



Un protocole pour une enquête par questionnaire anonyme au sens du Règlement européen

Marie Plessz

► To cite this version:

Marie Plessz. Un protocole pour une enquête par questionnaire anonyme au sens du Règlement européen. Bulletin de Méthodologie Sociologique / Bulletin of Sociological Methodology, 2020, 145 (1/January 2020), pp.100-110. <10.1177/0759106319888706>. <hal-02463372>

HAL Id: hal-02463372

<https://hal.science/hal-02463372v1>

Submitted on 31 Jan 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.



HAL Authorization

Un protocole pour une enquête par questionnaire anonyme au sens du règlement européen

SOUMIS le 09/10/2019

Auteur : Marie Plessz

Affiliation : CMH (EHESS, ENS, CNRS), Inra

coordonnées : marie.plessz@inra.fr , CMH, 48 bd jourdan, 75014 Paris

Résumé

L'entrée en vigueur du Règlement général sur la protection des données personnelles (RGPD) dans l'Union européenne en 2018 a conduit de nombreux sociologues à s'interroger sur la concordance entre leurs pratiques d'enquêtes et le cadre législatif. Enseigner comment faire une enquête par questionnaire implique de former les étudiants à ces aspects déontologiques et juridiques des pratiques d'enquête. Aussi je me suis faite accompagnée par une juriste déléguée à la protection des données pour vérifier que la petite enquête par questionnaire (10 minutes) que je faisais réaliser chaque année par mes étudiants de master était conforme à la législation. Ce travail a débouché sur un protocole pour une enquête anonyme (par téléphone et auto-administré sur internet) qui est reproduit ici. Il est précédé d'une mise en contexte sur la législation sur les données personnelles ; sur la façon dont se déroule l'enquête pour laquelle j'ai conçu ce protocole ; enfin sur le choix d'anonymiser complètement l'enquête.

Comment citer ce document :

Plessz, M. (2020). Un protocole pour une enquête par questionnaire anonyme au sens du Règlement européen. Bulletin of Sociological Methodology/Bulletin de Méthodologie Sociologique, 145 (1/January 2020), 100-110. , DOI : 10.1177/0759106319888706

Un protocole pour une enquête par questionnaire anonyme au sens du règlement européen

Dans le cadre d'un master de sociologie quantitative, j'ai dispensé pendant trois ans une formation à l'enquête quantitative par questionnaire. Un aspect de cette formation est nécessairement la déontologie de l'enquête et le respect de la réglementation sur la protection des données¹. Concernant la protection des données, les premières années je me contentais d'assurer à mes étudiants que nous faisons une enquête anonyme, et qu'en conséquence nous étions en règle par rapport à la loi. Mais était-ce bien vrai ? Quand est entré en vigueur du règlement général sur la protection des données personnelles ou RGPD (Règlement européen 2016-679) en mai 2018, j'ai voulu faire le point. J'ai suivi des formations, échangé avec des Déléguées à la protection des données (DPD, auparavant conseillères informatique et liberté, CIL)², et suis parvenue à la réponse suivante :

- oui, si mon enquête est anonyme, je suis en règle car la loi ne s'applique pas ;
- mais l'anonymat au sens de la loi est une condition exigeante (Groupe de travail « Article 29 » sur la protection des données, 2014 ; InSHS *et al.*, 2019, 25). Pour que mon enquête soit anonyme, il faut qu'à aucun moment les répondants ne soient identifiables, ni directement, ni indirectement, y compris en croisant des réponses ou en utilisant des traces numériques (adresses IP...).

Assurer l'anonymisation des données au sens de la loi nécessite de scruter tout le processus de production, traitement et archivage des données d'enquêtes et non simplement le contenu du questionnaire. Dans mon cas, cela a débouché sur le protocole que je reproduis ci-dessous en le commentant.

Auparavant, j'apporte des éléments de contexte sur la législation sur les données personnelles d'une part ; sur la façon dont se déroule l'enquête pour laquelle j'ai conçu ce protocole d'autre part ; enfin sur le choix d'anonymiser complètement l'enquête.

Le RGPD et la recherche en France

Le RGPD encadre la collecte et le traitement des données personnelles dans tous les domaines (services publics, commerciaux, sur informatique ou sous format papier, à des fins lucratives ou non). Il définit des données personnelles comme « toute information se rapportant à une personne physique identifiée ou identifiable »³. Il reprend de nombreux aspects de la directive européenne 95-46 elle-même fortement inspirée par la législation française, la Loi informatique et libertés de 1978, ⁴.

