Traitement des données personnelles appliqué au domaine de la recherche

Selon la loi « Informatique et Libertés » du 6 janvier 1978 modifiée en 2018 et le règlement général sur la protection des données (RGPD)

Depuis la mise en vigueur du règlement général sur la protection des données, plus communément appelé “RGPD”, le traitement des données à caractère personnel fait l’objet d’une forte attention. Qu’elle se place dans le cadre d’un sondage en ligne ou d’une expérience in-situ, la collecte de ces données est courante dans de très nombreux domaines de recherche. Afin d’y voir plus clair sur les pratiques à adopter, voici un document recensant les informations nécessaires pour faire une collecte de données en toute légalité.

Table des matières

Définitions

Avant de se lancer dans la présentation des différentes lois il est important de définir les termes principaux :

Le traitement des données

Il y a "traitement" de données personnelles, lorsqu'un chercheur collecte des données personnelles (au cours d'un entretien, par exemple), mais aussi quand il se contente de consulter de telles données dans un document. Ainsi, sont des traitements de données personnelles : *"la collecte, l'enregistrement, l'organisation, la structuration, la conservation, l'adaptation ou la modification, l'extraction, la consultation, l'utilisation, la communication par transmission, la diffusion ou toute autre forme de mise à disposition, le rapprochement ou l'interconnexion, la limitation, l'effacement ou la destruction"* (art. 4 du Règlement européen du 27 avril 2016).

Les données à caractère personnel

Les données à caractère personnel sont toutes **les informations relatives à une personne physique identifiée**, ou qui peut être identifiée en croisant des données la concernant. Elles peuvent être le nom, le prénom, l'âge, le métier, une photo, l'adresse postale, l'adresse mail, le numéro de téléphone, la date de naissance, le numéro IP, tout numéro d'identification (ex : numéro de sécurité sociale), une empreinte digitale, des données de localisation, et tous "éléments spécifiques propres à son identité physique, physiologique, génétique, psychique, économique, culturelle ou sociale".

Les données sensibles

Parmi les données à caractère personnel il existe une catégorie dite "données sensibles" dont la collecte et le traitement sont d'ordinaire interdits. Cependant une dérogation pour la recherche scientifique existe. Les données sensibles sont toutes les

informations à caractère personnel qui révèlent l'origine raciale ou ethnique, les opinions politiques, les convictions religieuses ou philosophiques ou l'appartenance syndicale, ainsi que le traitement des données génétiques, des données biométriques aux fins d'identifier une personne physique de manière unique, des données concernant la santé ou des données concernant la vie sexuelle ou l'orientation sexuelle d'une personne physique.

La recherche scientifique

Comme nous venons de le voir, il existe quelques dérogations à la loi concernant le domaine de la recherche scientifique. En voici la définition :

"Aux fins du présent règlement, le traitement de données à caractère personnel à des fins de recherche scientifique devrait être interprété au sens large et couvrir, par exemple, le développement et la démonstration de technologies, la recherche fondamentale, la recherche appliquée et la recherche financée par le secteur privé. Il devrait, en outre, tenir compte de l'objectif de l'Union mentionné à l'article 179, paragraphe 1, du traité sur le fonctionnement de l'Union européenne, consistant à réaliser un espace européen de la recherche. Par «fins de recherche scientifique», il convient également d'entendre les études menées dans l'intérêt public dans le domaine de la santé publique." (art 159 du Règlement européen du 27 avril 2016)

Questions à se poser avant de collecter des données personnelles

La collecte de données à caractère personnel est très encadrée par la loi, voici les questions indispensables à se poser avant de commencer :

Finalité et objectif de la collecte des données

Quelle est l'objectif de ma collecte ? Qu'est ce que je veux savoir ? L'article 5 du RGPD prévoit que les données personnelles ne peuvent être collectées que pour des *"finalités déterminées, explicites et légitimes"* qui doivent en principe être définies en amont du traitement et être portées à la connaissance des personnes concernées (articles 13 et 14). Cependant, il n'est pas toujours possible de cerner entièrement la finalité du traitement des données à caractère personnel à des fins de recherche scientifique au moment de la collecte des donnée. Afin de respecter les droits des personnes concernées, il est indispensable de définir le domaine de recherche lié, même si ce domaine est large. Ainsi, les personnes concernées doivent pouvoir donner leur consentement uniquement pour ces domaines ou certaines parties du projet de recherche.

