

# Projet de guide pratique

Ouverture et réutilisation de données  
publiquement accessibles

*Publié pour consultation publique le 01/08/2023*

# Introduction

## Ouverture, partage et réutilisation des données : quels sont les enjeux ?

**Le potentiel et les enjeux de l'ouverture, du partage et de la réutilisation de données s'imposent de plus en plus dans le débat public.** La mise à disposition de données au plus grand nombre n'est pas seulement un enjeu technique mais surtout une question économique, scientifique, démocratique et politique. D'un point de vue juridique, de telles opérations soulèvent également des enjeux au regard de la protection des données à caractère personnel, sans pour autant que les textes applicables constituent un obstacle.

**Consciente depuis longtemps de ces évolutions, la CNIL se montre, de manière constante, favorable au principe de la circulation des données, tout en considérant que la protection de ces dernières est la condition d'une ouverture, d'un partage et d'une réutilisation acceptables socialement et soutenables éthiquement.**

Le mouvement réglementaire en faveur de l'ouverture et de la réutilisation des données s'est accéléré ces dernières années. La stratégie européenne pour les données a notamment pour objectif de faciliter le partage des données entre les acteurs, qu'ils soient publics ou privés.

Dans la continuité de [la directive du 20 juin 2019 concernant les données ouvertes et la réutilisation des informations du secteur public](#) (dite directive « ISP 2 »), qui a créé la notion d'« ensembles de données de forte valeur » devant être obligatoirement partagés<sup>1</sup>, deux textes européens traduisent cette volonté :

- le règlement sur la gouvernance des données (« *Data Governance Act* », applicable fin septembre 2023), qui, pour « *favoriser la disponibilité des données en vue de leur réutilisation en augmentant la confiance dans les intermédiaires de données et en renforçant les mécanismes de partage* », fixe les conditions de réutilisation de certaines données protégées détenues par des organismes du secteur public, les exigences applicables aux services d'intermédiation « neutres » de données, ainsi que le cadre d'enregistrement volontaire des organisations « altruistes » en matière de données ;
- la proposition de règlement sur les données (« *Data Act* », publié en février 2022), qui, pour assurer en particulier un meilleur équilibre dans la distribution de la valeur issue de l'exploitation des données, envisage une approche plus contraignante en prévoyant notamment leur partage avec les consommateurs et entre les entreprises dans le contexte du développement de l'Internet des objets, ainsi que l'accès par des organismes publics aux informations détenues par des acteurs privés en cas de nécessité exceptionnelle.

Sur le plan national, [le rapport Bothorel](#) sur la politique publique de la donnée, remis au premier ministre en décembre 2020, a suggéré que la CNIL réalise un « guide de conformité » sur le cadre juridique du partage volontaire des « données d'intérêt général » et l'application du droit d'accès et de réutilisation aux données du secteur privé reçues par les administrations.

Parallèlement, au cours de ces dernières années, la CNIL a eu à connaître **d'un nombre croissant de demandes d'avis sur des projets de texte, de demandes de conseil, de plaintes et de dossiers de contrôle ou de sanctions** concernant des traitements de données publiquement accessibles, tant par le secteur privé que public.

**Dans le prolongement des [travaux CADA-CNIL menés sur l'open data](#), et devant la montée en puissance des réutilisations de données librement accessibles sur internet, un groupe de travail interne (« GT ouverture, partage et réutilisation de données »), piloté par Madame Anne Debet, a été créé dans le but de clarifier, d'illustrer par des cas types concrets, et de décliner en bonnes pratiques immédiatement mobilisables, le cadre juridique applicable et les obligations qui en découlent ;** ce au bénéfice, d'une part des organismes souhaitant ou devant ouvrir les données qu'ils détiennent et, d'autre part, des réutilisateurs des données librement accessibles, du secteur public comme du secteur privé.

Ce guide est le fruit d'un état des lieux approfondi de la doctrine de la CNIL en la matière, enrichi d'une série de consultation de différents acteurs directement impliqués dans le développement des démarches d'ouverture, de

<sup>1</sup> En décembre 2022, la Commission européenne a adopté [le règlement d'exécution établissant la liste des ensembles de données de forte valeur](#) relevant des six catégories thématiques visées par la directive (« géospatial », « observation de la terre et environnement », « météorologiques », « statistiques », « sociétés et propriété de l'entreprise », « mobilité »), ainsi que les modalités de leur publication et de leur réutilisation. Applicable en avril 2024, il a pour principal objectif « *que les données publiques présentant le potentiel socio-économique le plus élevé soient mises à disposition aux fins de réutilisation avec un minimum de restrictions légales et techniques et gratuitement* ».

partage et de réutilisation des données (partenaires institutionnels, associations et entreprises, avocats, chercheurs).

**Ces premiers travaux se concentrent sur l'ouverture des données sur internet et la réutilisation des données publiquement accessibles.**

## De quoi parle-t-on ?

Le périmètre de ce guide porte sur :

- **La diffusion publique de données personnelles**, qui consiste en la mise à disposition du public, en ligne et dans un format ouvert, aisément réutilisable et exploitable par machine, de données se rapportant à des personnes physiques identifiées ou potentiellement identifiables, à des fins de réutilisation par des tiers (par exemple, au moyen d'une interface de programmation applicative – API –, ou par le biais d'un fichier au format texte CSV). Cette **opération d'ouverture des données, souvent désignée sous le terme anglais de mise en « open data »**, est essentiellement réalisée par des personnes publiques (p. ex. : la [diffusion des déclarations des responsables publics](#) par la Haute Autorité pour la transparence de la vie publique), ouvrant leurs données en application de législations spéciales ou des dispositions du code des relations du public avec l'administration (CRPA qui organise le régime général du droit d'accès et de réutilisation des données publiques), mais peut également être le fait de personnes privées, souhaitant notamment partager avec le plus grand nombre des données « d'intérêt général ». Elle porte généralement sur un ensemble de données relativement important et structuré, qui constitue alors une base de données.

Il est à noter que la notion d'ouverture de données ne comprend pas les publications de données sur Internet accessibles à tous mais dont la vocation première n'est pas de permettre la réutilisation (p. ex. : publication de contenus sur les réseaux sociaux par les individus eux-mêmes, d'articles de blogs, etc.). Elle ne comprend pas non plus les cas de partage de données en « circuit fermé » ou « semi-ouvert », qui correspondent aux transmissions de données à des tiers individuellement autorisés, figurant ou non sur une liste limitative (p. ex. : le partage de données entre administrations à des fins de simplification des démarches pour l'utilisateur, le partage avec les pouvoirs publics de « données d'intérêt général » par des acteurs privés, la mise à disposition d'API à destination des chercheurs, la commercialisation de bases de données commerciales, etc.). Ces partages feront l'objet de travaux ultérieurs.

Dans le présent guide, l'expression « diffuseur de données » s'entendra uniquement des organismes qui mettent des données personnelles à disposition du public, généralement en utilisant la toile (*web*).

- **La réutilisation de données personnelles publiquement accessibles**, notion plus large, qui concerne bien sûr les données diffusées en open data précédemment évoquées, mais aussi toutes les autres données librement **accessibles en ligne**, par toute personne, gratuitement ou non, avec ou sans création de compte.

Autrement dit, sont visées aussi bien les données ouvertes à des fins d'exploitation par des tiers, que les données « simplement » publiées par un individu sur un site web (p. ex. : discussions sur un forum). De telles réutilisations de données peuvent être le fait de personnes publiques (p. ex. : une université procédant à un travail de recherche sur les impacts pour la vie privée des publications librement accessibles en ligne), comme de personnes privées (p. ex. : une entreprise qui exploite des données publiées sur les réseaux sociaux à des fins de veille réputationnelle).

## À qui s'adresse ce guide ?

**Ce guide s'adresse à la fois :**

- **aux diffuseurs de données sur Internet**, personnes physiques ou morales, publiques ou privées, qui mettent en ligne à disposition du public des données personnelles d'individus (par exemple, l'INPI et l'INSEE qui diffusent en open data des données d'entreprises) ;
- **aux réutilisateurs de données publiquement accessibles**, personnes physiques ou morale, publiques ou privées, collectant les données en vue d'une exploitation de celles-ci pour leur propre compte (par exemple, l'exploitation de la base SIRENE de l'INSEE par une entreprise pour établir un tableau référençant ses concurrents au sein de sa zone géographique).

Enfin, **ce guide pourra également intéresser les personnes concernées par les traitements mis en œuvre par les diffuseurs et réutilisateurs** : elles y trouveront, en particulier, des informations sur les droits dont elles disposent pour conserver la maîtrise des usages qui sont faits de leurs données.

# Table des matières

---

Ce guide pratique, qui a vocation à s'enrichir au fil de l'eau en fonction des usages, propose :

- **une grille d'analyse générale, sous la forme de « fiches principes »**, permettant à tout diffuseur et réutilisateur de données personnelles de cheminer le plus rapidement et efficacement possible sur les questions structurantes et éléments de réponse pertinents ;
- **des fiches « cas d'usage »**, illustrant comment les diffuseurs ou réutilisateurs peuvent se mettre en conformité de manière concrète et opérationnelle.

## Sommaire des fiches principes

Fiche principe n°1 : Quelle qualification juridique pour les diffuseurs et réutilisateurs de données ? .....	5
Fiche principe n°2 : Comment identifier la base légale de son traitement ? .....	8
Fiche principe n°3 : comment informer les personnes concernées.....	16
Fiche principe n°4 : Quels sont les droits des personnes sur leurs données ?.....	22
Fiche principe n°5 : Comment garantir la minimisation des données traitées ?.....	27
Fiche principe n°6 : Comment garantir l'exactitude, la sécurité et la conservation limitée des données ? .....	32

## Sommaire des fiches cas d'usage

Fiche cas d'usage n°1 : la diffusion et la réutilisation des données d'entreprises déclarées à l'administration..	37
Fiche cas d'usage n°2 : la réutilisation de données publiquement accessibles aux fins de diffusion d'annuaires « enrichis » de professionnels .....	45
Fiche cas d'usage n°3 : la réutilisation de données publiquement accessibles à des fins de constitution ou d'enrichissement de fichiers destinés à la prospection commerciale.....	54
Fiche cas d'usage n°4 : La réutilisation de données publiquement accessibles à des fins de recherche scientifique (hors santé) .....	61
Fiche cas d'usage n°5 : le moissonnage de données publiquement accessibles par des autorités publiques dans le cadre de leurs missions .....	72

# Fiche principe n°1 : Quelle qualification juridique pour les diffuseurs et réutilisateurs de données ?

Toute personne physique ou organisme traitant des données personnelles doit au préalable s'interroger sur sa qualification au sens du RGPD, qualification dont vont dépendre ses obligations. Pour un traitement donné, il est possible d'être responsable de traitement, sous-traitant ou responsable de traitement conjoint. Il incombe aux acteurs de déterminer leur qualification au cas par cas. La présente fiche rappelle les définitions de ces différentes notions et explique comment les appliquer en cas de diffusion publique d'une base de données sur Internet et de réutilisation de données publiquement accessibles sur Internet.

## Le responsable de traitement

**Le responsable de traitement** est la personne physique ou morale, l'autorité publique, le service ou un autre organisme qui **détermine les finalités et les moyens du traitement, c'est-à-dire qui décide du « pourquoi » et du « comment » de l'utilisation de données personnelles.**

Dans les cas où les données sont diffusées en open data, les acteurs seront généralement responsables de traitement distincts, avec le diffuseur des données d'un côté et le réutilisateur de l'autre.

S'agissant de la diffusion publique d'une base de données, le responsable de traitement est en principe celui qui décide de la mise en ligne ou, lorsqu'il s'agit d'une obligation légale, celui sur qui pèse cette obligation.

### S'agissant de la réutilisation de données publiquement accessibles sur Internet :

- Lorsqu'une personne décide de réutiliser des données sur Internet et détermine, par là même, la finalité (l'objet de la réutilisation) et la caractéristique principale du traitement (utiliser des données publiées sur Internet) : elle est en principe responsable de traitement, y compris si elle recourt à un prestataire. Cela est valable que les données aient ou non été initialement diffusées en open data.

Il en va ainsi de la personne qui décide de constituer une base de données à partir de données publiquement accessibles pour l'exploiter commercialement auprès de différents clients. Cette personne et ses clients pourraient alors être qualifiés de responsables de traitement distincts, à condition que chaque traitement puisse être séparé de l'autre (et que leurs finalités et moyens ne soient pas déterminés conjointement comme cela est détaillé ci-dessous).

- Lorsqu'une personne confie une mission à un prestataire sans lui en imposer les moyens et que c'est ce prestataire qui choisit, pour atteindre l'objectif, de réutiliser des données personnelles publiées sur Internet, c'est en principe le prestataire qui doit être qualifié de responsable du traitement.

### Focus : qu'est-ce qu'une finalité ?

La finalité est le but précis pour lequel les données sont traitées. Elle doit être déterminée, explicite et légitime.

Exemples :

- il peut s'agir de la publication des données dans un annuaire public à des fins d'information du public.
- il peut s'agir de la réutilisation de données publiquement accessibles à des fins de prospection commerciale entre professionnels.

Si une analyse au cas par cas est toujours nécessaire afin de qualifier un responsable de traitement, certaines situations laissent peu de doutes.

Exemples de responsables de traitement :

- une collectivité territoriale qui doit publier un document administratif contenant des données personnelles dans le cadre de ses obligations légales ;
- une entreprise d'édition juridique qui décide de collecter des coordonnées d'avocats dans un annuaire publié par le CNB afin de les prospector par courriel.

## Le responsable conjoint du traitement

**Lorsque deux responsables du traitement, ou plus, déterminent conjointement les finalités et les moyens du traitement**, ils sont responsables conjoints du traitement.

La qualification peut être plus délicate à opérer en présence de plusieurs acteurs exerçant une influence sur la détermination des finalités et des moyens du traitement. Il convient alors de déterminer, entre autres, si les acteurs traitent les données pour des finalités propres et distinctes ou pour une finalité commune.

**En matière de partages ouverts, c'est-à-dire quand il s'agit de publier ou de réutiliser des données publiées, la question de la responsabilité conjointe peut se poser lorsqu'un intermédiaire intervient** entre un diffuseur et un réutilisateur de données.

### Focus sur les licences de réutilisation

Le diffuseur des données peut choisir d'encadrer les réutilisations futures par le biais d'une licence, notamment afin d'assurer une certaine prévisibilité aux personnes concernées de l'usage qui pourra être fait de leurs données. Cela ne le rendra pas nécessairement responsable de ces réutilisations.

D'une part, l'autorisation ou l'interdiction par le diffuseur de certaines catégories de réutilisation ne permet pas de considérer que la finalité de la réutilisation a été conjointement déterminée si le réutilisateur conserve un degré d'influence autonome sur ses propres traitements, à commencer par la décision d'utiliser ou non les données.

D'autre part, déterminer conjointement la finalité n'est pas suffisant si les moyens sont définis distinctement.

De manière générale, prévoir une licence de réutilisation est une pratique à encourager de la part des diffuseurs, y compris en dehors des cas de responsabilité conjointe (pour lesquels le RGPD impose la conclusion d'un « accord »). Lorsque les réutilisateurs sont des responsables de traitement distincts, il reste en effet pertinent de leur rappeler l'étendue de leur responsabilité et de leurs obligations en matière de protection des données personnelles.

