages potentiels puissent, dans la plupart des cas, être véritablement objectivés, du fait même de la difficulté à les évaluer précisément.

En référence aux travaux d'Erving Goffman ou d'Howard Becker, il paraît heuristique d'analyser la manière dont certains acteurs, notamment ceux chargés de ces tâches, peuvent être amenés à adhérer publiquement à cet impératif de manière relativement stricte, non pas nécessairement parce qu'ils les jugent pleinement légitimes, mais parce que la non–conformité est perçue comme risquée, notamment sur le plan réputationnel et professionnel[56]. Il faudrait ainsi examiner l'hypothèse selon laquelle certains acteurs peuvent être conduits à se conformer publiquement à ces normes d'anonymisation, afin d'éviter d'être étiquetés comme déviants ou non professionnels — ce qui correspond à une posture d'hyperconformité préventive. Autrement dit, ces normes s'imposeraient, dans une certaine mesure, indépendamment des situations empiriques, leur légitimité étant présupposée plutôt que discutée, leur remise en question exposant à une disqualification professionnelle, voire morale, au nom de l'adhésion aux valeurs de la science ouverte ou du respect du cadre éthico-juridique.

Un apport important à cet égard des travaux de Howard Becker et Erving Goffman, et plus largement de l'interactionnisme symbolique, est d'inviter à dis-

50 Un des acquis majeurs de la sociologie des organisations ou des institutions est d'avoir montré que ces dernières, contre une vision homogène ou « monolithique », sont traversées de tensions, de jeux de pouvoir entre sous-groupes, de désajustements entre règles formelles et pratiques réelles, et de sous-cultures multiples.

51 On pourrait multiplier les exemples ; sur le cas des DPO, voir : Louvion Alexis, « De l'empêcheur de tourner en rond au superhéros : le travail de mise en conformité des Data Protection Officers », *Droit et société*, n°121, 2025, pp.455-478.

52 Cf. Abbott Andrew, *The System of Professions*, *op. cit*.

53 Sous l'influence plus générale de l'exacerbation du principe de précaution. Sur ce point, voir par exemple : Zaccaï Edwin, Missa Jean-Noël (dir.), *Le principe de précaution. Significations et conséquences*, Bruxelles, Éditions de l'université de Bruxelles, 2000.

54 Cette posture est sans doute plus ou moins accentuée selon les sujets, thèmes ou objets d'enquêtes, les caractéristiques des populations enquêtées, les risques perçus de dommage pour ces derniers, etc.

55 Cf. Blien Uwe, Muller Walter & Wirth Heike, « L'anonymat des données individuelles : études des risques d'identification », *Sociétés contemporaines*, n°14-15, 1993, pp.23-42.

56 À titre très personnel, j'ai pu moi-même ressentir, à plusieurs reprises, un certain sentiment d'inachèvement, voire un malaise lié à l'impression diffuse de ne pas savoir si je faisais bien ou mal, dans un contexte dépourvu de repères stabilisés.

tinguer façade publique et pratiques effectives. Il est ainsi possible de respecter formellement les normes officielles de la science ouverte tout en développant, en coulisses, des arrangements pragmatiques avec les règles. Ces accommodements concernent sans doute une partie non négligeable des opérateurs de l'anonymisation ou des chercheurs déposants, susceptibles d'adopter des formes de conformité apparente ou d'hypocrisie normative[57]. Il ne s'agit ici que d'hypothèses, tant il est difficile de mobiliser pleinement ce cadre sans enquête dédiée. Mais au fil des années, des rencontres et des discussions informelles, mon impression est d'avoir observé à plusieurs reprises, auprès d'acteurs différents, des flottements, des malaises ou des tensions qui pourraient être analysés à travers ce prisme. Comme le souligne Becker, il peut exister, dans un domaine donné d'activité, des acteurs pleinement conformes, mais l'intérêt analytique de cette approche réside ici dans cette zone intermédiaire, où les alignements normatifs ne sont ni évidents ni homogènes.

Analyser sociologiquement les pratiques d'anonymisation des publications

Une dernière piste d'analyse consiste à explorer les normes disciplinaires à l'œuvre dans l'anonymisation des publications scientifiques. Cette perspective fait écho à l'analyse des rapports de pouvoir à l'œuvre dans la négociation de l'anonymisation avec les enquêtés, qui engagent la position relative du sociologue[58]. Il s'agit ici d'examiner les modalités et la variabilité des normes scientifiques sous–jacentes aux régimes de preuve, afin d'éclairer ce qui se joue, souvent de manière invisible, dans l'anonymisation des données.