De quelles données ai-je besoin ? (Principe de minimisation)

À partir de l'objectif visé, il est important de se poser la question des données dont on a besoin. Ce principe est appelé "le principe de minimisation". Il concerne le fait de ne collecter que ce qui est nécessaire et seulement si c'est vraiment nécessaire. En effet, les données récoltées doivent être *"adéquates, pertinentes et limitées à ce qui est nécessaire au regard des finalités pour lesquelles elles sont traitées".* Le mieux est d'éviter de collecter trop de données à caractère personnel. Dans une enquête par

questionnaire, il faut toujours éviter de collecter la date de naissance exacte qui est directement identifiante. L'année ou même une tranche d'âge suffisent généralement pour l'analyse. De même, une information sur le lieu de résidence (nom de commune, code postal, ….) peut permettre indirectement une identification dans les cas où la population est très restreinte (très petites communes par exemple). Il faut se demander, dans ce cas, si une échelle plus globale (canton, département, région, iris, agglomération, …) peut convenir. Se posent les mêmes questions pour une dénomination très précise d'une profession.

Durée de conservation des données

Le RGPD prévoit à son article 5 que les données ne peuvent être conservées *"sous une forme permettant l'identification des personnes concernées"* que pendant *"une durée n'excédant pas celle nécessaire au regard des finalités pour lesquelles elles sont traitées"*. Cependant il existe une dérogation pour la recherche : "les données peuvent être conservées au-delà de la durée qui a été nécessaire pour atteindre la finalité de recherche (par exemple, au-delà de la durée d'un projet de recherche déterminé) du moment qu'elles sont ensuite conservées uniquement pour être utilisées à des fins de recherche."

Qui sera la personne en charge de ces données ?

Dans toutes les collectes, il est demandé de désigner un délégué à la protection des données (DPO) qui sera en charge de la protection des données à caractère personnel. Cette personne devra s'assurer de la bonne conservation des données et du respect des droits des participants.

Comment seront traitées les données ?

Afin de respecter les droits des sujets, il est indispensable de spécifier la manière dont les données seront traitées : analyse de vidéos, analyse de conversations enregistrées, traitement automatique de logs, fouille de données, traitement sur des serveurs externes (API)… Les traitements doivent être portés à la connaissance des participants.

Réglementation pour un traitement légal des données

La mise en place d'une collecte de données à caractère personnel doit respecter deux obligations : l'obligation d'informer les personnes concernées et l'obligation de protéger ces données.

Obligation d'informer les personnes concernées (en cas de collecte directe ou de réutilisation d'un corpus)

Le droit d'information est un des droits fondamentaux des usagers. Il est donc impératif que le responsable des données informe les personnes concernées de

l'existence du traitement de leurs données ainsi que la finalité de cette collecte et les droits dont ils disposent.

"Conformément à la loi « Informatique et Libertés » du 6 janvier 1978, vous disposez d'un droit d'accès, de rectification et d'effacement des données qui vous concernent. Pour l'exercer, adressez-vous à « ..... »."

Le consentement doit être obtenu après rappel des droits des participants. Nous verrons plus tard les informations nécessaires pour un consentement éclairé légal.

En cas de collecte :
Dans le cas d'une collecte de données (au cours d'un entretien, par exemple), c'est en pratique le chercheur qui fournira ces informations aux personnes concernées.

En cas d'utilisation d'un corpus :
Si le chercheur mène sa recherche dans des corpus existants, sans collecter les données auprès des personnes concernées, le chercheur est dispensé de contacter ces personnes pour les informer qu'il va traiter leurs données personnelles à des fins de recherche scientifique ou historique.

En cas de simple consultation de documents contenant des données :
Un chercheur qui consulte des documents contenant des données personnelles, n'est pas tenu d'en informer les personnes concernées : soit parce qu'il a obtenu la dérogation ou l'autorisation nécessaire de la part de l'auteur de ce document pour consulter, à des fins de recherche scientifique ou historique, les documents administratifs ou d'archives publiques ou d'archives privées ; soit parce que ces documents, qui ne sont ni documents administratifs, ni archives publiques, ni archives privées, comprennent des données personnelles qui ont été collectées licitement et loyalement et que le chercheur les consulte à des fins de recherche scientifique ou historique.