1 La déontologie et le respect d'une loi peuvent être vus comme complémentaires. On pourrait imaginer une enquête conforme au RGPD et moralement discutable, sur le fond ou dans la manière dont elle est conçue. En outre, selon les sujets et terrains d'enquête, d'autres aspects de la protection des données peuvent entrer en jeu : secret des affaires (enquêtes sur des entreprises), sûreté de l'État, droit à l'image...

2 Je remercie tout particulièrement Nathalie Gandon, DPD de l'INRA, pour sa disponibilité et son dialogue constructif.

3 Définition et exemples sur le site web de la CNIL, <https://www.cnil.fr/fr/rgpd-de-quoi-parle-t-on> consulté le 03/10/2019. Voir aussi le guide publié récemment par le CNRS (InSHS *et al.*, 2019).

4 Une modification importante est l'inversion de la charge de la preuve, c'est-à-dire qu'il faut pouvoir rendre compte à tout moment du fait que l'on respecte la loi, ce qui implique par exemple de suivre la procédure interne proposée par l'établissement pour le traitement de données personnelles avant de commencer la collecte, mais

L'obligation faite aux chercheurs en sciences sociales de se mettre en conformité avec la loi sur les données personnelles n'est donc ni nouvelle, ni destinée à nous empêcher de faire de la recherche – bien au contraire, la recherche publique est par principe une base légale permettant de procéder à un traitement de données personnelles, contrairement à l'usage commercial des données.

Une enquête pour former les étudiants à la recherche en sciences sociales

L'enquête pour laquelle j'ai préparé ce protocole a lieu chaque année, sur un sujet différent (voir par exemple Zaoui, Bilbaut et al, 2016). Elle est conçue, testée, collectée et analysée par des étudiants de Master de sciences sociales, encadrés par les enseignant.e.s (Séverine Gojard, Anne Lhuissier et moi-même de 2015 à 2018, Raphaël Dhuot et moi-même pour l'année 2019-2020). Deux techniques de passation de l'enquête sont employées :

- les étudiants administrent le questionnaire par téléphone, en tirant au sort des numéros dans l'annuaire des pages blanches (accessible en ligne) et en complétant le questionnaire dans l'interface en ligne de Limesurvey ;
- en complément, nous achetions environ 400 réponses à une société d'enquêtes en ligne (*access-panel*) : les membres du panel reçoivent un lien personnalisé, se connectent à notre enquête et remplissent le questionnaire directement.

Je ne reviendrai pas sur les motivations pédagogiques (et les limites) de cette double stratégie. Je voudrais simplement souligner les convergences entre quelques principes de la loi et nos objectifs pédagogiques dans la formation à l'enquête par questionnaire. La loi encourage à recueillir des données personnelles pour des finalités bien définies (pas « au cas où ça servirait un jour ») et de façon proportionnée à ces finalités. Or une enquête par questionnaire doit anticiper le traitement statistique et l'analyse sociologique des réponses. Notre enquête est très courte (passation téléphonique de 10 à 15 minutes, passation sur internet environ 8 minutes), notre échantillon est limité (environ 600 réponses au total). Chaque question doit donc être pesée soigneusement : À quel aspect de notre problématique répond-elle ? allons-nous pouvoir en exploiter les réponses ? Les modalités de réponses proposées sont-elles pertinentes pour l'analyse, sur le plan sociologique et statistiques (éviter les modalités trop rares, les confusions qui mènent aux non-réponses, les questions ouvertes trop rarement exploitées...). Ces précautions vont dans le sens de la loi : ne pas collecter plus que nécessaire.