**En cas de responsabilité conjointe, chaque responsable de traitement doit s'assurer de la licéité du traitement, notamment en définissent de manière transparente leurs obligations respectives dans le cadre d'un « accord ».**

La forme de cet accord n'est pas précisée par le RGPD. L'essentiel est que les parties s'engagent mutuellement, dans le cadre d'un contrat notamment, sur « qui fait quoi ».

Il est à noter que les personnes concernées peuvent exercer leurs droits auprès de n'importe lequel des responsables de traitement conjoint.

## Le sous-traitant

Le **sous-traitant agit pour le compte du responsable de traitement**. Il ne peut traiter les données pour son propre compte ([sauf exception](#)), et ne doit les traiter que sur instructions documentées du responsable de traitement.

Exemples de sous-traitants :

- Le prestataire du diffuseur de données :
- L'éditeur d'un logiciel (proposé à l'utilisation comme service, en mode « SaaS ») d'aide à l'organisation et à la publication des données agirait en qualité de sous-traitant dès lors qu'il ne décide ni de la finalité (pourquoi le logiciel est utilisé en l'espèce) ni des moyens (quelles données, publiées auprès de quels destinataires, pour combien de temps, etc.) du traitement de son client ;
- le fournisseur d'une plateforme de partage de données en ligne, telle que la DINUM qui édite et développe la plateforme data.gouv.fr, et qui ne fait qu'héberger et mettre à disposition les données que tout responsable de traitement décide de partager par ce biais, dès lors que ce dernier conserve la maîtrise de la finalité et des moyens de la publication ou du partage ;

- Le prestataire du réutilisateur de données :
  - L'éditeur d'une interface de programmation applicative de moissonnage de données permettant de récupérer des données en ligne, dès lors qu'il ne détermine pas la finalité et les moyens de chaque utilisation (tels que les données collectées).

### Indifférence de l'accès aux données ou de leur exploitation effective

L'absence d'utilisation ou d'exploitation effective des données personnelles, voire l'absence d'accès aux dites données par un acteur n'exclut pas nécessairement la qualification de responsable de traitement, puisque cela n'empêche pas que ce dernier puisse décider à la fois des finalités et des moyens essentiels du traitement (CJUE, arrêt témoins de Jéhovah du 10 juillet 2018 dans l'affaire C-25/17). À titre d'exemple, cela a pu être le cas dans une sanction publique de la CNIL (délibération de la formation restreinte n°SAN-2021-012 du 26 juillet 2021) qui a considéré qu'un cabinet de conseil avait établi un fichier de *lobbying* pour le compte de son client, ce dernier étant responsable de traitement.

À l'inverse, le fait qu'un acteur centralise des données qu'il collecte sur Internet pour en donner l'accès à ses différents clients, dans le cadre de prestations de services distinctes, est généralement un indice décisif pour retenir sa responsabilité. Il sera en effet souvent difficile de considérer qu'il opère pour le compte de ses clients dans cette situation, contrairement au cas où il constituerait une base de données par client, où la sous-traitance serait plus facile à retenir.

### Focus sur les contrats d'adhésion et les moyens non essentiels

Le fait de recourir à un logiciel de traitement de données personnelles conçu par un autre acteur et sur lequel son utilisateur ne peut qu'effectuer certains paramétrages (voire aucun) ne dispense pas ce dernier de sa qualité de responsable de traitement si c'est lui qui décide d'utiliser des données personnelles avec ce logiciel, pour une certaine finalité (CJUE, deuxième chambre, 29 juillet 2019, Fashion ID GmbH & Co. KG, C-40/17).

Un tiers peut donc déterminer des moyens de traitement indépendamment du traitement opérationnel (p. ex. : matériel ou logiciel utilisé pour le traitement, méthode de stockage des données, etc.) ou fournir un service « clé en main » sans pour autant être responsable de traitement. Si ce tiers est ensuite conduit, sur instruction de son client, à manipuler lui-même les données personnelles pour le traitement, il agira en tant que sous-traitant.

À l'inverse, l'établissement d'un cahier des charges, notamment dans le cadre d'un appel d'offre ou dans la commande publique, n'exclut pas que le répondant puisse être responsable de traitement. Tel sera le cas si le contrat avec le prestataire ne fixe que l'objectif de la prestation en laissant ce dernier entièrement libre de déterminer la finalité et les moyens du traitement envisagé.

Pour plus d'informations, voir le guide sur la [responsabilité des acteurs dans la commande publique](#).

**En situation de sous-traitance, un contrat doit être conclu entre le sous-traitant et son responsable de traitement.** Ce contrat doit contenir toutes les mentions prévues par l'article 28 du RGPD. Par ailleurs, le responsable de traitement doit s'assurer que son sous-traitant présente des garanties suffisantes pour la conformité des traitements (pour en savoir plus : [guide du sous-traitant](#)).

## Fiche principe n°2 : Comment identifier la base légale de son traitement ?

La base légale d'un traitement est ce qui donne le droit à un organisme de traiter des données personnelles. L'identification d'une base légale est donc une première étape indispensable pour assurer la conformité d'un projet de partage de données ou de réutilisation de données publiquement accessibles sur Internet.

### L'obligation d'identifier une base légale

Pour être licite, tout traitement doit se fonder sur l'une des six « bases légales » (ou « fondements juridiques ») prévues par le RGPD.

**Une base légale valable doit ainsi être identifiée en amont de la mise en œuvre du traitement, au cas par cas, de manière adaptée à la situation et au type de traitement en cause** (statut de son responsable, objectifs poursuivis, obligations réglementaires, enjeux pour les personnes concernées, etc.).

Attention : le fait que des données soient « publiquement accessibles », en étant notamment **en libre accès sur Internet, ne signifie pas qu'il est possible de se passer d'une base légale ou de réutiliser ces données sans conditions.**

**En matière de diffusion publique d'une base de données et de réutilisation de données publiquement accessibles, les principales bases légales envisageables sont :**

- **[l'obligation légale](#)** : le traitement est imposé à l'organisme, public ou privé, par des dispositions législatives ou réglementaires ;
- **[la mission d'intérêt public](#)** : le traitement est nécessaire à l'exécution d'une mission d'intérêt public telle que définie par des dispositions légales ;
- **[l'intérêt légitime](#)** : le traitement est nécessaire à la poursuite d'intérêts légitimes de l'organisme qui traite les données ou d'un tiers, sous réserve que ne prévalent pas les droits et intérêts des personnes dont les données sont traitées ;
- **[le consentement](#)** : la personne a consenti au traitement de ses données, de manière libre, spécifique, éclairée et univoque.

L'identification d'une base légale appropriée est d'autant plus importante qu'elle a aussi des conséquences sur les droits des personnes concernées.

### Pour approfondir

- [Les bases légales, cnil.fr](https://www.cnil.fr)

### Questions à se poser pour identifier sa base légale

#### Pour les diffuseurs de données sur Internet

Si les conditions propres à la base légale envisagée par chacune des questions ci-dessous ne sont pas remplies, l'organisme doit, soit modifier les paramètres de son projet de traitement pour parvenir à les respecter, soit rechercher une autre base légale.

1. *La diffusion publique de données envisagée est-elle nécessaire au respect d'une [obligation légale](#) ?*

Pour que la base légale « obligation légale » puisse être retenue, il faut :

- qu'un texte suffisamment clair et précis exige de l'organisme la mise en œuvre du traitement dans les conditions projetées ;
- qu'il n'existe, en d'autres termes, pas de moyen moins intrusif (p. ex. : mise à disposition des données sur demande ou diffusion limitée à telles et telles données) permettant de respecter les dispositions légales en cause.

## Illustrations

Ce sera notamment le cas des traitements suivants :

- la diffusion par l'INPI du registre national des entreprises, en application de l'[article L123-52 du code de commerce](#) ;
- la diffusion par la Haute Autorité pour la transparence de la vie publique des [données portant sur les déclarations de situation patrimoniale des membres du gouvernement](#), ainsi que sur les déclarations d'intérêts de ces derniers et des élus européens, nationaux et locaux, en vertu de l'article 5 de la loi n° 2013-907 du 11 octobre 2013 relative à la transparence de la vie publique ;
- la diffusion par l'administration fiscale des informations relatives aux ventes de biens immobiliers intervenues au cours des cinq dernières années (« [base DVF](#) »), en vertu de l'article R\*. 112 A-1 du livre des procédures fiscales ;
- la diffusion par la CNIL de la dénomination et des coordonnées professionnelles des organismes ayant désigné un délégué à protection des données, ainsi que des moyens de contacter ce dernier (« [base DPO](#) »), conformément à l'article 83 du décret d'application de la loi Informatique et Libertés.

2. *Si non, et que je suis un organisme public ou privé poursuivant une [mission d'intérêt public](#), cette diffusion de données est-elle nécessaire à l'exécution d'une telle mission ?*

La possibilité de se fonder sur la base légale « mission d'intérêt public » suppose :

- que la mission dans laquelle s'inscrit le traitement soit prévue par un texte normatif applicable au diffuseur (par exemple un décret fixant les missions d'un organisme) ;
- que la diffusion des données permette d'exercer *spécifiquement* cette mission (ce n'est pas le cas si elle vise un objectif sans rapport particulier avec celle-ci ou trop éloigné de ses particularités), de manière pertinente et appropriée.

S'il ne s'agit pas d'une obligation, plus ces dispositions sont précises, plus il est facile pour les organismes concernés de recourir à cette base légale.

## Illustrations

Est concernée par cette base légale la diffusion en open data, par les administrations, de documents administratifs contenant des données personnelles lorsque cette mise à disposition sans anonymisation préalable est :

- autorisée par le code des relations avec le public et l'administration ([art. D. 312-1-3](#)) ;
- sans être imposée par un texte (simple faculté).

Exemples :

- la publication par une commune, dans un format ouvert, de l'organigramme de ses services pouvant comporter l'identité de ses responsables, ou encore des permis de construire pouvant mentionner le nom et l'adresse du pétitionnaire, ainsi que ceux de l'architecte.
- la mise à disposition par l'INSEE du répertoire des entreprises et de leurs établissements (« [base SIRENE](#) ») en application de l'[article R. 123-232 du code de commerce](#).

Lorsque la mise en ligne des données personnelles contenues dans des documents administratifs n'est pas autorisée par un texte, seul le recueil de consentements valables auprès des personnes concernées permettra à l'administration d'assurer la licéité de son traitement (voir à ce sujet le [guide élaboré par la CADA et la CNIL sur l'ouverture des données publiques](#)).

3. *Si non, et que je suis un organisme privé, ce traitement est-il nécessaire à la poursuite d'un [intérêt légitime](#) ?*

Pour que « l'intérêt légitime » justifie la diffusion de données en *open data*, il faut que :

- **l'intérêt poursuivi soit bien légitime**, qu'il s'agisse de l'intérêt propre du diffuseur ou de l'intérêt de tiers, tels que d'éventuels nouveaux destinataires : ce pourra notamment être le cas lorsque la

diffusion ouverte est réalisée pour satisfaire l'intérêt du public à trouver une information ou encore pour permettre des recherches scientifiques ;

- **le traitement doit être nécessaire à la poursuite de cet intérêt** : pour le satisfaire, il ne doit pas exister d'autres moyens moins invasifs (p. ex. : publication de statistiques anonymes ou partage de données avec une liste restreinte de destinataires) que la mise à disposition sur Internet des données personnelles ;
- **cet intérêt ne porte pas une atteinte disproportionnée aux intérêts, droits et libertés des personnes concernées.**

Pour vérifier cette dernière condition, il faut notamment prendre en compte :

- **l'importance des bénéfices que l'on peut anticiper du traitement**, pour le responsable de traitement ou pour les tiers ;
- **les attentes raisonnables des personnes concernées**, qui devraient pouvoir « naturellement » anticiper le traitement d'ouverture de leur données (pas de surprise quant aux finalités, modalités de mise en œuvre et conséquences), y compris quand l'organisme envisage d'ouvrir les données après les avoir collectées ;
- **les risques qui en résultent pour les personnes** (p. ex. : atteinte à la réputation, impact émotionnel), notamment au regard des catégories de données en cause (p. ex. : données sensibles ou se rapportant à des mineurs) et de toutes les réutilisations (p. ex. : suivi ou surveillance des activités, profilage) rendues possibles (qu'elles soient autorisées ou non par le diffuseur des données) ; ces risques sont très différents selon que les données diffusées sont ou non déjà disponibles, même de façon éparse, sur Internet : si c'est le cas, les risques sont beaucoup plus faibles ; dans le cas contraire, il convient de faire une appréciation fine des risques que l'accessibilité aux données personnelles peut faire courir aux personnes concernées ;
- **les garanties pertinentes pour réduire ces risques**, telles que la pseudonymisation des données, l'accès à celles-ci par un moteur de recherche dédié ou des interfaces de programmation applicatives (API), la mise en place de systèmes bloquant l'indexation des données nominatives par les moteurs de recherche externes et le moissonnage en ligne (« *web scraping* »), ou encore des CGU interdisant certaines réutilisations (voir les fiches principe [n°5](#) et [6](#) du présent guide dédiées aux mesures à adopter pour garantir la minimisation, l'exactitude et la sécurité des données traitées).

### Cas particulier : la diffusion publique des données sur Internet non envisagée au moment de leur collecte

Si cette ouverture n'est pas prévue par le droit en vigueur, l'organisme qui veut y procéder devra :

- soit recueillir le consentement préalable des personnes concernées ;
- soit pouvoir démontrer que cette publication répond à son intérêt légitime et qu'elle est compatible avec le contexte initial de la collecte des données (au regard notamment des attentes des personnes concernées ou des catégories de données traitées), conformément à l'article 6.4 du RGPD.

En pratique, la mise à disposition sur Internet d'une base de données personnelles sur le fondement de l'intérêt légitime du diffuseur correspond à deux hypothèses principales :

- le cas où cette mise à disposition est un prolongement acceptable de l'activité du diffuseur, et correspond notamment à son activité commerciale normale ; *par exemple, un éditeur qui publie des articles d'auteur sur son site peut légitimement décider de mettre à disposition une base de données contenant les références de tous les articles, avec le nom de l'auteur, des mots-clefs, etc.* ;
- le cas où un acteur, sans y être légalement obligé ou que cela corresponde à une mission d'intérêt public, choisit de mettre à disposition des tiers une base de données qu'il a constituée, notamment pour permettre à des chercheurs et à des programmes de recherche et développement d'y accéder.

Cette dernière démarche peut notamment correspondre aux concepts de « données d'intérêt général » ou « d'altruisme de la donnée ». Elle participe indéniablement au développement de l'économie numérique et est au cœur de l'intelligence artificielle. Mais elle fait aussi courir des risques aux personnes si la base de données mise en libre disposition contient des données personnelles.

Ainsi, il est indispensable que le diffuseur commence par se demander s'il est possible de diffuser une base de données anonymisées (le sont par exemple celles publiées par les entreprises gestionnaires des réseaux de transport et de distribution d'électricité et de gaz naturel, en application et dans les conditions prévues par le décret n°2017-486), pour lesquelles la réidentification des personnes n'est pas possible par des moyens raisonnables.

À défaut, la CNIL estime que l'intérêt légitime du diffuseur nécessite en principe des mesures de pseudonymisation. La licéité de l'ouverture des données à des fins de réutilisation s'apprécie au cas par cas. Elle dépend notamment du caractère déjà public ou non des données contenues dans la base de données.

Enfin, si la balance des intérêts en présence rend illégale (hors consentement des personnes concernées) une telle ouverture, le responsable de traitement peut envisager une mise à disposition non publique, par exemple réservée à des chercheurs qu'il autorise individuellement à accéder à la base (ce cas d'usage est également encadré par le RGPD mais ne correspond pas à l'objet de ce guide).