Qu'est ce qu'un consentement éclairé ?

Le consentement éclairé doit être obtenu avant le début de la collecte. Il est lié au droit d'information (cf. dernière partie) et doit contenir :

- les coordonnées du délégué à la protection des données de l'organisme, ou d'un point de contact sur les questions liées à la protection des données personnelles
- les finalités du traitement auquel sont destinées les données
- ce qui autorise l'organisme à traiter ces données
- les tiers qui auront accès aux données
- la durée de conservation des données
- Les modalités d'accès à des droits des usagers et la possibilité d'introduire une réclamation à la CNIL
- L'utilisation des données hors de l'UE
- la base juridique du traitement de données (c'est-à-dire ce qui autorise légalement le traitement : il peut s'agir du consentement des personnes concernées, du respect d'une obligation prévue par un texte, de l'exécution d'un contrat, etc.)

Obligation de prendre les mesures de protection nécessaires

Le responsable du traitement doit vérifier que toutes les mesures nécessaires sont prises, au sein de l'organisme et chez les sous-traitants, pour assurer la protection des données personnelles au cours de leur traitement c'est-à-dire pendant leur collecte, leur consultation, jusqu'à leur effacement le cas échéant.

- Le traitement des données personnelles doit être **mené conformément à sa finalité** : seules les données nécessaires à la réalisation de la finalité peuvent être collectées.
- **Il est tenu d'effacer les données personnelles**, aussitôt qu'elles ne sont plus nécessaires aux recherches pour lesquelles elles ont été collectées ou dès que les personnes concernées le demandent. (Toutefois, les données collectées peuvent être conservées si leur suppression risquerait "de rendre impossible ou d'entraver sérieusement la réalisation des finalités" du traitement).
- **Le fichier qui les contient doit être protégé** et n'être accessible qu'aux personnes autorisées.

Une étude d'impact doit être menée si le traitement présente un risque (données sensibles), enfin les données doivent être effacées dès qu'elles ne sont plus nécessaires à la réalisation de la finalité du fichier.

En cas d'utilisation d'un corpus :
Un chercheur qui consulte des documents contenant des données à caractère personnel, le fait sur autorisation ou dérogation. Il a certainement dû fournir, à cette occasion, un engagement écrit de ne pas copier ni diffuser ces données personnelles. C'est en effet une obligation de la part du responsable du corpus, de mettre en œuvre toutes les mesures destinées à protéger les données personnelles (art. 29 du Règlement européen du 27 avril 2016).

Signalement au CIL

Dans certains cas (à définir), les enquêtes collectant des données à caractère personnel doivent être inscrites au registre du Correspondant informatique et liberté (CIL) de l'établissement (dans une UMR le CIL par défaut est celui de l'employeur du Directeur de l'unité de recherche). Il n'y a pas de déclaration à faire à la CNIL sauf dans certains cas, notamment la collecte de données sensibles ou le transfert des données hors de l'union européenne.
CIL de l'université Lyon 1 : cil@univ-lyon1.fr

Stockage des données

Comment stocker et conserver les données ?

Le cycle de conservation des données à caractère personnel peut être divisé en trois phases successives distinctes : la base active, l'archivage intermédiaire et l'archivage définitif.

La base active
C'est la durée d'utilisation courante des données ou autrement dit, la durée nécessaire à la réalisation de la finalité du traitement.

L'archivage intermédiaire
Il peut être justifié que les données personnelles soient conservées pour des durées plus longues en archivage intermédiaire distinctement de la base active, avec accès restreint, dans la mesure où, sous réserve de garanties appropriées pour les droits et libertés des personnes concernées, certaines données peuvent être traitées exclusivement à des fins archivistiques dans l'intérêt public, à des fins de recherche scientifique ou historique ou à des fins statistiques. Dans le cas d'un archivage intermédiaire, le responsable du fichier doit veiller à ne conserver que les données nécessaires et permettre aux personnes concernées de faire valoir un droit en justice : **un tri doit donc être effectué parmi la totalité des données collectées pour ne garder que les seules données indispensables.**

Pour les archives intermédiaires, le choix du mode d'archivage est laissé à l'appréciation du responsable du fichier. Des données peuvent ainsi être archivées :