Explorer cette question conduit à s'intéresser aux pratiques mises en œuvre dans les publications, qui participent plus largement à définir ce qui doit être conservé pour faire preuve — un référentiel susceptible de varier selon les disciplines, les contextes nationaux, ou encore d'évoluer dans le temps[59]. Or, il existe très peu de travaux dans cette perspective, en particulier en langue française, adoptant une approche à la fois transdisciplinaire, comparative et longitudinale. C'est précisément l'objectif du projet ETAPSS[60], qui vise à étudier la manière dont les enjeux de confidentialité et d'anonymisation sont traités dans l'écriture des articles publiés en sciences sociales. Il s'agit de mettre au jour et d'interroger les normes — qu'elles soient explicites ou qu'elles irriguent plus implicitement la production des articles scientifiques — dans un large spectre disciplinaire : anthropologie, économie, épidémiologie, géographie sociale, linguistique, science politique, sciences de l'information et de la communication, sociologie, démographie.

Empiriquement, cette démarche suppose un examen à grande échelle des pratiques, celles de milliers d'auteur·es, dans un domaine où les prescriptions sont rarement explicitées ou enseignées, et principalement transmises par socialisation diffuse. Le projet s'appuie notamment sur l'analyse d'articles publiés dans des revues à comité de lecture, envisagées comme des collectifs professionnels relativement stables, vecteurs centraux de la socialisation académique, de la validation des critères scientifiques et cibles privilégiées des politiques de science ouverte. Il adopte également une perspective comparative internationale — le cas de la France étant mis en regard de ceux du Royaume–Uni, des États–Unis et du Canada — afin de mieux saisir la spécificité des configurations nationales et leurs évolutions éventuelles. Cette approche intègre en outre une dimension longitudinale, en remontant au moins jusqu'au début des années 2000, afin d'identifier d'éventuels effets des dynamiques, plus ou moins précoces, de la science ouverte sur ces pratiques.

CONCLUSION

Ce texte a proposé une réflexion sur l'anonymisation des données qualitatives à partir d'un retour d'expérience situé, principalement fondé sur le cas de beQuali et, plus largement, sur celui des entrepôts de données en sciences sociales. Il met en évidence le caractère central, mais problématique de l'anonymisation dans les dispositifs de diffusion élargie de données, y compris sous accès contrôlé. Loin de relever de procédures stabilisées, l'anonymisation apparaît comme une activité d'arbitrage située, incertaine et coûteuse, qui engage des compromis entre protection des enquêtés, valeur scientifique des matériaux et soutenabilité du travail. Ces contraintes contribuent à limiter concrètement le développement du partage de données qualitatives, en favorisant les enquêtes les plus faciles à anonymiser et en laissant souvent de côté les terrains les plus sensibles ou les plus fortement

57 Cf. Goffman Erving, *La mise en scène de la vie quotidienne*, Paris, Minuit, 1973.

58 Sur ce point, voir : Roux Sébastien, « La transparence du voile. Critique de l'anonymisation comme impératif déontologique », *in* Sylvain Laurens, Frédéric Neyrat (dir.), *Enquêter de quel droit ? Menaces sur l'enquête en sciences sociales*, Bellecombe-en-Bauges, Éditions du Croquant, 2010, pp.137-151. Voir également, plus récemment : Riou Camille. « Les enjeux de l'anonymisation dans une enquête auprès d'expert·e·s de la question des violences sexuelles », *art. cit.* Pour une première analyse longitudinale des pratiques des sociologues français, voir : Coulmont Baptiste, « Le petit peuple des sociologues Anonymes et pseudonymes dans la sociologie française », *Genèses*, n°107, 2017, pp.153-175.

59 Mais aussi selon les méthodes et objets d'enquête, les types de populations étudiés, etc.

60 Étudier le travail d'anonymisation dans les publications en sciences sociales (ANR-24-RESO-0004-01), que je coordonne, avec Selma Bendjaballah accessible sur : https://www.sciencespo.fr/cdsp/fr/projets/projets-en-cours/etapss/

contextualisés, notamment ethnographiques. Elles expliquent aussi sans doute, en partie, le décalage persistant entre les incitations croissantes au partage et le faible nombre de dépôts effectifs, en particulier au regard du volume d'enquêtes produites chaque année. Si des alternatives existent — circulation restreinte, dispositifs d'accès sécurisé —, elles ne suppriment pas complètement la question de l'anonymisation, mais en déplacent les modalités et les responsabilités. Dès lors que l'on vise une forme de diffusion, même encadrée, l'anonymisation demeure un point de passage obligé, sans solution pleinement satisfaisante ni stabilisée. Dans ce contexte, l'article invite à déplacer le regard : plutôt que de chercher à formaliser des « bonnes pratiques » universelles, il paraît plus heuristique de considérer l'anonymisation comme un ensemble de pratiques situées, à analyser dans leurs contextes d'élaboration. Cela suppose de développer des enquêtes empiriques approfondies sur ces pratiques, mais aussi de construire des espaces de discussion collective permettant de confronter les expériences et de rendre visibles les arbitrages à l'œuvre.