#### 4. Si non, est-il possible de recueillir un [consentement](#) valable ?

Pour recueillir valablement le consentement des personnes concernées quant à la diffusion de leurs données en *open data*, plusieurs conditions doivent être respectées :

- **le consentement doit être libre** : tout d'abord les personnes doivent pouvoir choisir sans contrainte d'accepter ou non l'ouverture de leurs données, mais aussi de changer d'avis librement sans avoir à subir de conséquences négatives (p. ex. : exclusion de l'accès au service ayant motivé la fourniture des données) ;
- **le consentement doit être spécifique** : en particulier, s'il intervient dans un contexte contractuel, il doit être distinct de l'accord contractuel (l'acceptation des conditions générales d'utilisation du service ne vaut ainsi pas acceptation du traitement de mise à disposition des données sur Internet) ;
- **le consentement doit être éclairé** : son recueil doit être accompagné de la fourniture en amont d'un certain nombre d'informations à la personne concernée.

Au-delà des obligations liées à la [transparence](#), il convient d'avertir les personnes des conséquences de l'ouverture de leurs données sur Internet, et notamment des risques engendrés, qui dépendront du contexte, de la nature des données, des modalités de leur diffusion, etc.

- **le consentement doit être univoque** : c'est-à-dire donné par une déclaration ou tout autre acte positif clair, sans ambiguïté ; la jurisprudence refuse les cases précochées pour l'expression du consentement<sup>2</sup>.

### Illustrations

À titre d'exemple, la constitution et la diffusion en *open data* d'une base de données d'échantillons de voix à des fins de recherche linguistique peuvent se fonder sur le consentement des participants. Il convient alors de prévoir un mécanisme adapté de collecte mais aussi de gestion des consentements permettant d'apporter la preuve de leur recueil.

<sup>2</sup> [Cour de justice de l'Union européenne, grande chambre, 1<sup>er</sup> octobre 2019, Planet49 GmbH, C-673/17, curia.europa.eu](#)

## Pour les réutilisateurs de données publiquement accessibles sur Internet

À noter : si les conditions propres à la base légale envisagée par chacune des questions ci-dessous ne sont pas remplies, l'organisme doit, soit modifier les paramètres de son projet de traitement pour parvenir à les respecter, soit rechercher une autre base légale.

### 1. Cette réutilisation de données personnelles est-elle nécessaire au respect d'une obligation légale ?

Pour que la base légale « obligation légale » puisse être retenue, il faut :

- qu'un texte suffisamment clair et précis exige de l'organisme, la mise en œuvre du traitement dans les conditions projetées ;
- qu'il n'existe, en d'autres termes, pas de moyen moins intrusif permettant de respecter les dispositions légales en cause.

### Illustrations

La CNIL a reconnu la validité de cette base légale pour la recherche par les assureurs des bénéficiaires de contrats d'assurance-vie non réclamés, notamment sur le web ou sur des réseaux sociaux publiquement accessibles, à condition que cette recherche ne soit pas automatisée et ne conduise pas à une collecte massive<sup>3</sup>.

**Néanmoins, l'obligation légale pourra rarement être invoquée en pratique**

### 2. Si non, est-elle soumise obligatoirement au consentement des personnes concernées en vertu d'une disposition légale ?

Certains textes imposent de fonder le traitement sur la base légale du consentement. Il en va ainsi, sauf exceptions, des traitements ayant pour objet la prospection par voie électronique (art. L34-5 du code des postes et communications électroniques), ou encore des traitements emportant, avec ou son profilage préalable, des prises de décisions automatisées aux effets significatifs pour les personnes (art. 22 du RGPD).

### 3. Si non, est-elle nécessaire à l'exercice d'une mission d'intérêt public qui m'a été légalement confiée ?

La possibilité de se fonder sur la base légale de la « mission d'intérêt public » suppose :

- que la mission dans laquelle s'inscrit le traitement soit prévue par un texte normatif applicable au réutilisateur ;
- que la réutilisation des données permette d'exercer *spécifiquement* cette mission (pas le cas si elle vise un objectif sans rapport particulier avec celle-ci ou trop éloigné de ses particularités), de manière pertinente et appropriée.

S'il ne s'agit pas d'une obligation, plus ces dispositions sont précises, plus il est facile pour les organismes concernés de recourir à cette base légale.

### Illustrations

C'est le cas du pôle d'expertise de la régulation numérique (PEReN) qui dispose d'un pouvoir de réutilisation des données publiquement accessibles sur les sites des opérateurs de plateforme afin de réaliser des expérimentations ayant notamment pour objet de concevoir ou évaluer des outils techniques destinés à la régulation des opérateurs de plateformes en ligne<sup>4</sup>.

### 4. Si non, la réutilisation de données peut-elle se fonder sur la poursuite d'un intérêt légitime ?

La réutilisation doit répondre à un intérêt légitime du réutilisateur et ne pas porter une atteinte excessive aux droits et intérêts des personnes concernées par les données. Les bénéfices attendus de la réutilisation, pour le

<sup>3</sup> Il s'agit de respecter la loi n° 2014-617 du 13 juin 2014 relative aux comptes bancaires inactifs et aux contrats d'assurance vie en déshérence.

<sup>4</sup> Conformément au décret n° 2022-603 du 21 avril 2022 fixant la liste des autorités administratives et publiques indépendantes pouvant recourir à l'appui du pôle d'expertise de la régulation numérique et relatif aux méthodes de collecte de données mises en œuvre par ce service dans le cadre de ses activités d'expérimentation.

responsable de traitement ou pour les tiers, les **attentes raisonnables** des personnes concernées par le projet de traitement, **ainsi que les impacts/risques qui en résultent et les garanties envisageables pour les réduire**, sont les critères déterminants pour apprécier si l'intérêt légitime du réutilisateur peut valablement justifier la mise en œuvre du projet.

### Trois cas de figure doivent être distingués :

- **Premier cas** : les données mises à disposition du public le sont en vertu de la législation sur l'open data, qui prévoit leur libre réutilisation. Dans ce cas, si la finalité poursuivie par le réutilisateur est elle-même légale, le traitement reposera presque toujours sur un intérêt légitime. On ne peut exclure que, malgré la mise à disposition des données pour réutilisation, certaines formes de traitement particulièrement intrusives portent aux intérêts des personnes concernées une atteinte excessive, qui les rendrait illégales sans le consentement des de celles-ci.
- **Deuxième cas** : la personne qui a diffusé la base de données l'a fait volontairement, pour permettre la réutilisation, mais celle-ci n'est pas prévue par le cadre légal. Dans ce cas, l'intérêt légitime du réutilisateur est le plus souvent acquis si la diffusion de la base était elle-même légale. Il convient en effet de tenir compte de ce que la mise en ligne de la base de données n'est elle-même légale que si le diffuseur a pu estimer que cette publication, qui expose les données à toutes formes de réutilisation, ne porte pas une atteinte disproportionnée aux personnes, ou, à défaut, a recueilli le consentement des personnes concernées (en tout état de cause, les personnes auront dû, en principe, être préalablement informées de la diffusion). Malgré cela, il convient de s'interroger à chaque fois, pour vérifier si l'utilisation qui est faite de la base peut se fonder sur un intérêt légitime ou nécessite le consentement des personnes.
- **Troisième cas** : des données personnelles ont été publiées sur internet mais ne l'ont pas été dans une perspective d'open data, mais simplement dans le cadre de l'activité du responsable de traitement (données figurant sur les réseaux sociaux, données d'articles de presse, données présentes sur un site de petites annonces etc.). Si le caractère d'ores et déjà public de la donnée doit être pris en compte pour apprécier l'atteinte que leur réutilisation pourrait porter à la vie privée des personnes concernées, et les risques auxquels elles pourraient être exposés, l'intérêt légitime du réutilisateur doit être apprécié au cas par cas.

### Dans le cadre de la mise en balance par le réutilisateur des intérêts en présence, plusieurs éléments peuvent être pris en compte, en particulier :

- le contexte de la collecte des données, leur finalité initiale, et le contexte de la publication des données en ligne (p. ex. : publication prévue ou non par un cadre légal, intervenant sur un site d'échanges d'ordre personnel ou professionnel) ;
- les catégories de personnes (p. ex. : personnalités publiques, mineurs) et de données concernées, plus ou moins sensibles (p. ex. : données professionnelles ou données relatives à la vie privée) ;
- les éventuelles restrictions découlant d'une licence de réutilisation ou des conditions générales d'utilisation d'un site web (p. ex. : interdiction de procéder à des moissonnages de données en ligne à des fins de prospection commerciale) ;
- les mesures susceptibles d'être adoptées pour limiter l'impact du traitement sur les personnes (ex. : anonymisation à bref délai, pseudonymisation, droit d'opposition inconditionnel).

### Illustrations

Par exemple, « l'intérêt légitime » pourrait valablement fonder :

- la rediffusion, par des sociétés spécialisées dans l'information légale et financière, des données d'entreprises diffusées par les administrations chargées de leur mise à disposition au public (INPI et INSEE) ;
- l'exploitation, à des fins de réalisation d'études sur l'évolution des prix de l'immobilier, de la base de données « demandes de valeurs foncières / DVF » obligatoirement diffusée par la DGFIP et faisant état des transactions immobilières des cinq dernières années, sans identification directe des personnes concernées) ;

- la collecte par un employeur de données publiées sur un réseau social professionnel afin de contacter un candidat potentiel, dès lors que les personnes concernées, en les partageant sur un tel site s'attendent raisonnablement à ce type de réutilisations ([voir la fiche n°14 du guide de la CNIL sur le recrutement](#)) ;
- l'exploitation, par l'établissement de statistiques intégrant des mécanismes de pseudonymisation et d'anonymisation, de données publiées sur des blogs, forums, réseaux sociaux, etc., à des fins de mesure de l'opinion sur un sujet, une personne ou un produit spécifique (détection de tendances, corrélations) ;
- la réutilisation, dans le strict respect des droits des personnes concernées (droit de rectification, en particulier), des données publiées sur des sites institutionnels aux fins d'améliorer la transparence des conditions d'exécution par les élus de leur mandat politique (élaboration et publication de contenus synthétisant leurs activités : taux de présence, nombre de rapports, sens de leurs votes sur différentes thématiques, etc.).

À l'inverse, « l'intérêt légitime » n'apparaît pas pouvoir valablement fonder :

- l'exploitation des coordonnées publiques des DPD / DPO, diffusées en open data par la CNIL (art. 83 du décret d'application de la loi Informatique et Libertés), à des fins de communication sur des sujets sans lien avec leur activité professionnelle (p. ex. : prospection politique) ;
- la réutilisation des coordonnées électroniques publiées par des utilisateurs de sites de petites annonces entre particuliers à des fins de commercialisation de bases de prospection (p. ex. : auprès d'agences immobilières, s'agissant d'annonces dédiées à la location / vente de logements) ;
- le recours à un procédé particulièrement intrusif et massif qui permet de récupérer les images présentes de façon éparse sur Internet de millions d'internautes en France, afin d'alimenter un système de reconnaissance faciale pouvant être utilisé par des États à des fins policières (voir à ce sujet la sanction publique rendue par la CNIL, [n° SAN-2022-019 du 17 octobre 2022](#)) ;
- le traitement mis en œuvre par une entreprise éditrice d'un site web compilant de multiples informations sur d'innombrables personnes (photos, vidéos, posts, CV, liens vers les pages web où elles apparaissent, etc.), tirées de diverses sources ouvertes (réseaux sociaux, blogs, etc.), et les restituant aux internautes sous forme de profils à partir de la saisie de données nominatives dans un moteur de recherche.

**À noter : tous les cas de rediffusion de données déjà publiées ne pourront pas se fonder sur la base légale de l'intérêt légitime.** Il existe effectivement une différence d'enjeux et d'impacts entre l'établissement de profils nominatifs de particuliers, sur la base d'une agrégation de toutes les données en ligne les concernant, et la « simple » rediffusion de données d'entreprises par ailleurs diffusées par des administrations dans le cadre de leurs missions légales.

Pour ce type de traitements, auxquels les personnes concernées ne sauraient raisonnablement s'attendre, seul le recueil de consentements valables auprès de celles-ci devrait permettre de garantir leur licéité.

#### 5. Si non, puis-je recueillir un [consentement](#) valable auprès des personnes concernées ?

Certains textes imposent le recueil d'un consentement. Il en va ainsi, sauf exceptions, des [traitements ayant pour objet la prospection par voie électronique](#) (art. L34-5 du code des postes et communications électroniques), ou encore des [traitements emportant, avec ou son profilage préalable, des prises de décisions automatisées](#) aux effets significatifs pour les personnes (art. 22 du RGPD). Ce consentement peut alors servir de base légale au traitement.

Pour être valable, il doit respecter un certain nombre de [conditions](#).

S'agissant des modalités de recueil du consentement, les possibilités suivantes méritent d'être évoquées.

- Dans l'hypothèse où l'organisme souhaitant réutiliser les données dispose du moyen de contacter les personnes concernées (p. ex. : utilisateurs de réseaux sociaux ayant un profil public comportant leurs adresses mail ou un outil de contact accessible par toute personne, blogueurs mettant à disposition un

formulaire de contact), il pourra, sans difficultés, solliciter directement auprès d'elles le recueil de leur accord.

- Dans l'hypothèse inverse, l'organisme souhaitant réutiliser les données peut demander à celui les ayant initialement publiées d'informer lui-même les personnes concernées de son projet de traitement et d'inviter celles n'y étant pas opposées à se rapprocher du réutilisateur pour lui fournir leur consentement. Cela suppose toutefois que l'organisme ayant publié les données soit en lien avec les personnes concernées et accepte de jouer le rôle d'intermédiaire (notamment, compte tenu du fait qu'il ne serait pas légitime à communiquer directement les données de contact qu'il détient au réutilisateur).
- Il existe enfin des cas où le consentement d'une personne à une réutilisation de ses données pour certaines finalités pourrait être donné par celle-ci au moment de la mise en ligne des informations la concernant, par exemple au moyen d'une case à cocher lors de la création de son compte sur un réseau social ou sur un site de petites annonces. Une attention particulière doit alors être portée, d'une part à la liberté du consentement, qui doit souvent pouvoir être accordé ou refusé finalité par finalité (le consentement à la publication n'emporte pas un consentement à toute forme de réutilisation) et, d'autre part, aux informations devant être portées à la connaissance des personnes, qui incluent notamment l'identité de chaque responsable de traitement réutilisateur. Cette option implique elle aussi une collaboration entre le diffuseur et le réutilisateur des données.

PROJET

## Fiche principe n°3 : comment informer les personnes concernées

Le RGPD impose d'informer les personnes concernées de l'ouverture ou de la réutilisation de leurs données sur internet. La CNIL fait le point sur les mesures à prendre pour respecter cette obligation et précise les cas dans lesquels une information publique générale pourra suffire.

### Pourquoi assurer la transparence des traitements ?

Le principe de transparence oblige les organismes collectant des données personnelles à en **informer les personnes afin qu'elles comprennent les usages** qui seront faits de leurs informations (pourquoi, comment) **et soient en mesure d'exercer leurs droits** (opposition, accès, rectification, etc.). Il contribue ainsi à la loyauté des traitements et à l'établissement de relations de confiance entre les organismes qui en sont responsables et les individus qu'ils concernent.