- dans une base d'archive spécifique, distincte de la base active, avec des accès restreints aux seules personnes ayant un intérêt en raison de leurs fonctions (par exemple, le service du contentieux)
- ou dans la base active, à condition de procéder à un isolement des données archivées au moyen d'une séparation logique (gestion des droits d'accès et des habilitations) pour les rendre inaccessibles aux personnes n'ayant plus d'intérêt à les traiter

L'archivage définitif
L'archivage définitif concerne la conservation définitive des données. Il est possible de conserver définitivement des données personnelles s'il y a un intérêt public. Dans ce cas, ces données doivent être transmises puis gérées par les services des archives habilités dans le respect du Livre 2 du Code du Patrimoine.

Les bonnes questions à se poser

- Jusqu'à quand ai-je vraiment besoin des données pour atteindre l'objectif fixé ?
- Ai-je des obligations légales de conserver les données pendant un certain temps ?
- Dois-je conserver certaines données en vue de me protéger contre un éventuel contentieux ? Lesquelles ?
- Jusqu'à quand puis-je faire valoir ce recours en justice ?
- Quelles informations doivent être archivées ? Pendant combien de temps ?
- Quelles sont les règles de suppression des données.
- Quelles sont les règles d'archivage des données ?

Pseudonymisation et anonymisation

La pseudonymisation ou anonymisation des données concerne le traitement de données à caractère personnel de telle façon que celles-ci ne puissent plus être attribuées à une personne concernée précise sans avoir recours à des informations supplémentaires, pour autant que ces informations supplémentaires soient conservées séparément et soumises à des mesures techniques et organisationnelles afin de

garantir que les données à caractère personnel ne sont pas attribuées à une personne physique identifiée ou identifiable. Dans le cas de la pseudonymisation, le responsable des données va attribuer un pseudo au participant tandis que dans le cas d'une anonymisation, il sera impossible de relier entre elles les données d'un même participant. Il est souvent difficile de savoir comment anonymiser correctement des données.

Comment évaluer une solution d'anonymisation ?
Une solution d'anonymisation doit être construite au cas par cas et adaptée aux usages prévus. Pour aider à évaluer une bonne solution d'anonymisation, l'article 29 propose trois critères :

- L'individualisation : est-il toujours possible d'isoler un individu ?
- La corrélation : est-il possible de relier entre eux des ensembles de données distincts concernant un même individu ?
- L'inférence : peut-on déduire de l'information sur un individu ?

Ainsi : un ensemble de données pour lequel il n'est possible ni d'individualiser ni de corréler ni d'inférer est a priori anonyme ; un ensemble de données pour lequel au moins un des trois critères n'est pas respecté ne pourra être considéré comme anonyme qu'à la suite d'une analyse détaillée des risques de ré-identification.

Comment choisir les techniques d'anonymisation ?
Ces techniques se regroupent autour de deux grand principes : transformer les données pour qu'elles ne se réfèrent plus à une personne réelle et généraliser les données de façon à ce qu'elles ne soient plus spécifiques à une personne mais communes à un ensemble de personnes.

Les droits des participants

Voici les 5 droits fondamentaux des participants à la collecte des données :

Droit d'information

L'organisme qui collecte des informations sur des usagers doit proposer une information claire sur l'utilisation de données et sur les droits des usagers.

Voici les informations indispensables à connaître :

- les coordonnées du délégué à la protection des données de l'organisme, ou d'un point de contact sur les questions liées à la protection des données personnelles
- l'utilisation qui sera faite des données
- ce qui autorise l'organisme à traiter ces données
- les tiers qui auront accès aux données
- la durée de conservation de vos données
- Les modalités d'accès à des usagers à leurs droits et la possibilité d'introduire une réclamation à la CNIL
- L'utilisation des données hors de l'UE
- la base juridique du traitement de données (c'est-à-dire ce qui autorise légalement le traitement : il peut s'agir du consentement des personnes concernées, du respect d'une obligation prévue par un texte, de l'exécution d'un contrat, etc.)