Le choix d'une enquête anonyme au sens de la loi

J'ai fait le choix de rendre l'enquête complètement anonyme. J'aurais pu assumer la collecte de données personnelles, suivre la procédure de déclaration interne de l'établissement accompagnant mon unité pour la mise en œuvre du RGPD⁵ plutôt qu'un protocole pour enquête anonyme et me mettre en conformité avec la loi (rédiger une note d'information, assurer la protection des données et leur archivage⁶ après une durée déterminée...). Deux raisons m'ont fait choisir la première option. Tout d'abord, nos étudiants passent beaucoup de temps à collecter les données, puis à les traiter sur leur ordinateur personnel, ils conservent en général une copie des données à l'issue du cours. Si ces données sont identifiantes, il y a là une incompatibilité avec la loi. Je ne me vois pas traquer mes étudiants pour les obliger à détruire toutes les copies de ces données qui seraient encore en leur possession après la fin des cours. Or le respect de la loi sur les données personnelles m'imposerait de

dispense de déclarer tous les traitements à la CNIL. On note aussi l'augmentation des pénalités financières en cas d'infraction : la loi française prévoyait jusqu'à 300 000 euros d'amende, le RGPD prévoit jusqu'à 4% du chiffre d'affaires mondial pour les infractions graves commises par des entreprises.

5 Par exemple pour les unités Inra il faut remplir le « questionnaire de contexte SSI-IL ». La plateforme Opidor facilite la rédaction d'un plan de gestion des données. <https://opidor.fr/> consulté le 03/10/2019.

6 La loi impose la destruction *ou l'archivage*. La destruction des données de la recherche ne peut être imposée en France car elle entrerait en contradiction avec le code du patrimoine. Ceci est rappelé dans le guide du CNRS mentionné plus haut.

ne pas laisser les étudiants emporter avec eux un fichier de données personnelles pour les réutiliser ensuite pour une autre finalité.

Ensuite les enquêtes que nous avons réalisées jusque-là étaient très proches d'être anonymes au sens de la loi. La perte d'information sociologiquement pertinente allait donc être très faible. Je m'en suis rendue compte quand je me suis demandée si j'étais en mesure d'assurer le droit d'accès, de rectification et de suppression des données personnelles de mes enquêtés : pour la quasi-totalité des répondants je suis incapable de les retrouver dans la base car je n'ai pas assez d'informations personnelles. Or il ne faut pas collecter plus de données personnelles seulement pour respecter les obligations légales. Mais cela ne suffit pas pour que l'enquête soit anonyme, du fait des risques de réidentification indirecte mentionnés plus haut. Par exemple, par recoupement d'information, certains répondants peuvent être identifiables tandis que d'autres ne le sont pas ; autre exemple, si les étudiants me remettent une liste des numéros de téléphones appelés avec des dates et heures d'appel, et que les dates et heures de passation sont enregistrées dans la base de données de l'enquête, l'enquête n'est plus anonyme.

Conclusion

Aujourd'hui, avant d'entreprendre un traitement de données personnelles (non anonyme) nous devrions établir un document regroupant tous les éléments permettant de montrer à tout moment notre conformité (principe d'*accountability*). Paradoxalement, pour m'assurer que je réalisais un traitement de données anonymes, j'ai été amenée à faire presque le même travail. En effet, pour s'assurer qu'un traitement est anonyme aux yeux de la loi, il faut scruter chaque étape du processus, il faut se pencher sur l'enquête du point de vue du contenu, de l'organisation, et des caractéristiques techniques des moyens informatiques mis en œuvre. Les sociologues connaissent de très bonnes techniques d'anonymisation leur déontologie et leur savoir-faire sont des ressources essentielles pour produire des enquêtes sociologiquement pertinentes et respectueuses de la loi, mais le dialogue avec un ou une DPD sont également essentiels.

À l'issue de ce processus il me semble qu'anonymiser au sens de la loi une enquête par questionnaire de dix minutes est atteignable sans grave compromis sur la qualité sociologique et statistique des résultats. Les impératifs pédagogiques (outils de suivi des passations, accès des étudiants aux données) ajoutent des contraintes, mais j'ai réussi dans la situation où je me trouvais, à trouver des solutions satisfaisantes. De toute évidence, plus l'enquête est détaillée, plus les risques de réidentification sont importants. Il en va de même si l'enquête porte sur un bassin de répondants réduit⁷.