**Ce principe s'applique à tout traitement de données personnelles, que les données soient :**

- **directement recueillies auprès des personnes concernées** : par exemple, lors du renseignement en mairie d'un formulaire administratif, dans le cadre d'un entretien téléphonique, de l'ouverture en ligne d'un compte utilisateur, etc. ; ou
- **indirectement collectées** : collecte de données en libre accès sur Internet via le téléchargement de fichiers, le recours à des outils de moissonnage de données ou l'utilisation d'interfaces de programmation applicatives mises à disposition de réutilisateurs par les plateformes en ligne ; obtention d'informations auprès de partenaires institutionnels/commerciaux, , réutilisation d'une base de données déjà constituée, etc.

**Lorsque le responsable de traitement n'a pas directement collecté les données personnelles auprès des personnes concernées, il est dispensé de l'obligation d'informer ces personnes si cette information est impossible en pratique ou requerrait des efforts disproportionnés.**

### Quelles informations fournir et à quel moment ?

#### Pour les organismes diffusant une base de données sur Internet

**Lorsque la diffusion est prévue dès le moment où les données sont collectées, il faut informer systématiquement les personnes concernées.** Les organismes doivent ainsi leur indiquer, au moment où ils recueillent les données :

- leur **identité** et **coordonnées** (favoriser les différents modes de communication : adresses postale et électronique, téléphone, etc.), ainsi que les **moyens de contacter leur [délégué à la protection des données](#)** s'ils en ont désigné un ;
- la **finalité** et la **base légale du traitement d'ouverture en ligne des données** (p. ex. : l'« obligation légale » ou la « mission d'intérêt public » en cas de diffusion de données par une administration agissant en application du CRPA ou d'une législation spéciale), **avec, pour les organismes privés, des précisions sur l'« intérêt légitime » qu'ils poursuivent** si leur traitement se fonde sur celui-ci (p. ex. : les raisons pour lesquelles un réseau social met à disposition des chercheurs une API pour faciliter le retraitement des données diffusées). [Voir la fiche « cas d'usage » n°4 du guide](#) ;
- les **destinataires ou catégories de destinataires** des données (p. ex. : tous les internautes, les chercheurs, etc.), avec, le cas échéant, des précisions quant au **transfert** envisagé de celles-ci vers un pays tiers à l'Union européenne.

En principe, il est également nécessaire de préciser :

- la **durée du traitement** des données ou, lorsque ce n'est pas possible, les critères utilisés pour la déterminer
- l'existence de leurs **droits**, tel que le droit d'opposition (à mentionner clairement et séparément de toute autre information) si le traitement est fondé sur l'intérêt légitime ou la mission d'intérêt public,

ou de retirer son consentement à tout moment (si le traitement est fondé sur le consentement des personnes concernées). [Voir la fiche « principe » n°4 du guide.](#)

- le **droit d'introduire une réclamation auprès de la CNIL.**

Lorsque la diffusion concerne une base de données déjà constituée, l'information à fournir porte sur les mêmes éléments. En revanche, certains cas de dispense d'information peuvent trouver à s'appliquer (v. ci-dessous).

## Pour les organismes réutilisant des données en libre accès sur Internet

**Lorsque l'information individuelle des personnes est matériellement possible**, ce qui nécessite de disposer de leurs données de contact, le réutilisateur est en principe tenu de procéder à cette information. Elle doit avoir lieu aussi tôt que possible, et au plus tard lors de la première prise de contact avec les intéressés ou lors de la première communication des données à un autre destinataire s'il y en a une, et dans tous les cas **dans un délai ne dépassant pas un mois après la collecte**. Par ailleurs, lorsque l'information individuelle n'est pas possible, le réutilisateur doit en principe publier une information générale, collective (par exemple sur un site web), dans les meilleurs délais.

L'information doit comporter les points suivants :

- leurs **identité** et **coordonnées** (favoriser les différents modes de communication : adresses postale et électronique, téléphone, etc.), ainsi que les **moyens de contacter leur délégué à la protection des données** s'ils en ont désigné un ;
- la **finalité** et la **base légale de leur traitement avec, le cas échéant, des précisions sur « l'intérêt légitime » le fondant** (p. ex. : l'intérêt de l'entreprise à mesurer son « e-réputation » à partir d'une analyse des commentaires la concernant publiés en ligne). [Voir la fiche principe n°2 du guide.](#)
- les **catégories de données utilisées** (p. ex. : identité, coordonnées, images, posts sur les réseaux sociaux)
- les **destinataires ou catégories de destinataires** des données, avec, le cas échéant, des précisions quant au **transfert** envisagé de celles-ci vers un pays tiers à l'Union européenne.

En principe, il est également nécessaire de préciser :

- la **durée de traitement** des données, ou, lorsque ce n'est pas possible, les critères utilisés pour déterminer cette durée ;
- **la ou les source(s) accessible(s) au public** (p. ex. : sites web) **d'où elles proviennent**. **À noter** : lorsque plusieurs sources ont été exploitées et qu'il n'est pas possible d'attribuer les données d'une personne à une source en particulier, des informations générales sur ces sources doivent être fournies ; ces informations générales devraient toutefois préciser les différentes sources en cause lorsqu'elles sont connues (p. ex. : noms des différents réseaux sociaux sur lesquels les données auront été collectées) ;
- le cas échéant, la **mise en œuvre d'une décision entièrement automatisée**, les informations utiles à la compréhension de l'algorithme, **et les conséquences pour la personne concernée** ;
- l'existence de leurs **droits**, tel que le droit d'opposition (à mentionner clairement et séparément de toute autre information) si le traitement est fondé sur l'« intérêt légitime » ou la « mission d'intérêt public », ou de retirer son consentement à tout moment (si le traitement est fondé sur le consentement des personnes concernées). [Voir la fiche « principe » n°4 du guide.](#)
- le **droit d'introduire une réclamation auprès de la CNIL.**

## Comment la délivrer en pratique ?

**Les personnes concernées ne doivent pas rencontrer de difficultés dans l'accès à l'information comme dans sa compréhension** : elle doit être bien distinguée des autres indications sans lien avec la protection des données (support distinct des CGU), être aussi succincte et claire que possible (vocabulaire simple, phrases courtes, style direct, etc.) et adaptée aux conditions d'interaction avec les personnes.

Il existe ainsi **plusieurs moyens pour la fournir** : elle peut figurer sur le formulaire en ligne utilisé par le diffuseur des données pour recueillir celles-ci, ou transmise dans un document dédié en cas de collecte en face-

à-face ; être mentionnée dans les courriels ou courriers adressés par un réutilisateur des données lors de son premier contact avec les personnes concernées ; être délivrée via un message vocal pré-enregistré, etc.

Pour atteindre l'objectif de concision et de bonne lisibilité, il **peut être procédé à une information en plusieurs niveaux, priorisant les informations essentielles que sont l'identité du responsable du traitement, les finalités et les droits des personnes.**

**Exemple :** ces informations sont données à la personne concernée directement sur la page d'inscription aux services proposées par une plateforme de communication en ligne, et sur cette même page, un lien renvoie vers une notice d'information complète.

À noter qu'en fonction du contexte, d'autres informations pourraient être délivrées prioritairement, telles que les conséquences prévisibles de la diffusion des données personnelles (p. ex. : indexation par des moteurs de recherche externes), ou les caractéristiques importantes de leur réutilisation (p. ex. : réalisation d'un profilage, transmission des données à des partenaires commerciaux).

## Quels sont les cas dans lesquels la délivrance d'une information individuelle n'est pas obligatoire ?

**Dans certains cas prévus par les textes, les responsables de traitements peuvent ne pas procéder à une information individuelle des personnes concernées.**

**Pour déterminer si l'information individuelle des personnes est requise ou non, il convient de se référer aux fiches particulières déclinant les règles générales à chaque type de réutilisation.**

### Cas n° 1 : la personne concernée a déjà obtenu les informations

**Par exemple,** cette hypothèse correspond au cas où les données ont été collectées sur Internet par un laboratoire de recherche pour un suivi de cohorte à l'année « n » ; les chercheurs les réutilisent pour les comparer avec les résultats de « n+1 » et les personnes concernées avaient bien été averties de la durée du suivi lors de la première collecte.

**Attention :** le fait que, dans le cadre de leur CGU ou politique de confidentialité, des réseaux sociaux et autres plateformes de communication en ligne signalent la possibilité d'une réexploitation des données par des tiers ne permet pas à leurs réutilisateurs de considérer les personnes comme étant « déjà informées » : l'information fournie doit en effet être particulière à chaque traitement des données mis en œuvre par chaque organisme concerné.

### Cas n°2 : la fourniture des informations rendrait impossible ou compromettrait gravement la réalisation des objectifs poursuivis

**Par exemple :** cette exception peut notamment concerner **des travaux de recherche scientifique** impliquant l'analyse de données diffusées sur les réseaux sociaux, si l'information des personnes concernées risque de les conduire à modifier leur comportement en ligne et, ainsi, de biaiser les travaux en question.

Lorsque cette exception apparaît mobilisable, **le responsable du traitement doit prendre des mesures appropriées** pour protéger les droits, libertés et intérêts légitimes des personnes concernées (ce point, qui concerne également le cas n°3, est précisé ci-dessous).

### Cas n°3 : l'information se révèle impossible ou exigerait des efforts disproportionnés

Cet argument est souvent invoqué par les organismes privés qui réutilisent des données en libre accès sur Internet, dans la mesure où ils ne disposent généralement pas des données de contact des personnes dont ils

collectent les données et où ceux qui les détiennent ne sont pas autorisés à les communiquer à toute personne souhaitant les exploiter. **Une analyse au cas par cas est à réaliser**, tenant compte du contexte spécifique de chaque traitement.

**Concernant les efforts disproportionnés, deux cas de figure peuvent être schématiquement distingués.**

*Le réutilisateur des données dispose des coordonnées des personnes concernées, ou peut facilement et légalement y accéder*

**Dans cette hypothèse, le caractère disproportionné d'une information directe et individuelle est plus difficilement caractérisable.**

Le caractère proportionné s'apprécie en mettant en rapport :

- **d'une part, l'atteinte portée à la vie privée des personnes dont les données sont traitées** ; par exemple, le fait d'utiliser des données publiques pour une étude interne à un organisme porte une atteinte moindre à la vie privée des personnes que le fait de les republier ou de les enrichir, en les rendant facilement accessibles ;
- **d'autre part, la difficulté et le coût d'une information individuelle.**
  - Lorsqu'une information par voie électronique est possible, elle est généralement requise.
  - Lorsque les données disponibles sont des données postales, une analyse au cas par cas est nécessaire, en prenant en compte notamment le coût de l'information, la faiblesse ou l'importance des risques que le traitement peut faire encourir aux personnes, le fait que les personnes peuvent ou non raisonnablement s'attendre au traitement de leurs données.

Il faut se garder de tout raisonnement automatique : ce n'est pas parce que les personnes sont très nombreuses que le responsable de traitement est systématiquement dispensé de l'obligation d'informer, il faut prendre en compte les autres paramètres et notamment l'intrusivité du traitement.

**Par exemple :**

- l'extraction par une entreprise des données de profils de réseaux sociaux (photos, établissements scolaires, employeurs, etc.) pour enrichir un service d'annuaire téléphonique ne peut légalement intervenir sans une information individuelle préalable (si 25 millions de personnes sont concernées, celle-ci ne constitue pas un « effort disproportionné » compte tenu du caractère intrusif du traitement et de la détention des coordonnées, a jugé le Conseil d'État – arrêt du 12 mars 2014, [n° 353193](#)) ;
- de même, et bien que les personnes puissent s'y attendre davantage, l'enregistrement par des cabinets de recrutement de données publiées sur les réseaux sociaux professionnels, à des fins d'enrichissement de bases de candidatures, doit s'accompagner de la délivrance d'une information à chacune des personnes concernées (voir la fiche n°14 du [guide de la CNIL sur le recrutement](#)).

**À l'inverse, certains réutilisateurs pourront se prévaloir de cette exception, notamment :**

- si l'atteinte portée à la vie privée est particulièrement faible ;
- ou si les données ne sont pas conservées sous une forme permettant l'identification directe des personnes concernées.

**Par exemple :**

- j'ai préparé une note interne pour mon entreprise avec des éléments biographiques sur un parlementaire, à partir d'informations disponibles sur Internet : je n'ai pas besoin de l'informer de ce traitement, qui porte à sa vie privée une atteinte particulièrement faible, quand bien même je dispose de ses données de contact.
- une équipe de recherche réalise un projet de recherche scientifique sur les contenus du web social, intégrant des mesures [d'anonymisation ou de pseudonymisation](#) à bref délai ; ou une étude sur le

fonctionnement des institutions démocratiques impliquant l'exploitation des listes publiques des personnes auditionnées par les parlementaires : ici encore, l'exception à l'information pourra être invoquée, compte tenu du nombre de personnes concernées et des contraintes importantes pour la recherche systématique de leurs coordonnées.

*Le réutilisateur des données ne dispose pas des coordonnées des personnes, ou seulement d'anciennes informations, à l'exactitude incertaine (plus de 10 ans, par exemple).*

**Dans cette hypothèse, le caractère disproportionné d'une information individuelle pourra plus facilement être reconnu.**

En particulier :

- lorsque les caractéristiques du traitement (finalité, portée, nature des données, garanties apportées, attentes raisonnables des personnes, etc.) ne l'impose pas (p. ex. : exploitation, à des fins d'études statistiques sur l'évolution des prix de l'immobilier, de la base de données « demandes de valeurs foncières / DVF » obligatoirement diffusée en *open data* par la direction générale des finances publiques et comprenant, à l'adresse, les transactions immobilières des cinq dernières années, sans identification directe des personnes concernées) ;
- lorsque l'atteinte portée à la vie privée par le traitement est particulièrement faible ;
- lorsque soit l'accès aux moyens de contact des personnes concernées ne paraît pas aisé ou pas souhaitable (p. ex. : personnes utilisant des pseudonymes, données pseudonymisées par le responsable du traitement initial), soit l'envoi induirait un coût trop lourd à supporter (p. ex. : nombre élevé de personnes et absence d'adresses mail facilement et légalement accessibles).

#### **Attention**

Afin de procéder à l'information, le réutilisateur sera parfois conduit à demander des coordonnées de contact à l'éditeur du site sur lequel il a récupéré les informations. Le fait que l'éditeur procède lui-même à l'information pour le compte du réutilisateur, qu'il y soit légalement tenu ou qu'il accepte d'y procéder, ne pose a priori pas de difficulté. En revanche, la transmission des coordonnées de contact au réutilisateur, qui constitue elle-même un traitement soumis au RGPD, n'est souvent pas légitime et nécessite le consentement des personnes.

#### **À retenir**

**Tout réutilisateur de données souhaitant s'appuyer sur cette exception prévue à l'article 14.5 du RGPD doit :**

- **vérifier la réalité du caractère « impossible » ou « disproportionné » de la délivrance d'une information individuelle, en mettant en balance, dans ce deuxième cas, l'importance de l'atteinte à la vie privée (compte tenu des mesures envisagées pour la réduire : réduction du nombre de données collectées et de leur durée de conservation, garanties d'un niveau élevé de sécurité, [réalisation d'une AIPD](#), [anonymisation ou pseudonymisation des données](#), etc.), les efforts qu'une telle information requerrait et les effets que son absence pourrait avoir sur les personnes ;**
- **pouvoir en justifier à tout moment, notamment en documentant son analyse, conformément au principe de responsabilité.**

**Dans tous les cas, il doit également procéder systématiquement à la délivrance d'une information publique générale et complète, en n'hésitant pas à multiplier les supports de diffusion lorsque les personnes concernées sont nombreuses (publications sur le site web du réutilisateur, sur les comptes de réseaux sociaux qu'il utilise, dans un journal, etc.).**

*Cas n°4 : le traitement est mis en œuvre aux fins d'expression universitaire, artistique ou littéraire, ou d'exercice, à titre professionnel, de l'activité de journaliste*

L'article 80 de la loi Informatique et Libertés prévoit que le droit à l'information peut être écarté pour les « traitements mis en œuvre aux fins [...] d'expression universitaire, artistique ou littéraire », lorsqu'une telle dérogation « est nécessaire pour concilier le droit à la protection des données à caractère personnel et la liberté d'expression et d'information ».