Et selon le cas :

- l'existence d'une prise de décision automatisée ou d'un profilage, les informations utiles à la compréhension de l'algorithme et de sa logique, ainsi que les conséquences pour la personne concernée
- le fait que les données sont requises par la réglementation, par un contrat ou en vue de la conclusion d'un contrat
- les intérêts légitimes poursuivis par le responsable du traitement ou par un tiers (exemple : prévention de la fraude)
- le droit au retrait du consentement à tout moment
- la faculté d'accéder aux documents autorisant le transfert de données hors de l'Union européenne (exemples : clauses contractuelles types de la Commission européenne)

Au moindre problème dans la collecte des données (destruction, perte, altération ou vol des données) l'usager doit être mis au courant.

Droit d'accès et de portabilité

À tout moment, l'usager peut demander à l'organisme qui détient des données sur lui à ce qu'on les lui communique pour en vérifier le contenu en échange d'une pièce d'identité.

Tout comme dans le cas du droit à l'information, l'organisme doit être capable de donner :

- les finalités d'utilisation de ces données,
- les catégories de données collectées,
- les destinataires ou catégories de destinataires qui ont pu accéder à ces données,
- la durée de conservation des données ou les critères qui déterminent cette durée,
- l'existence des autres droits (droit de rectification, d'effacement, de limitation, d'opposition),
- la possibilité de saisir la CNIL,
- toute information relative à la source des données collectées si celles-ci n'ont pas directement été récoltées auprès de vous,
- l'existence d'une prise de décision automatisée, y compris en cas de profilage, et la logique sous-jacente, l'importance et les conséquences pour vous d'une telle décision,
- l'éventuel transfert de vos données vers un pays tiers (non-membre de l'UE) ou vers une organisation internationale.

Le droit à la portabilité
Le droit à la portabilité offre la possibilité de récupérer une partie des ses données dans un format lisible par une machine. Libre à l'usager de stocker ailleurs ses données portables ou les transmettre facilement d'un système à un autre, en vue d'une réutilisation à d'autres fins.

Droit de rectification

L'usager peut demander la rectification des informations inexactes ou incomplètes le concernant en échange d'une pièce d'identité. Il permet d'éviter qu'un organisme n'utilise ou ne diffuse des informations erronées sur lui.

Droit d'opposition

L'usager peut s'opposer à tout moment à ce qu'un organisme utilise certaines de ses données. Cependant, dans le cas de la recherche scientifique il existe une dérogation lorsque la recherche a été publiée.
"Lorsque des données à caractère personnel sont traitées à des fins de recherche scientifique ou historique ou à des fins statistiques en application de l'article 89, paragraphe 1, la personne concernée a le droit de s'opposer, pour des raisons tenant à sa situation particulière, au traitement de données à caractère personnel la concernant, à moins que le traitement ne soit nécessaire à l'exécution d'une mission d'intérêt public."

Droit d'effacement

L'usager peut demander la suppression des informations le concernant en échange d'une pièce d'identité.

Cependant, la recherche publique bénéficie d'un régime dérogatoire :
"Les États membres devraient être autorisés à prévoir, dans des conditions spécifiques et moyennant des garanties appropriées pour les personnes concernées, des dispositions particulières et des dérogations concernant les exigences en matière d'information et les droits à la rectification, à l'effacement, à l'oubli, à la limitation du traitement, à la portabilité des données et le droit d'opposition lorsque les données à caractère personnel sont traitées à des fins archivistiques dans l'intérêt public, à des fins de recherche scientifique ou historique ou à des fins statistiques."
Un responsable de traitement de données à des fins de recherche scientifique peut donc refuser de faire droit à une demande d'effacement, mais il ne s'agit pas d'une faculté discrétionnaire : **il doit être en mesure de prouver que cette suppression empêche la recherche projetée ou la compromet gravement.** Dans ce cas il faudra trouver un compromis avec lui comme par exemple une meilleure anonymisation de ses données.

Sources

https://www.cnil.fr/fr/limiter-la-conservation-des-donnees
https://www.cnil.fr/fr/les-droits-pour-maitriser-vos-donnees-personnelles
http://www.cil.cnrs.fr/CIL/spip.php?article2869
https://ethiquedroit.hypotheses.org/1717
https://www.cnil.fr/fr/reglement-europeen-protection-donnees/chapitre3#Article17

Note de l’auteur

Ce document n’ayant pas été réalisé par une juriste mais par une jeune designer-chercheuse il se pourrait que des informations fausses se soient glissées dedans. Pour me faire part de vos remarques ou propositions de modifications n’hésitez pas à me contacter ici : alice.blot@univ-lyon1.fr