Faire une enquête anonyme a un gros avantage dans un cadre pédagogique : les étudiants peuvent quitter la formation avec une copie des données, ce qui ne serait pas le cas si les données étaient identifiantes. Il ne faut toutefois pas surestimer les obstacles qui se dresseraient si l'on voulait collecter des données personnelles pour la recherche. En particulier, si les données ne sont pas sensibles (nous en donnons la définition plus bas), le consentement exprès n'est pas nécessaire pour collecter des données personnelles car la recherche publique constitue une base légale au traitement. Il faut bien sûr respecter les autres aspects de la loi : déclaration du traitement au DPD de l'établissement⁸, information des personnes (notice ou page web d'information), droit d'accès, rectification et suppression, stockage sécurisé, respect de la finalité, de la proportionnalité et de la durée prévue pour le traitement, archivage en fin de traitement. Les difficultés émergent vraiment dans le cadre d'enquêtes ethnographiques en immersion, a fortiori si le sujet de l'enquête est une « donnée sensible » par nature (positions politiques, santé, sexualité, religion, syndicalisme...) ⁹. Il me semble

7 Il peut s'agir d'un critère géographique ou d'une autre caractéristique qui serait un critère d'inclusion (ex : enquête auprès de mères aveugles).

8 Dans une UMR c'est le DPD de l'institution de rattachement du Directeur d'unité (DU) si celui-ci n'a pas désigné un autre DPD. Il convient donc de se renseigner auprès de son DU.

9 Les enquêtes ethnographiques sont absentes du guide du CNRS. D'après mes échanges avec des personnes impliquées dans la mise en place du RGPD pour la recherche au CNRS, la réflexion est en cours au moment de la

que pour une telle enquête, il est essentiel de dialoguer le plus tôt possible avec un.e délégué.e à la protection des données.

rédaction de cet article. Il me semble important que les chercheurs et chercheuses soient impliquées dans cette réflexion afin de la rendre aussi pertinente que possible par rapport aux enjeux de l'enquête ethnographique.

Comment citer ce document :

Plessz, M. (2020). Un protocole pour une enquête par questionnaire anonyme au sens du Règlement européen. *Bulletin of Sociological Methodology/Bulletin de Méthodologie Sociologique*, 145 (1/January 2020), 100-110. , DOI : 10.1177/0759106319888706

Protocole pour une enquête par questionnaire anonyme

Préambule

Le cours « Enquête par questionnaire » fait partie du tronc commun du parcours « Quantifier en sciences sociales » du master Sciences sociales (PSL-EHESS-ENS) (spécialité Sociologie et statistique du master Sociologie générale de l'EHESS jusqu'en 2019). Dans ce séminaire, chaque année les étudiants encadrés par leurs enseignants construisent un questionnaire pour répondre à une problématique sociologique, et administrent ce questionnaire par téléphone en tirant au sort des numéros de téléphone dans l'annuaire téléphonique. Ils réalisent ensuite des analyses statistiques pour répondre à leur problématique de recherche sociologique. Cette collecte de données a donc un double objectif, pédagogique (comment faire une enquête par questionnaire correcte du point de vue sociologique, statistique et de la protection des données personnelles) ; et scientifique puisqu'elle vise à éclairer une question de recherche définie d'après les travaux sociologiques existants.

Ce document présente les procédures suivies et précautions prises pour que ce traitement de données soit anonyme, c'est-à-dire qu'il ne soit pas un traitement de données personnelles, définies comme directement ou indirectement identifiantes.

Ce document a été rédigé par Marie Plessz, avec l'aide de la Déléguée à la protection des données de l'INRA, Nathalie Gandon¹⁰.

Construction du questionnaire

Les enseignants et étudiants définissent une problématique, les thématiques du questionnaire et listent les informations dites « sociodémographiques » à recueillir, selon leur pertinence pour le sujet (par exemple le sexe). Ils formulent ensuite chaque question et les modalités de réponse qui seront proposées au répondant.

Comme beaucoup d'enquêtes sociologiques le questionnaire recueille des « informations sociodémographiques ». Le risque d'identification y est plus important. Des précautions sont prises pour assurer à la fois la pertinence sociologique des informations recueillies et l'impossibilité d'identification directe ou indirecte :

- Aucune information directement identifiante n'est recueillie (par exemple nom, prénom, adresse)
- Pour les informations pouvant être indirectement identifiantes on s'assure de les recueillir dans des catégories suffisamment larges pour éviter l'identification indirecte. On ne recueille aucun numéro dit de référence (ex. numéro de sécurité sociale).

Un numéro d'enquête ou un pseudonyme que l'on retrouverait dans une table de correspondance nominative serait aussi identifiant. La pseudonymisation n'est pas l'anonymisation.

- Enfin les enquêtes portent sur la France entière, ce large bassin d'enquête réduit le risque d'identification indirecte par croisement des informations sociodémographiques.