Une dérogation identique est prévue pour les journalistes professionnels (au sens de l'article [L7111-3 du code du travail](#)) qui se livrent à des activités d'investigation « en sources ouvertes ».

La protection des données personnelles doit en effet être conciliée avec la liberté d'expression et d'information. Une information minimale sur le traitement doit être fournie dans la publication, et en particulier l'identité du responsable de traitement. Cette exigence doit se combiner avec les règles spécifiques régissant les publications de presse.

## **Autres cas prévus par des dispositions légales constituant une mesure nécessaire et proportionnée pour garantir des objectifs d'intérêt public particulièrement importants**

**Par exemple : en application de l'article 48 de la loi Informatique et Libertés**, le droit à l'information individuelle ne s'applique pas dans la mesure où une telle limitation est nécessaire au respect des fins poursuivies par ce traitement :

- aux traitements portant sur des données indirectement collectées et utilisées lors d'un traitement de l'État intéressant la sécurité publique ;
- aux traitements mis en œuvre par les administrations ayant pour mission de contrôler ou recouvrer des impositions, ou d'effectuer des contrôles de l'activité de personnes physiques ou morales pouvant donner lieu à la constatation d'une infraction ou d'un manquement, à des amendes administratives ou à des pénalités.

### **À noter :**

**Les textes instaurant les traitements concernés et autorisant une telle dérogation doivent contenir des dispositions spécifiques**, prévues par l'article 23 du RGPD, telles que les finalités du traitement en cause, l'identité de son responsable (ou l'identification des catégories de responsables), les catégories de données traitées et leur durée de conservation, les risques pour les droits et libertés des personnes concernées, ainsi que les garanties destinées à prévenir les abus.

## **Références**

- [Articles 12, 13 et 14 du RGPD](#)
- [Conformité RGPD : comment informer les personnes et assurer la transparence ?, cnil.fr](#)
- [L'information, design.cnil.fr](#)
- Voir les exemples de mentions d'information figurant dans les fiches de ce guide dédiées à certains cas d'usage

## Fiche principe n°4 : Quels sont les droits des personnes sur leurs données ?

Les personnes dont les données sont partagées sur Internet, ou collectées en ligne à des fins de réutilisation, disposent de plusieurs droits sur celles qui les concernent. Ces droits vont leur permettre de conserver la maîtrise de leurs données et il appartient aux organismes responsables des traitements de les respecter et d'en faciliter l'exercice

### De quels droits s'agit-il ?

En plus de devoir être informées de la mise en œuvre des traitements (voir à ce [sujet la fiche n° 3 de ce guide](#)), les personnes concernées par les traitements de diffusion publique de données sur Internet et de réutilisation de données publiquement accessibles disposent de **différents droits qui constituent un ensemble de leviers d'action concrets leur permettant de contrôler les usages (objectifs poursuivis, conditions pratiques)** qui sont faits de leurs données.

#### Attention :

**Le choix de la base légale** (voir à ce sujet [la fiche « principe » n°2 de ce guide](#)) **conditionne l'exercice de certains droits.**

Pour aider les organismes dans l'identification des droits applicables à leurs traitements, la CNIL a établi le tableau récapitulatif suivant. Les conditions d'exercice et de respect de ces droits, comme les cas dans lesquels il est exceptionnellement permis d'y déroger, sont détaillés plus bas.

Droits de la personne	Accès	Rectification	Opposition	Effacement	Limitation	Portabilité
Base légale						
Consentement	✓	✓	✗ <sup>(1)</sup>	✓	✓	✓
Respect d'une obligation légale	✓	✓	✗	✓ / ✗ (2)	✓	✗
Mission d'intérêt public	✓	✓	✓	✓ / ✗ (2)	✓	✗
Intérêt légitime	✓	✓	✓	✓	✓	✗

(1) : à noter que si la personne concernée ne peut pas s'opposer elle peut retirer son consentement au traitement.

(2) : le droit peut être écarté par le responsable de traitement dans un certain nombre de cas.

**En matière de partage et de réutilisation de données publiquement accessibles, les principaux droits applicables sont les suivants.**

- **le droit d'accès** : les personnes peuvent demander à l'organisme la confirmation qu'il détient des données les concernant, en obtenir une copie pour en vérifier le contenu

**et solliciter certaines informations relatives aux caractéristiques du traitement mis en œuvre (finalités poursuivies, durées de conservation, identité des destinataires ou catégories de destinataires, sources des données, etc.) ;**

**Exemple :** une personne peut demander à un cabinet de recrutement s'il a collecté sur le web des données la concernant et, le cas échéant, solliciter la transmission d'une copie de celles-ci (dans un document qui les synthétise, notamment), ainsi que de toute information disponible sur les sites exploités et opérations réalisées (en particulier, existence éventuelle d'une [prise de décision automatisée et/ou d'un profilage](#) et conséquences associées).

- **[le droit de rectification](#) :** les personnes peuvent demander à l'organisme à ce que soient corrigées les données inexactes les concernant, ou à ce que soient complétées celles qui sont en lien avec la finalité du traitement ;

**Exemple :** un élu dont les données, diffusées sur des sites institutionnels, sont réutilisées par une association s'étant donné pour mission de faciliter l'accès au fonctionnement des institutions démocratiques, en particulier en établissant des statistiques sur l'exercice des mandats électifs, pourra demander à ce que soient corrigées les données statistiques les concernant et faisant l'objet d'une publication (p. ex. : taux de présence à l'assemblée, nombre d'amendements déposés, niveau de fidélité à la ligne de son parti).

- **[le droit de retirer son consentement](#) :** lorsqu'un traitement de données est fondé sur le consentement préalable des personnes, celles-ci peuvent, à tout moment, sans justification particulière et via une modalité simple et équivalente à celle utilisée pour le recueillir, retirer leur consentement ;

**Exemple :** l'utilisateur d'un site web de petites annonces relatives à des biens immobiliers, qui a mis en ligne une annonce en laissant ses coordonnées et qui a consenti à la réutilisation de ses données à des fins de prospection commerciale par des agences immobilières, doit pouvoir retirer son consentement quand il le souhaite ([voir à ce sujet la fiche principe n°2 de ce guide](#)).

- **[le droit d'opposition](#) :** dans de nombreux cas (voir le tableau en page précédente), et notamment lorsque le traitement des données n'est pas fondé sur une obligation légale<sup>5</sup>, les personnes peuvent s'opposer à tout moment, et pour des raisons tenant à leur situation particulière, au traitement de leurs données ; l'organisme doit alors cesser celui-ci, sauf s'il démontre qu'il existe des motifs légitimes et impérieux pour le poursuivre ou qu'il est nécessaire pour la constatation, l'exercice ou la défense de droits en justice ;

**Exemple :** une personne figurant dans les fichiers d'un organisme ayant pour activité la représentation d'intérêts (p. ex. : fichier de contact, cartographie des influenceurs), et dont les données ont été collectées en ligne (forums de discussion, réseaux sociaux, blogs, etc.) au titre de leur participation active au débat public sur des sujets cibles de l'organisme, doivent pouvoir s'opposer au traitement de leurs informations auprès de celui-ci (voir à ce [sujet la fiche principe n°3 de ce guide](#)).

À noter : si la finalité est la prospection, le droit d'opposition est discrétionnaire.

- **[le droit à l'effacement](#) :** les personnes peuvent obtenir la suppression de leurs données, dans un certain nombre de cas, notamment quand ces données ne sont plus nécessaires au regard de l'objectif poursuivi, font l'objet d'un traitement illicite, sont traitées à des fins de prospection ou ont été collectées par une plateforme de communication en ligne, en particulier lorsqu'elles étaient mineures, ou encore quand ces personnes ont retiré leur consentement, ou se sont opposées au traitement et qu'il n'existe pas de motif légitime impérieux justifiant la poursuite celui-ci ;

---

<sup>5</sup> Voir en annexe la liste d'exemples de diffusions de données en *open data* pour lesquelles le droit d'opposition n'est pas applicable.

**Exemple :** les courtiers en données (« data-brokers ») spécialisés dans la collecte de données sur Internet, à des fins de commercialisation de celles-ci auprès d'entreprises ou de partis politiques, sont tenus de supprimer les informations se rapportant aux personnes qui en font la demande.

- **le droit à la limitation :** les personnes peuvent demander à ce que leurs données soient temporairement « gelées » dans un certain nombre de cas, notamment lorsqu'elles exercent leurs droits d'opposition ou de rectification ; au cours du délai dont dispose l'organisme pour y répondre, celui-ci ne pourra pas les utiliser (sauf exceptions) ;

**Exemple :** une personne qui s'oppose à la poursuite de la diffusion d'une donnée la concernant par une administration peut demander à celle-ci la cessation de la publication le temps qu'elle recherche d'éventuels motifs lui permettant de ne pas définitivement mettre un terme à une telle diffusion.

- **le droit à la portabilité des données :** lorsque les données sont traitées sur le fondement du consentement ou du contrat, les personnes peuvent recevoir celles qui les concernent et qu'elles ont fournies au responsable de traitement, les réutiliser et/ou les transmettre à un autre responsable de traitement.

**Exemple :** une personne consent à la mise à disposition en open data des données qu'elle a fournies à une entreprise, notamment dans une logique d'altruisme. Elle peut alors recevoir ces données dans un format pertinent (c'est-à-dire structuré, couramment utilisé et lisible par ordinateur) et même demander à ce qu'elles soient transmises directement à un autre organisme lorsque cela est techniquement possible.

## Quelles sont les conditions d'exercice de ces droits ?

Ils doivent pouvoir être exercés sur simple demande, écrite ou orale, la personne concernée pouvant justifier de son identité par tout moyen.

**Attention :** ne demander la fourniture d'une pièce d'identité que s'il existe un doute raisonnable.

Ce ne sera pas le cas, en particulier, si la personne exerce ses droits depuis un compte utilisateur créé au préalable, ou en utilisant la même adresse courriel que celle qu'elle a toujours utilisée pour ses contacts avec l'organisme.

De plus, l'exercice de l'ensemble des droits est gratuit. Toutefois, des frais raisonnables peuvent exceptionnellement être demandés dans certains cas d'exercice du droit d'accès (p. ex. : demande d'une copie supplémentaire).

## Comment les respecter en pratique ?

L'organisme doit mettre en place un dispositif qui non seulement garantit l'effectivité des droits des personnes concernées, mais également facilite leur exercice.

### Exemples de bonnes pratiques

- Fournir un formulaire de contact, un numéro de téléphone et/ou une adresse de messagerie dédié(e) à l'exercice des droits ;
- Si l'organisme dispose d'un site web intégrant des comptes utilisateurs, donner la possibilité aux personnes de les exercer à partir de leur espace personnel ;
- Si le traitement consiste dans une mise à disposition publique de données à des fins de réutilisation par des tiers (p. ex. : *open data* des collectivités publiques), permettre aux personnes concernées de s'opposer « d'emblée » à certains usages de leurs données, notamment à des fins commerciales, et communiquer sur ces oppositions auprès des éventuels réutilisateurs (p. ex. : publication d'un fichier d'opposition, ou encore d'un « étiquetage » des informations en cause directement dans la base de données).

**L'organisme doit garantir qu'il répondra aux demandes dans les meilleurs délais et, en principe, au plus tard dans un délai d'un mois.** Ce délai peut être prolongé de deux mois en raison de la complexité de la demande ou du nombre de demandes que l'organisme aurait reçu de cette même personne. Dans ce cas, l'organisme doit informer la personne concernée des raisons de cette prolongation dans le délai d'un mois.

**Recommandation :** mettre en place une procédure interne prévoyant les conditions de gestion et de suivi des demandes d'exercice des droits.

### Points de vigilance concernant le traitement des demandes de droit d'accès

Les données doivent être communiquées :

- sous une forme électronique d'usage courant, lorsque la demande est faite en ligne, à moins que la personne concernée souhaite qu'il en aille autrement (p. ex. : fourniture orale ou par courrier postal) ;
- sous une forme compréhensible (les codes, sigles et abréviations utilisés doivent être explicités) ;
- via la transmission des supports originaux ou d'un document faisant la synthèse des données traitées (p. ex. : quelles informations pour quelles finalités) ;
- dans le respect des droits des tiers (secret des affaires, propriété intellectuelle, droit à la vie privée, secret des correspondances).

**Exemple :** un cabinet de recrutement, qui collecte sur des réseaux sociaux professionnels les profils susceptibles d'intéresser certains de leurs clients, doit veiller à ce que les informations se rapportant aux personnes ayant pu commenter ces profils ne soient pas communiquées aux individus auxquels ces profils se rapportent.

### Point de vigilance concernant le traitement des demandes de rectification, d'effacement et de limitation

À moins qu'une telle information soit impossible ou exige des efforts disproportionnés, l'organisme doit **informer chaque destinataire auquel il a communiqué les données des rectifications, effacements d'informations et limitations du traitement auquel il a procédé** en réponse aux demandes de personnes concernées ([article 19](#) du RGPD).

En particulier, **lorsqu'un diffuseur de données en ligne est tenu d'effacer les données qu'il a rendues publiques**, il doit, en tenant compte des technologies disponibles et coûts de mise en œuvre, prendre les mesures raisonnables pour informer les tiers qui réutilisent ces données que la personne concernée a demandé l'effacement de tout lien vers ces informations, ou de toute copie ou reproduction de celles-ci ([article 17](#) du RGPD).

**Une telle information peut être livrée par différents moyens, y compris techniques.**

**Exemple :** un dispositif spécifique, permettant au détenteur de données d'informer les réutilisateurs des suites apportées aux demandes des personnes concernées, devrait être prévu dès la conception de l'outil organisant le partage (API dédiée à la communication d'informations relatives à l'exercice des droits, balisage des données, association de métadonnées aux données, etc.).

## Dans quels cas est-il possible d'y déroger ?

**De façon générale, l'organisme peut refuser de donner suite aux demandes dont il est saisi lorsqu'il est en capacité de démontrer :**

- qu'elles sont manifestement infondées ou excessives, notamment en raison de leur caractère répétitif ;
- ou qu'elles se rapportent à des données ne permettant pas l'identification des personnes concernées. Cependant, les personnes concernées pourront dans certains cas fournir des

informations complémentaires permettant à l'organisme de leur « rattacher » des données et de faire ainsi suite à leur demande.

**Attention** : si le responsable du traitement ne donne pas suite à la demande formulée par la personne concernée, il doit informer celle-ci, au cours du mois suivant la réception de la demande, des motifs de son inaction et de la possibilité d'introduire une réclamation auprès d'une autorité de contrôle et de former un recours juridictionnel.

**De plus, certains droits connaissent des limites, qui vont découler de textes encadrant spécifiquement certains traitements (p. ex. : [décret n° 2021-148 du 11 février 2021](#)) ou directement du RGPD et de la loi Informatique et libertés ». Ainsi par exemple, le droit à l'effacement ne s'applique pas dans certaines hypothèses, notamment quand le traitement des données est nécessaire :**

- **à l'exercice du droit à la liberté d'expression et d'information :**

p. ex. : dans certains cas, un journaliste peut refuser de supprimer des données figurant dans un de ses articles publiés en ligne au nom de la liberté dont il dispose en la matière, de même qu'un moteur de recherche peut refuser le [déréférencement](#) de certains contenus au nom du droit du public à accéder à l'information.

- **à la constatation, à l'exercice ou à la défense de droits en justice :**

p. ex. : une entreprise qui, dans le cadre de la préparation ou de la gestion d'un contentieux, a collecté des données publiées en ligne pour disposer de preuves sur la diffamation publique dont elle aurait été l'objet, peut ne pas donner suite à une demande d'effacement.

- **à la réalisation de statistiques, ou de [recherches scientifiques](#) ou historiques,** dans les conditions prévues par le RGPD et sous réserve que l'application d'un tel droit rende impossible ou compromette gravement la mise en œuvre de ces dernières.

Sous la même réserve que précédemment, **les finalités « statistiques » et « recherches » permettent également de déroger au droit d'opposition, au droit à la limitation, ainsi qu'au droit d'accès**, à condition, pour celui-ci, que les données soient conservées pour une durée limitée et sous une forme excluant manifestement tout risque d'atteinte à la vie privée (p. ex. : les personnes sont identifiées par un code attribué aléatoirement).

#### **Point de vigilance**

Dans le cadre de l'information individuelle ou générale devant être effectuée par l'organisme sur la mise en œuvre du traitement (voir la fiche dédiée), il convient d'indiquer aux personnes concernées le fait que leurs droits font l'objet de restrictions. Les motifs du refus de l'exercice d'un droit aux personnes concernées qui en ont fait la demande devraient également être expliqués, dans un langage compréhensible du grand public.

#### **Références**

- [Articles 12, 15 à 21 du RGPD](#)
- [Articles 49 à 56 de la loi Informatique et Libertés](#)

## Fiche principe n°5 : Comment garantir la minimisation des données traitées ?

### Le principe de minimisation : de quoi s'agit-il ?

Ce principe impose que les données personnelles traitées soient « *adéquates, pertinentes et limitées à ce qui est nécessaire au regard des finalités poursuivies* » (article 5.1.c du RGPD).

Pour garantir le respect de ce principe, en particulier au moyen des mesures développées ci-dessous, il est essentiel que, **préalablement à la mise en œuvre de leur traitement, les organismes qui diffusent ou qui réutilisent des données publiquement accessibles sur Internet :**

- **aient clairement défini l'objectif** poursuivi ;
- **se soient assurés de la légitimité de cette finalité** (art. 5.1.b du RGPD), notamment au regard d'obligations légales leur incombant, ou encore de l'intérêt que le traitement présente pour des tiers ou pour eux-mêmes ; par exemple :
  - l'intérêt du public à connaître des données détenues par une administration et associées à la gestion d'un service public, que ce soit pour contrôler les conditions d'exercice de l'action publique ou pour développer des services innovants d'ordre économique ou social ; ou
  - l'intérêt d'une entreprise à exploiter des données librement accessibles sur des réseaux sociaux professionnels à des fins de recrutement ;
- **identifient les catégories de données utiles et proportionnées** pour satisfaire ces obligations/intérêts, et qu'ils pourront par conséquent légalement traiter ; ce, sous réserve dans certains cas du respect complémentaire de certaines dispositions spécifiques, comme celles relatives aux conditions de traitement des données sensibles (p. ex. : données se rapportant à la santé, à la vie sexuelle, aux opinions politiques, convictions religieuses, etc.) ;
- **définissent les mesures techniques et organisationnelles** propres à garantir la bonne application du principe de minimisation, et à pouvoir la démontrer conformément aux principes « *de responsabilité du responsable du traitement* » et de « *protection des données dès la conception et par défaut* » ([art. 24 et 25](#) du RGPD).

### En pratique : quelles mesures adopter pour le respecter ?

#### Vous êtes un diffuseur de données

**La mise à disposition du public d'une base de données constitue une opération particulière au regard du principe de minimisation :** alors que ce principe interdit généralement de conserver des données personnelles « au cas où elles seraient utiles », « pour une éventuelle réutilisation », il est dans ce cas possible de mettre à disposition, et donc de conserver, des données en vue d'une utilisation incertaine par des tiers réutilisateurs. Cela n'est évidemment envisageable que si la mise en ligne de la base est elle-même légale, notamment parce qu'elle répond à une obligation légale ou ne porte pas aux droits des personnes une atteinte disproportionnée (v. fiche « Base légale »).

**Même dans ce cadre très particulier, les diffuseurs sont tenus de fournir certains efforts pour minimiser les données personnelles,** en envisageant les mesures qui suivent.

*Anonymiser les données, lorsque cela est possible*

L'anonymisation est un traitement qui consiste à **rendre impossible en pratique toute identification de la ou des personnes concernées à partir d'un jeu de données, par quelque moyen que ce soit et de manière irréversible**. Elle permet le partage et la réutilisation de « gisements » de données sans aucun risque d'atteinte à la vie privée des intéressés et, par conséquent, sans que la législation relative à la protection des données personnelles ne trouve à s'appliquer.

L'anonymisation devrait ainsi être envisagée toutes les fois où :

- il n'y a pas de disposition légale imposant la diffusion des données en cause sous une forme directement ou indirectement identifiante ;
- l'anonymisation n'aurait pas pour effet de nuire à la satisfaction de l'objectif poursuivi.

#### À noter :

- **L'anonymisation des données sera obligatoire dans certains cas, notamment pour l'open data des données publiques**

L'article L. 312-1-2 du code des relations entre le public et l'administration (CRPA), qui fixe le régime général de l'ouverture des données publiques, prévoit ainsi que les données diffusées sont, par principe et sauf exceptions, anonymisées (voir le [guide CADA-CNIL](#) et [la fiche pratique dédiée à ce sujet](#)).

- **Le caractère anonyme d'un jeu de données ouvert est souvent difficile à caractériser**

Si les données sont mises à la disposition du public, dans de nombreux cas, le risque d'identification est élevé, notamment du fait de la multiplicité des informations disponibles, sur Internet en particulier, et des capacités croissantes des techniques de recoupement de celles-ci. Les risques de réidentification doivent être évalués avec attention, en particulier lorsqu'il ne s'agit pas seulement de statistiques agrégées.

Le caractère anonyme des informations n'aura, par exemple, pas pu être établi pour les cas suivants d'ouverture en ligne de données :

- publication par un fournisseur d'accès internet, pour des réutilisations notamment à des fins de recherche scientifique, d'une base de données rassemblant 20 millions de recherches effectuées sur son site par 650 000 utilisateurs ; si chaque identifiant (nom d'utilisateur, adresse IP, etc.) avait été remplacé par un nombre choisi aléatoirement, l'historique de recherche d'un individu en dit beaucoup sur celui-ci (localisation, sujets d'intérêt, âge, profession, etc.), tant et si bien que certains internautes ont pu être réidentifiés ;
- publication par un organisme public d'une cartographie de la France divisée en carrés de 200 mètres de côté, auxquels étaient associées des données socio-démographiques telles que l'imposition moyenne des habitants ; dans la mesure où certains de ces carrés, situés dans des territoires peu peuplés, ne comptaient qu'un seul foyer fiscal, il avait ensuite été aisé, au moyen de simples recoupements de données, de retrouver l'adresse et l'identité des propriétaires concernés.

*À défaut d'anonymisation préalable, il convient de tenir compte des éléments suivants*

- **Respecter les dispositions légales encadrant le champ des données à diffuser, le cas échéant**

Dans certains cas, un texte précise les données personnelles devant être publiées en open data ou, à l'inverse, dont la publication est interdite. Le **responsable de traitement concerné doit alors veiller à respecter le cadre fixé.**

**Exemples d'articles prévoyant la publication de certaines données :** article [L. 123-52](#) du code de commerce pour les données figurant dans le registre national des entreprises diffusé par l'INPI ; [article R\\*. 112 A-1 du livre des procédures fiscales](#) pour celles relatives aux dernières transactions immobilières diffusées par l'administration fiscale ; [article 83 du décret d'application de la loi Informatique et Libertés](#) pour celles se rapportant aux désignations de délégués à la protection des données diffusées par la CNIL.

Comme indiqué par la CADA et la CNIL dans le guide précédemment évoqué sur l'ouverture des données publiques, **lorsque la diffusion sans anonymisation constitue une faculté pour les administrations (art. D312-1-3 du CRPA), l'opportunité d'une telle diffusion doit être appréciée au regard de l'intérêt du public** à connaître de données revêtant un caractère personnel et des risques corrélatifs d'atteinte à la vie privée ou à la sécurité des personnes.

- **Par exemple** : la publication en ligne d'organigrammes ou annuaires d'agents publics ne devraient pas comporter les données se rapportant à ceux ayant un lien immédiat avec le public et dont la révélation de l'identité serait de nature à menacer leur intégrité physique, ou n'exerçant pas de responsabilités particulières.
- **Attention** : la publication des documents administratifs détenus par les services publics d'archives doit intervenir dans les conditions prévues par le CRPA (point 9 de l'article D. 312-1-3) et faire l'objet d'une attention particulière dès lors que peuvent être concernées des données sensibles ; en particulier, toute publication anticipée par rapport à ce que le texte prévoit est soumise à autorisation préalable de la CNIL (v. pour ex. sa [délibération n° 2020-045 du 23 avril 2020](#)).

Pour garantir le caractère à la fois pertinent et non excessif des données mises à disposition, il convient de s'intéresser, sans pouvoir prétendre à l'exhaustivité, aux finalités de réutilisations susceptibles d'être poursuivies par ceux qui les exploiteront. Les organismes publics et privés peuvent ainsi opportunément associer à leur travail de sélection les réutilisateurs potentiels de ces données (p. ex. : acteurs de la recherche scientifique, de l'élaboration des politiques publiques, startups).

- **Limiter, autant que possible, le caractère identifiant des données diffusées, en recourant notamment aux techniques de pseudonymisation**

La pseudonymisation consiste à traiter les données de telle façon que celles-ci ne puissent plus directement être attribuées à une personne, en protégeant les informations permettant d'établir indirectement ce lien. Le plus souvent, il s'agit de remplacer les informations les plus identifiantes d'un jeu de données (noms et prénoms, etc.) par des données indirectement identifiantes (alias, numéro séquentiel, etc.). Elle permet ainsi de **mettre à disposition de tiers de nombreuses données à une maille individuelle, tout en limitant le risque d'identification des individus concernés**, et donc d'une exposition de leur vie privée (pour en savoir plus : [Recherche scientifique \(hors santé\) : enjeux et avantages de l'anonymisation et de la pseudonymisation](#)).

À noter que **la diffusion d'un jeu de données pseudonymisées, comme sa réutilisation, reste soumise au respect du cadre juridique de la protection des données personnelles**. En effet, contrairement à l'anonymisation, la pseudonymisation est une opération réversible : il est possible de retrouver l'identité d'une personne si l'on dispose d'informations supplémentaires (accès à la table de correspondance ou recoupement des données diffusées avec d'autres données issues de sources publiques ou privées).

## **Vous êtes un réutilisateur de données publiquement accessibles**

*Lorsque cela est possible, ne collecter parmi les données en accès libre que celles qui sont indispensables à la réalisation de l'objectif poursuivi*

Pour ce faire, les réutilisateurs de données doivent d'abord **mener une réflexion sur le champ des données « strictement nécessaires »**. Ainsi par exemple, dans le domaine « RH », les données utilisées par les recruteurs à des fins de constitution de profils de candidats doivent se limiter à permettre l'appréciation de leurs aptitudes professionnelles et capacité à occuper le poste proposé (voir à ce sujet la fiche n°14 du guide « Recrutement » de la CNIL).

**Il leur faut également réfléchir aux modalités de leur collecte** : lorsque celles-ci ne sont pas « manuelles », elles doivent être propres à garantir que le recueil d'informations n'excèdera pas en pratique le champ préalablement défini.

**Lorsqu'il est impossible de calibrer précisément en amont le champ des données collectées** (p. ex. : téléchargement d'une base de données en open data comportant plus de données que celles strictement

nécessaires à la satisfaction du besoin du réutilisateur), **un tri doit être opéré le plus rapidement possible** entre les informations dont la conservation peut être justifiée et celles devant être supprimées.

- **Privilégier [l'utilisation d'API](#) toutes les fois où de tels outils sont mis à disposition par les diffuseurs et de nature à répondre aux besoins**

En plus de garantir, en principe, le fait que la collecte concernera des données se rapportant à des personnes ayant été mises en capacité de s'opposer à leur réutilisation (ou à certaines réutilisations) par des tiers, les API permettent de cibler davantage les données recueillies que le recours (par ailleurs souvent proscrit par les sites) à des outils de moissonnage des données (« *web scraping* ») mettant en œuvre des techniques d'extraction des contenus web, via des scripts ou programmes automatisés.

- **En cas de moissonnage de données**

En plus de pouvoir compromettre fortement le bon fonctionnement des sites « moissonnés » (par exemple en occasionnant des dénis de service si trop de requêtes sont adressées simultanément), et être interdits par d'autres réglementations (par exemple par des CGU qui s'appuieraient sur le droit des producteurs de bases de données), les outils utilisés sont susceptibles, en fonction de la façon dont ils sont développés, de collecter toutes sortes de données. Il est donc nécessaire de veiller à :

- **définir, en amont de la mise en œuvre du traitement, des critères précis et pertinents de collecte** afin d'opérer une distinction quant à la nature des données collectées ;
- **limiter le risque de recueil de données sensibles**, lors de la sélection des mots clés ou catégories de contenus, par exemple ;
- **supprimer toute donnée non pertinente, quel que soit son niveau de sensibilité, dès qu'elle est identifiée comme telle** dans le cadre de l'exploitation des données ainsi collectées (un tri exhaustif n'étant pas toujours possible de manière automatisée).

*Anonymiser ou pseudonymiser les données quand leur exploitation sous une forme (directement) identifiante n'est pas requise*

Toutes les fois où un [traitement d'anonymisation ou de pseudonymisation](#) des données collectées en ligne peut être mis en œuvre, il convient d'y recourir.

De tels traitements, qui excluent ou limitent le risque d'identification des personnes concernées, **doivent être envisagés aussi bien au stade de la collecte des données qu'au stade de leur exploitation**, notamment par les acteurs de la recherche scientifique. Ceux-ci peuvent utilement consulter la [fiche pratique de la CNIL présentant les enjeux et avantages de ces techniques](#), ou encore sa [délibération n° 2018-151](#) relative à un projet de traitement ayant pour finalité une recherche sur les impacts pour la vie privée des publications d'informations librement accessibles sur les réseaux sociaux.

À noter : à l'inverse et logiquement, lorsque de telles mesures protectrices ont été adoptées par le diffuseur des données, il convient de ne pas mettre en œuvre de traitements ayant pour objet ou effet de permettre la réidentification des personnes concernées.