En complément du recueil des informations dans des catégories relativement larges, élargir ou « flouter » le bassin d'enquête renforcent l'anonymat des données. Exemple : dire que l'enquête a porté sur des départements français sans préciser lesquels, voire sur « la population résidant en France ». En effet une information précise (profession : gynécologue-obstétricien) n'est pas identifiante ramenée à la France entière mais peut l'être ramenée à une petite localisation géographique.

¹⁰ Figurent ici, dans le protocole original, les coordonnées professionnelles de la DPD.

Le tableau ci-dessous synthétise les précautions à prendre lors du recueil des informations sociodémographiques :

Notion visée	Interdit car identifiant	Éviter car risque d'identification indirecte	Préférer
Âge	Date de naissance	Année de naissance/âge en années	Classe d'âge en tranches
Résidence	Adresse exacte	Nom de la commune, code postal, département	Région, type de commune ou d'habitat
Position professionnelle		Intitulé exact de la profession Prudence avec la ligne « autre : précisez » : risque de recueillir des professions rares et identifiantes	Groupe professionnel (PCS)
Genre		Modalité « autre... » : ne pas demander de préciser. ou préférer « ne souhaite pas répondre ».	Si variable à 2 modalités (sexe) : Pas de problème particulier
Composition du ménage		Nombre exact d'enfants	Regrouper les nombres élevés d'enfants (4 et plus ...)

Pour les variables portant sur la thématique d'intérêt de l'enquête, les modalités de réponses proposées visent à obtenir des effectifs suffisants pour chaque modalité afin de permettre les analyses statistiques. On veillera donc à éviter des modalités trop spécifiques, à la fois du point de vue des analyses statistiques et de l'anonymat des réponses.

Au fil de mon travail avec la DPD j'ai appris qu'une information en apparence anecdotique mais précise pouvait être identifiante, même si elle ne peut pas être recoupée avec d'autres sources d'information publiques. Par exemple si un répondant indique qu'il s'est fait tatouer « Paradoxe mon amour » sur l'avant-bras, il peut devenir identifiable. Le risque d'identification s'étend en effet au cas où un enquêté se reconnaît dans l'enquête – à condition que les éléments identifiants ne conduisent qu'à une seule personne dans le bassin de l'enquête. En l'occurrence, peut-il y avoir deux personnes en France avec un tel tatouage ?

Données sensibles/particulières : par défaut, le protocole ne pose pas de questions sur les données sensibles.

La loi interdit, par principe, la collecte de données dites sensibles dans le cadre de traitement de données personnelles. Ces données sont des données « qui révèlent l'origine raciale ou ethnique d'une personne physique, ses opinions politiques, ses convictions religieuses ou philosophiques ou son appartenance syndicale, ainsi que les données génétiques, biométriques aux fins d'identifier une personne physique de manière unique, les données concernant la vie sexuelle ou l'orientation sexuelle, les données de santé ». Ces données sensibles sont parfois pertinentes pour certains thèmes de recherche sociologique, voire elles sont le thème de recherche.

Des dérogations à cette interdiction peuvent être accordées par exemple si la personne consent à cette collecte ou si le traitement est nécessaire à des fins de recherche scientifique (conformément à l'article 89 du RGPD).

Quand il s'agit de recueillir des variables « auxiliaires » (et non des informations très précises) des formulations de questions et des modalités de réponses non sensibles sont souvent possibles. Exemple de question : « comment décririez-vous votre état de santé sur une échelle de 1 à 8 ? » plutôt que « souffrez-vous d'une des pathologies

suivantes ? ». Exemple de modalité de réponses : « pour raisons religieuses » plutôt que « parce que je suis catholique ».

Si des informations sensibles sont fournies par les répondants dans des questions ouvertes, il faudrait de les reformuler dans des catégories plus larges (pour raisons religieuses, pour raison de santé).

Collecte des données par téléphone

Chaque étudiant reçoit des numéros de département géographique assortis d'un numéro généré aléatoirement. Il ouvre l'annuaire du département (disponible en ligne) à la page correspondant au numéro aléatoire. Il appelle le premier numéro de téléphone de cette page, puis le 10^{ème}, puis le 20^{ème}, etc. Chaque fois qu'un répondant accepte de participer à l'enquête, l'étudiant entre les réponses sur la plateforme Limesurvey (voir plus bas). Les étudiants reçoivent en outre un numéro d'enquêteur, qu'ils renseignent pour chaque questionnaire qu'ils enregistrent dans la base.