*Ne pas traiter de données « sensibles », sauf exception valablement mobilisable*

Les réutilisateurs de données publiquement accessibles doivent se montrer très vigilants à l'égard des informations faisant l'objet d'une protection renforcée, telles que les données dites « sensibles ». Il s'agit en particulier de celles qui révèlent **la prétendue origine raciale ou ethnique, les convictions politiques, religieuses ou philosophiques ou l'appartenance syndicale, et des données concernant la santé, la vie ou l'orientation sexuelle**.

En effet, parce que leur collecte et exploitation soulèvent par nature des risques élevés pour les droits et libertés des personnes concernées, le RGPD ([art. 9](#)) et la loi Informatique et Libertés ([art. 6](#) et [44](#)) posent un **principe d'interdiction** de traitement de ces données. Ce principe s'accompagne toutefois d'exceptions, limitativement énumérées et d'interprétation stricte.

Ainsi, **si différents projets peuvent légitimement impliquer le traitement de données de cette nature** (p. ex. : mise en œuvre d'opérations ciblées de communication politique ou de lobbying, constitution

d'une banque d'images/vidéos à des fins de recherche scientifique, réalisation d'une étude sur le comportement des personnalités politiques en matière de protection de l'intimité de leur vie privée), **que ce soit au stade de leur collecte en ligne ou par recoupements ultérieurs** (les données sensibles seront alors « inférées »), **leurs responsables doivent vérifier et être en capacité de démontrer, en particulier dans leur [analyse d'impact relative à la protection des données \(AIPD\)](#), qu'ils peuvent bien invoquer l'une ou l'autre des exceptions prévues par les textes**, parmi lesquelles celles exposées ci-dessous.

- **L'exception « recueil du consentement »**

Cinq critères cumulatifs doivent être remplis pour qu'un réutilisateur puisse s'en prévaloir : [le consentement doit être libre, spécifique, éclairé, univoque et explicite](#). Il doit faire preuve de vigilance quant à ces conditions de validité et documenter les conditions de recueil de l'accord des personnes concernées (p. ex. : mise en place de mécanismes permettant d'obtenir une déclaration d'acceptation expresse et dédiée à cette question).

- **L'exception « données manifestement rendues publiques »**

Dans ses lignes directrices sur le ciblage des utilisateurs des réseaux sociaux, le Comité européen de la protection des données fournit une liste d'éléments à considérer pour évaluer, suivant la méthode du « faisceau d'indices », si des données particulières relèvent bien de cette catégorie : le paramétrage par défaut de la plateforme de médias sociaux (la personne concernée a-t-elle fait une action pour changer ses paramètres privés par défaut en paramètres publics ?), la nature de la plateforme (a-t-elle vocation à connecter des connaissances proches avec la personne concernée, à nouer des relations intimes ?), l'accessibilité de la page concernée (la création d'un compte est-elle préalablement nécessaire ?), la visibilité de l'information sur leur caractère public (p. ex. : présence continue d'une bannière sur la page, bouton de publication informant que les informations seront rendues publiques) et le point de savoir si la personne concernée a elle-même publié les données ou si, à l'inverse, elles l'ont été par un tiers ou déduites par le réutilisateur.

**À noter : le fait que les données aient été manifestement rendues publiques par les personnes n'exonère pas les réutilisateurs de leur obligation de pouvoir justifier d'une base légale (voir la fiche dédiée à cette question).**

- **L'exception de « nécessité pour la recherche publique »**

Pour que des chercheurs puissent valablement se prévaloir de cette exception, ils doivent être en capacité de justifier, d'une part que la recherche en cause respecte certains critères [précisés dans le code de la recherche](#) (ce qui exclut la recherche privée) ; d'autre part, que ces utilisations sont rendues nécessaires « pour des motifs d'intérêt public important », après avis motivé et publié de la CNIL qu'il leur appartient donc de solliciter (pour plus de précisions sur le traitement des données sensibles dans le cadre de la recherche scientifique, voir [la fiche cas d'usage dédiée à ce sujet](#)).

**À noter : les réutilisateurs de données doivent également être vigilants à l'égard des informations qui, sans relever de l'article 9 du RGPD, présentent une sensibilité certaine, comme :**

- **les données relatives à des infractions et condamnations**, qui ne peuvent être légalement traitées que par certaines catégories d'acteurs, et en particulier les réutilisateurs des décisions de justice pseudonymisées sous réserve que leurs traitements n'aient ni pour objet ni pour effet de permettre la réidentification des personnes concernées (art. 46 de la loi Informatique et Libertés) ;
- **ou encore celles relatives à des mineurs**, massivement présents en ligne et auxquels le cadre juridique de la protection des données accorde une attention particulière, s'agissant notamment des conditions de prise en compte de [leurs droits](#).

## Fiche principe n°6 : Comment garantir l'exactitude, la sécurité et la conservation limitée des données ?

### Les principes à prendre en compte

Le RGPD impose que les données personnelles traitées soient « *exactes et, si nécessaire, tenues à jour* » (article 5.1.d).

Les diffuseurs et réutilisateurs de données doivent donc tenir compte de ce principe d'exactitude des données en veillant, tout au long de la vie de leur traitement, à ce que les informations qu'ils exploitent et qui sont erronées ou incomplètes soient effacées, rectifiées ou complétées sans tarder.

Le RGPD prévoit également que les données sont :

- **conservées sous une forme identifiable pour la seule durée nécessaire** à la satisfaction de l'objectif poursuivi par leur traitement ;
- **traitées dans des conditions de sécurité adaptées au niveau de risque** que le traitement présente pour les droits et libertés des personnes concernées.

Sans préjudice d'éventuelles obligations légales d'archivage, il convient ainsi de supprimer les données ou de les [anonymiser](#) dès qu'elles ont « rempli leur office » et, d'ici là, de les sécuriser autant que nécessaire pour garantir à la fois leur intégrité, confidentialité et disponibilité.

**Conformément aux principes chapeaux de « responsabilité du responsable du traitement » et de « protection des données dès la conception et par défaut » (art. 24 et 25 du RGPD)**, les organismes diffuseurs et réutilisateurs de données « ouvertes » doivent prendre en compte les principes précédemment évoqués :

- **le plus en amont possible, c'est-à-dire dès les premières réflexions quant aux moyens de mise en œuvre de leurs projets de traitements ;**
- **au travers d'une identification des mesures techniques et organisationnelles à adopter pour garantir et démontrer en pratique leur bonne application.**

Ces mesures sont à déterminer au regard du contexte, de la portée et de la finalité de ces traitements, comme des risques qui leur sont associés pour les droits et libertés des personnes concernées.

La réalisation d'une [analyse d'impact relative à la protection des données \(AIPD\)](#), qui est obligatoire pour les traitements à risques élevés, a vocation à guider les responsables de traitements dans l'identification de ces mesures.

### En pratique : quelles mesures adopter pour les respecter ?

#### Vous êtes un diffuseur de données

*Pour garantir l'exactitude des données diffusées*

Pour garantir le respect du principe d'exactitude, les diffuseurs de données devraient :

- Avoir veillé, préalablement à leur ouverture, à ce que les données aient fait l'objet de toutes les corrections nécessaires

**Bonne pratique :** dans le cas où le diffuseur des données dispose des coordonnées des personnes concernées, attirer leur attention, juste avant l'opération d'ouverture, sur l'existence de leurs droits d'accès et de rectification.

- Faciliter l'exercice à tout moment, par les personnes concernées, de leur droit de rectification

**Par exemple :** mise à disposition d'un formulaire de contact sur l'espace en ligne où les données sont diffusées, leur permettant de solliciter à tout moment la modification de leurs données, que ce soit pour les voir corrigées, mises à jour ou complétées.

- Dater les données diffusées (notamment via la fourniture de métadonnées)

Ce point est d'autant plus important pour les données publiques diffusées par les administrations que le CRPA oblige les réutilisateurs de ces données à mentionner la date de leur dernière mise à jour.

Par ailleurs, lorsque c'est pertinent au regard de la nature de la base de données mise à disposition et de l'activité du diffuseur, les diffuseurs pourront opportunément :

- Adopter des modalités techniques permettant de les mettre à jour régulièrement et de façon automatisée

**Par exemple :** les interfaces de programmation applicative (ou API, pour *Application Programming Interface*) permettent de rendre disponibles des données sous un format électronique facilement utilisable, tout en facilitant l'actualisation ultérieure des données : toute mise à jour réalisée sur les données sources (notamment pour prendre en compte une demande de rectification d'une personne) est automatiquement prise en compte et répercutée sur les données rendues accessibles.

Une telle recommandation ne trouve pas à s'appliquer dans certains cas, notamment lorsque les données contenues dans la base mise à disposition n'ont pas vocation à être mises à jour (p. ex. : bases de données diffusées par les services publics d'archives).

- Tracer les modifications opérées sur les données et en informer l'ensemble des destinataires, si possible et sauf efforts disproportionnés

Une telle disposition nécessite cependant de garder la trace de tous les utilisateurs, notamment personnes physiques, ayant accédé à la base de données, ce qui n'est pas toujours pertinent.

#### *Pour garantir la sécurité des données diffusées*

L'ouverture des données au public implique par définition une absence de confidentialité des données concernées. Pour autant, le diffuseur devrait **prendre les mesures nécessaires pour réduire les risques pour les personnes concernées**. Par ailleurs, la sécurité des données ne s'arrêtant pas à la préservation de leur confidentialité, des **mesures de protection de leur disponibilité et intégrité** doivent également être prises.

- Prévoir des modalités d'accès aux données adaptées à leur sensibilité

Dans le cadre de données mises à disposition du public, la vraisemblance des risques dépend surtout de la nature des données, mais aussi en grande partie de la capacité des tiers à exploiter ou à inférer des informations sur des individus pour les détourner de leur finalité initiale et/ou en en faire un traitement illégitime. Ainsi, comme déjà souligné dans le « guide CADA-CNIL », les diffuseurs de données sont invités à adopter des mesures de protection limitant la surface d'exposition des données personnelles. Il pourra s'agir, par exemple, de :

- l'utilisation à destination des moteurs de recherche externes de **règles d'indexation** (telles qu'un fichier « robots.txt ») **proscrivant celle portant sur des données identifiantes** ;
- l'adoption d'un **dispositif empêchant des moisonnages massifs de données non autorisés**, tel que la limitation du nombre de données auxquelles accède un même utilisateur ou la mise en place d'un système d'exclusion des robots (« captcha ») ;
- la mise en place d'une **authentification obligatoire pour l'accès à certaines données**, ou une **configuration des systèmes de recherche sur le site de consultation** destinée à limiter leur caractère potentiellement intrusif (p. ex. : exclusion des recherches en plein texte ou des requêtes touchant des données sensibles ; obligation de renseigner dans le moteur de recherche au minimum deux données telles que le nom de la personne et un numéro de référence du document, telle qu'un nom et un prénom, ou une date, un lieu de naissance ou de provenance).

À noter que les administrations devront toutefois veiller à ce que les mesures adoptées ne portent pas atteinte, lorsqu'il est applicable, au principe de publication en ligne des données publiques dans un format ouvert, aisément réutilisable et exploitable par un système de traitement automatisé ([arrêt CE, 27/09/2022, n° 450739](#)).

- La question de [l'utilisation d'une API](#)

Se pose la **question de savoir si la base de données mise à disposition du public doit être directement accessible** (notamment par téléchargement du fichier dans un format ouvert) **ou accessible uniquement via une API**. La solution à privilégier dépend du contexte.

- **Lorsque l'utilisation d'une API n'est pas justifiée par des besoins techniques (comme c'est le cas lorsque les données sont très régulièrement mises à jour, par exemple), il peut être préférable de donner directement accès aux données par téléchargement.** Cette solution ne permet pas de calibrer l'accès aux données en fonction des finalités de la réutilisation, ni de disposer des coordonnées des réutilisateurs pour les informer proactivement d'une correction des données ou d'une modification des droits exercées par les personnes concernées sur leurs données ;
- **Lorsque les données présentent une sensibilité particulière, il peut être nécessaire de recourir à une API pour organiser leur partage avec des tiers.** En effet, celles-ci permettent une meilleure supervision de ce dernier, d'une part en contrôlant les accès, la granularité des données accédées et, le cas échéant, les finalités d'utilisation des données ; d'autre part, en permettant la transmission sécurisée d'informations associées à l'échange de données (durée de conservation, gestion de l'exercice des droits, etc.).

- Vérifier la robustesse de l'anonymisation/pseudonymisation, le cas échéant, et assurer la sécurité des données qui ne sont pas diffusées

Afin de garantir la confidentialité des données qui ne seront pas partagées (données permettant l'identification des personnes), il importe que les diffuseurs veillent au **cloisonnement logique ou physique** entre celles-ci et les données ouvertes, prennent des **mesures techniques** (chiffrement, journalisation, limitation des accès, etc.) **et organisationnelles** (habilitations, authentification et sensibilisation du personnel, etc.) adaptées et **proposent un point de contact aux réutilisateurs** pour leur permettre de signaler tout risque sur la sécurité des données.

- Veiller à l'intégrité et la disponibilité des données partagées

Lorsqu'un diffuseur partage des données, les réutilisations prévues dépendent souvent de la disponibilité et de l'intégrité des données : s'il y a une défaillance sur l'un ou l'autre de ces aspects, cela peut avoir des conséquences en aval sur les services proposés par les réutilisateurs et, corrélativement, pour les personnes y recourant.

Afin de garantir l'intégrité des données mises à disposition de tiers, il importe ainsi que leurs diffuseurs adoptent des **mesures adaptées de traçabilité** (journalisation des actions d'accès, de modification et de suppression, analyse des traces, etc.) **et de vérification de l'intégrité** (signature des données, sécurité des systèmes physiques, etc.). Afin de garantir la disponibilité des données, ils devraient rédiger et tester fréquemment un **plan de reprise et de continuité d'activité informatique**, et prendre des mesures techniques et organisationnelles adaptées (redondance des sauvegardes, dispositif alternatif sécurisé en cas de défaillance du dispositif principal, etc.).

- S'assurer que les données ne sont pas diffusées pour une durée excessive

La limitation de la durée d'exposition **réduit la vraisemblance des risques** associés à une utilisation illégitime des données. Ainsi par exemple, lorsqu'une administration procède à la suppression ou à l'archivage définitif d'un document administratif, elle devrait simultanément mettre un terme à la publication dudit document.

À noter que le délai de diffusion est dans certains cas prévus par les textes et il convient alors de s'y conformer. Par exemple, en application du décret n°2013-1212 traitant de la diffusion des déclarations

de situation patrimoniale et déclarations d'intérêts adressées à la Haute Autorité pour la transparence de la vie publique, ces déclarations « *demeurent accessibles au public pendant la durée des fonctions ou du mandat au titre desquels elles ont été déposées. Toutefois, lorsque la déclaration est déposée après la fin des fonctions, les éléments demeurent accessibles six mois après la fin des fonctions* »).

À l'inverse, certaines dispositions légales prévoyant l'ouverture de données conduisent *de facto* à une mise en ligne sans limitation de durée (p. ex. : décisions de justice pseudonymisées diffusées dans un format ouvert en application de la loi n°2016-1321 pour une République numérique) : il est d'autant plus important, dans ces cas, de prendre des mesures de précaution sur les conditions de la mise en ligne et d'informer le mieux possible les personnes concernées.