Pour que l'enquête reste anonyme, aucun lien ne doit être possible entre les noms et numéros de téléphone présents dans l'annuaire d'une part, et les données récoltées dans Limesurvey d'autre part.

Le protocole suivant est proposé :

- l'étudiant peut noter les numéros avant de les appeler. Il ne note aucun nom.
- Si personne ne décroche à l'appel, l'étudiant peut noter le numéro de téléphone afin de renouveler l'appel (aucune réponse à l'enquête n'a été collectée), jusqu'à ce que quelqu'un décroche;
- Si quelqu'un décroche, à l'issue de l'appel l'étudiant doit effacer le numéro et simplement noter le résultat (refus ou réponse au questionnaire).
- quand il cesse la collecte des réponses au questionnaire, l'étudiant note simplement le nombre d'appels selon leur résultat (non décroché, rappelé et non décroché, refus, accepté).
- L'étudiant efface alors tous les numéros de téléphone (destruction des fichiers papier ou numériques).
- Pour l'évaluation des taux de non-réponse, l'étudiant transmet uniquement les nombres d'appels passés selon leur résultat.

Avant d'élaborer ce protocole, je demandais aux étudiants de m'envoyer la liste des numéros de téléphone appelés, afin d'évaluer les étudiants. En effet je tiens compte du nombre de questionnaires obtenus (en tenant compte des difficultés linguistiques des étudiants étrangers) et d'appels passés. Il y avait là un risque d'identification (beaucoup d'étudiants notaient les jours où ils avaient appelé et leur numéro d'enquêteur figure dans l'enquête). En rédigeant ce protocole j'ai réalisé que la liste des numéros ne m'est pas utile. Mon principal critère de notation est le nombre d'appels passés. Si un étudiant voulait tricher sur le nombre d'appels, il pourrait aussi inventer une liste de numéros de téléphone.

Il arrive que les étudiants doivent corriger des erreurs de saisie (ex : en fin de questionnaire le répondant veut modifier une de ses réponses). Seuls un à trois étudiants (et l'enseignant) ont accès à la base de données pour la modifier, et nous désactivons le retour en arrière dans le questionnaire car cette fonctionnalité génère parfois des bugs. Voici la procédure retenue pour corriger des erreurs. L'étudiant note l'heure et la date de l'appel. Il transmet ces informations et son numéro d'enquêteur à l'enseignant. Celui-ci peut alors retrouver et corriger la ligne correspondant à cet entretien téléphonique, tout en conservant l'anonymat de l'enquête.

Par définition un étudiant n'a contacté qu'un répondant à une heure donnée. Comme c'est l'étudiant et non le répondant qui se connecte à l'interface Limesurvey, le répondant ne peut pas être identifié par son adresse IP.

Collecte des données sur internet

Dans le cas de la passation sur Internet, le répondant se connecte directement à l'interface Limesurvey.

Jusqu'à présent la passation a été sous-traitée à une entreprise de sondage disposant d'un panel internet, ou *access panel*. Les membres du panel de l'entreprise reçoivent un lien pour répondre à l'enquête dans l'interface Limesurvey. Ce lien contient un numéro qui permet à l'entreprise de récompenser ses panelistes mais ne permet pas au commanditaire (nous) de les identifier. Ce numéro transite sur nos serveurs le temps de la collecte mais ne sera pas conservé à l'issue de la collecte et ne figurera pas sur les fichiers de données qui seront exportés de limesurvey. Le prestataire s'assure de la conformité de ses traitements vis-à-vis du RGPD (gestion du panel, protection des données, droit d'accès...). Et de notre côté, nous considérons que ce numéro transitant temporairement sur les serveurs informatiques et n'étant pas exporté avant sa destruction, les moyens pour réidentifier un membre du panel ne sont plus raisonnablement susceptibles d'être utilisés (cf. considérant 26 du RGPD).

Ce numéro pourrait présenter des risques de réidentification indirecte parce que le prestataire possède une table de correspondance entre numéro et membres du panel.

Dispositif informatique de collecte et sauvegarde des données

Pour préserver l'anonymat on ne collecte pas les adresses IP ou d'autres informations sur le système informatique du répondant ; ce faisant on bloque la possibilité pour les répondants de revenir et compléter le questionnaire plus tard. Le questionnaire étant bref, cela ne pose pas de difficulté.

Le logiciel Limesurvey est installé sur un serveur sécurisé. Dans le cas de l'installation INRA l'accès est protégé par un identifiant et un mot de passe fourni par l'Institut (protocole LDAP). Dans le cas de l'installation EHES, l'accès est protégé par un identifiant et un mot de passe fourni par l'établissement.

La DPD m'a informé que les services informatiques de l'INRA conservent les adresses IP, heures de connexion et URL de connexion de toutes les personnes qui se connectent à l'installation Limesurvey de l'INRA (on peut donc dire quelle adresse IP s'est connectée à mon enquête, à quelle heure). Il y a là un risque de rupture d'anonymat. La DPD a alerté les services informatiques de l'INRA sur ce problème pour voir si un service d'enquête sans collecte de l'adresse IP pouvait être proposé. En attendant, voici la position adoptée :

le risque de réidentification par exploitation de l'adresse IP est très faible dans notre cas, sachant que, de plus, nous n'avons pas accès aux journaux de connexion de la DSI. En m'appuyant encore une fois sur le considérant 26, je considère l'enquête comme anonyme.

L'enseignant (ou deux enseignants) responsable du séminaire a accès à LimeSurvey. Un à trois étudiants reçoivent également des identifiants pour pouvoir créer le questionnaire et participer au paramétrage de la plateforme et à l'extraction des données. Les autres étudiants n'ont pas accès aux données permettant de paramétrer l'enquête sur LimeSurvey.

Analyse des données

Conçue de cette manière, la base de données collectée dans Limesurvey n'est pas identifiante.

La base de données contenant les réponses aux questionnaires est exportée par un enseignant ou un des étudiants disposant d'un identifiant, et mise à la disposition des étudiants du séminaire dans un dossier partagé auquel les étudiants se connectent sur invitation de l'enseignant responsable.

Les étudiants peuvent copier cette base de données sur leur ordinateur personnel et la manipuler sans risque car l'identification des répondants est impossible. Ils peuvent conserver une copie des données de questionnaire à l'issue du séminaire, à des fins pédagogiques (pour pouvoir retravailler leurs programmes statistiques par exemple) ou pour pouvoir montrer à un futur employeur le travail accompli.

À l'issue du séminaire et après finalisation des travaux de restitution (éventuelle publication...) la base de données est archivée par l'enseignant responsable. Aucun numéro de téléphone ne peut être conservé par les étudiants, ni transmis aux enseignants.

Références

Groupe de travail « Article 29 » sur la protection des données, 2014 « Avis 05/2014 sur les Techniques d'anonymisation », Union européenne. https://www.cnil.fr/sites/default/files/atoms/files/wp216_fr.pdf, consulté le 03/10/2019.

InSHS, Isabelle André-Poyaud, Sandrine Astor, Fabrice Boudjaaba, Gaëlle Bujan, Béatrice Collignon, Frédéric Dubois, Emmanuel Kessous, Lionel Maurel, & Muriel Roger, 2019, « Les sciences humaines et sociales et la protection des données à caractère personnel dans le contexte de la science ouverte : guide pour la recherche ». https://inshs.cnrs.fr/sites/institut_inshs/files/pdf/guide-rgpd_2.pdf, consulté le 03/10/2019.

Règlement européen 2016-679, « Règlement (UE) 2016/679 du Parlement européen et du Conseil du 27 avril 2016 », Règlement relatif à la protection des personnes physiques à l'égard du traitement des données à caractère personnel et à la libre circulation de ces données, et abrogeant la directive 95/46/CE (règlement général sur la protection des données). <https://www.cnil.fr/fr/reglement-europeen-protection-donnees>, consulté le 03/10/2019.

Zaoui, Bilbaut, Francou, Oehmichen, Palin Sainte-Agathe, Reyes, Rossignol-Brunet, Viedrov, Villani, Yabuki, Gojard, Lhuissier, Plessz, 2016 « manger dehors : représentations et pratiques », Aliss notes n°2, https://www6.versailles-grignon.inra.fr/aliss/Media/Fichier/ALISS-Notes/AN2_MangerDehors_2016, consulté le 03/10/2019.