- Communiquer sur les conditions de réutilisation des données auprès des réutilisateurs

Les diffuseurs de données personnelles ne sont en principe pas responsables des conditions dans lesquelles ces données seront en pratique réutilisées. Pour autant, ils devraient en amont **faire œuvre de pédagogie auprès des potentiels réutilisateurs, en soulignant la présence de données personnelles dans les fichiers ou flux mis à disposition et en rappelant les obligations « informatique et libertés » liées à leur exploitation.**

Les administrations recourant à la « [licence ouverte](#) » de réutilisation d'informations publiques (celle de référence établie par le Gouvernement), qui se limite à pointer l'éventuelle nécessité de respecter le cadre juridique de la protection des données personnelles, **peuvent ainsi utilement compléter son contenu de précisions sur le sujet.** Par exemple, sur la plateforme « data.gouv.fr » où elle diffuse son [jeu de données sur les désignations de DPO](#), la CNIL mentionne « l'avertissement » suivant :

« *Toute réutilisation de données publiées qui auraient la nature de données personnelles (numéro de téléphone, adresse de courrier électronique, etc.) suppose préalablement, de la part du réutilisateur, la vérification du complet respect de ses obligations prévues par le RGPD, notamment en termes d'information des délégué(e)s concerné(e)s et de respect de leurs autres droits définis par le règlement européen. À défaut, le réutilisateur s'exposerait notamment aux sanctions prévues par le RGPD* ».

**Un renvoi au présent guide, qui évoque de façon détaillée les conditions de réutilisation des données publiquement accessibles, peut désormais être effectué.**

Enfin, pour limiter les risques de réutilisation des données à des fins illégitimes, les diffuseurs de données peuvent également **apporter, par voie contractuelle** (conditions générales d'utilisation, licences de réutilisation *ad hoc* établies dans le respect des conditions prévues à l'[article D. 323-2-2](#) du CRPA, etc.), **des restrictions pertinentes à leur retraitement**, telles que : l'interdiction de réidentifier les personnes concernées, d'inférer des données sensibles, ou encore de les réutiliser à certaines fins (par exemple, à des fins de prospection commerciales ou à des fins autres que la recherche scientifique).

Cela sera d'ailleurs obligatoire dans certains cas en vertu de textes : par exemple, l'article R. 112 A-3 du livre des procédures fiscales énonce ainsi que les conditions d'utilisation du fichier des mutations immobilières / [base « DVF »](#), diffusé en open data par la DGFIP, « *prévoient, d'une part, que les traitements portant sur la réutilisation des informations mentionnées à l'article R\*. 112 A-1 ne peuvent avoir ni pour objet ni pour effet de permettre la réidentification des personnes concernées et, d'autre part, que ces informations ne peuvent faire l'objet d'une indexation sur les moteurs de recherche en ligne* ».

## **Vous êtes un réutilisateur de données publiquement accessibles**

*Pour garantir l'exactitude des données traitées*

Les réutilisateurs doivent prendre toutes les mesures raisonnables pour ne pas traiter des données obsolètes au regard des finalités poursuivies par leur traitement. Il s'agit notamment, en cas d'exploitation de fichiers téléchargeables mis à disposition du public par les administrations sur « data.gouv.fr » ou d'autres espaces de diffusion en ligne, de tenir compte du rythme d'actualisation de ces fichiers tel que généralement mentionné, ou encore de mettre à jour régulièrement les informations « privées » collectées sur des plateformes en ligne, sachant que les enjeux pour les personnes concernées peuvent être très importants. Par exemple, l'absence de mise à jour des données collectées peut conduire à la rediffusion, par différentes sociétés réutilisatrices,

d'informations que les personnes concernées avaient choisi de ne plus diffuser sur leurs profils publics de réseaux sociaux (voir à ce sujet les délibérations de la CNIL [n° 2011-203](#) et [n° 2012-156](#)).

- **Recourir quand c'est possible à des API**, en ce qu'elles permettent de rafraîchir les données de manière régulière et automatisée, comme le souligne la recommandation de la CNIL n° [2023-050](#) dédiée à ces dispositifs.
- **En cas d'utilisation de données de réseaux sociaux/blogs, en particulier à des fins de rediffusion de celles-ci, veiller au caractère très resserré des délais de mise à jour**, compte tenu du fait que les préjudices d'image et de réputation sont susceptibles d'être d'autant plus grands que les informations et images en cause évoluent par nature très rapidement.

*Pour garantir leur usage dans un temps limitée et des conditions sécurisées*

Comme n'importe quel responsable de traitement, les réutilisateurs de données publiquement accessibles doivent veiller au respect des principes de sécurité et de durée limitée de conservation des données.

Notamment parce que la sécurité des données ne s'arrête pas à la préservation de leur confidentialité (l'article L322-1 du CRPA soumet ainsi la réutilisation des informations publiques « à la condition que ces dernières ne soient pas altérées » et que « leur sens ne soit pas dénaturé »), le caractère public des données n'a pas d'incidence sur les conditions de mise en œuvre de ces principes.

**Dans ce contexte, les réutilisateurs de données publiquement accessibles sont invités à se reporter aux dossiers thématiques disponibles sur le site de la CNIL.** Ils y trouveront de nombreuses informations utiles quant à [la méthodologie à employer](#) et aux [mesures techniques et organisationnelles à adopter](#).

## Fiche cas d'usage n°1 : la diffusion et la réutilisation des données d'entreprises déclarées à l'administration

**Les données « d'entreprises »** regroupent toutes les informations relatives aux personnes morales de droit privé, par exemple les dénominations, formes juridiques, chiffre d'affaire et coordonnées des sociétés, ainsi que les informations relatives aux individus qui y sont rattachés, comme les mandataires sociaux des entreprises.

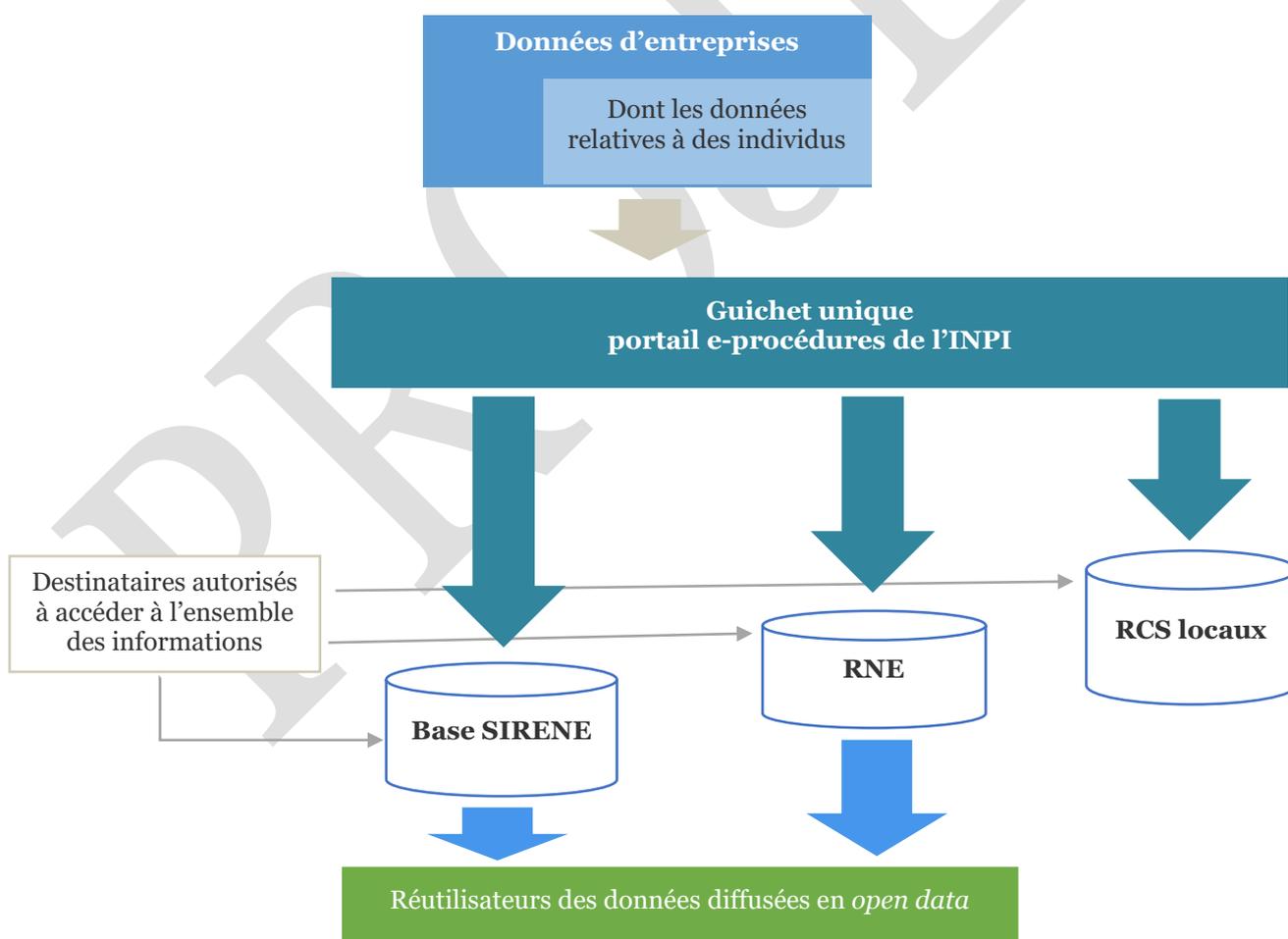
**Le RGPD ne s'applique pas aux noms, formes juridiques et coordonnées des sociétés, même lorsqu'elles incluent le nom d'une personne physique** (p. ex. : « Garage Michel Durand » SAS).

**Il s'applique en revanche à toutes les informations qui se rapportent à un individu identifié ou identifiable**, même lorsqu'elles sont traitées à titre professionnel ou qu'elles ont été diffusées en open data<sup>6</sup>. Cela inclut notamment : les noms, prénoms et adresses des personnes physiques lorsque celles-ci sont des autoentrepreneurs (entreprises individuelles), bénéficiaires effectifs ou des représentants légaux, y compris lorsque ces derniers ont le statut d'autoentrepreneur ou lorsqu'elles exercent une profession libérale sans être membre d'une structure d'exercice (ex : les données relatives à Michel Durand, en tant que gérant de la société « Garage Michel Durand »).

**La diffusion ou la réutilisation de bases de données qui contiennent des données à caractère personnel sont donc soumises aux conditions décrites ci-dessous.**

### La diffusion des données d'entreprises

Schéma du fonctionnement du guichet unique



<sup>6</sup> CJUE, deuxième chambre, 9 mars 2017, Camera di Commercio contre X, [C398/15](#), considérant 34

Depuis le 1<sup>er</sup> janvier 2023, la réalisation de formalités d'entreprises (constitution, modification, cessation) doit être réalisée auprès du guichet unique tenu par l'INPI. Cela implique la transmission de documents relatifs à la société, mais aussi à ses mandataires sociaux et ses bénéficiaires effectifs. Ces données d'entreprises sont amenées à figurer dans différents répertoires publics prévus par le droit français et européen. Ces textes autorisent l'accès à ces registres par certains destinataires institutionnels (comme l'administration fiscale) mais aussi la diffusion d'une partie de ces données en open data.

## L'encadrement juridique de la collecte et de la diffusion des données d'entreprises

En France, ces données d'entreprises sont essentiellement collectées et traitées par l'INPI et l'INSEE au titre de leurs obligations légales pour constituer deux registres, respectivement le **registre national des entreprises** (RNE) et la base communément appelée **SIRENE** (pour système national d'identification et du répertoire des entreprises et de leurs établissements).

Les données d'entreprises figurant dans ces registres sont ensuite diffusées en open data, c'est-à-dire publiées à des fins de réutilisation par des tiers par différents moyens, notamment via des interfaces de programmation (appelées [API](#)).

**Attention : lorsque les données d'entreprises comportent des données personnelles, elles ne doivent pas toutes être publiées par les administrations.**

Le code des relations entre le public et l'administration (CRPA) prévoit que les données personnelles contenues dans les documents administratifs doivent en principe faire l'objet d'une anonymisation avant leur diffusion en open data (article L. 312-1-2 alinéa 2 du CRPA), sauf disposition contraire.

Par exception, des règles spécifiques prévoient l'ouverture d'une partie des données d'entreprises, y compris lorsqu'elles contiennent des données personnelles. Ces dispositions distinguent les données pouvant être diffusées, et celles qui n'ont pas à être accessibles au public (l'accès étant réservé à des organismes dont la mission le rend nécessaire, par exemple pour des contrôles ou des enquêtes).

En application de la réglementation, seules les données suivantes sont diffusées en open data :

Données personnelles du RNE diffusées par l'INPI (article L. 123-52 du code de commerce)	Données personnelles de la base SIRENE diffusées par l'INSEE (article R. 123-232 du code de commerce)
<p>La diffusion des informations relatives à l'identité et au domicile des personnes physiques mentionnées dans le registre est limitée aux :</p> <ul style="list-style-type: none"> <li>• nom, nom d'usage, pseudonyme, prénoms,</li> <li>• mois, année de naissance, et</li> <li>• commune de résidence.</li> </ul> <p>Les documents comptables couverts par une déclaration de confidentialité ne sont pas non plus diffusés en open data.</p> <p>La diffusion des pièces annexées au Registre National des Entreprises, pouvant contenir des données à caractère personnel relatives aux dirigeants et associés des sociétés. Ces dernières peuvent différer selon le type de société et la formalité effectuée, on peut identifier par exemple :</p> <ul style="list-style-type: none"> <li>• les statuts en vigueur au jour du dépôt, paraphé, daté et signé par les associés ;</li> <li>• les procès-verbaux de décisions prises en assemblée générale.</li> </ul>	<p>Outre les données relatives à la société et son activité (ex : catégorie juridique, natures des activités exercées, date de la cessation d'activité, etc.), les données des auto-entrepreneurs personnes physiques suivantes sont mises à disposition du public :</p> <ul style="list-style-type: none"> <li>• le nom ainsi, le nom d'usage et le pseudonyme, les prénoms,</li> <li>• l'adresse du siège ou d'un établissement.</li> </ul> <p>Cela exclut donc la diffusion des autres renseignements de la base SIRENE, à savoir :</p> <ul style="list-style-type: none"> <li>• les renseignements portant sur les représentants légaux de personnes morales,</li> <li>• les nationalité, sexe, date et lieu de naissance, date de décès, adresse électronique et numéro de téléphone des personnes physiques,</li> <li>• et de ceux indiquant que l'adresse de la société correspondent au domicile personnel de l'autoentrepreneur ou du dirigeant.</li> </ul>

**Les autres données personnelles inscrites dans ces bases ne peuvent être publiées, y compris lorsqu'elles sont contenues dans les pièces justificatives** fournies à l'appui des demandes d'inscription ou de dépôt versées aux dossiers des registres (comme les pièces d'identité).

Les administrations doivent impérativement effectuer un tri entre les données et pièces justificatives transférées aux organismes ayant le droit à en connaître et celles qui peuvent être diffusées auprès du public.

**À noter :** Le RNE et la base SIRENE sont les principales bases de données d'entreprises diffusées en open data, mais elles ne sont pas les seules. D'autres bases similaires sont publiquement accessibles et susceptibles de contenir des données personnelles. Il s'agit notamment des annuaires d'ordre professionnels (en particulier des professionnels de la santé ou des avocats), ou encore de la base ICPE (installations classées protection de l'environnement) qui font l'objet d'encadrements spécifiques.

## L'information et les droits des personnes sur les données figurant dans ces répertoires

Comme tout traitement de données personnelles, les personnes concernées doivent être informées par les administrations concernées de la diffusion de leurs données en open data conformément aux exigences posées par l'article 13 du RGPD en cas de collecte directe et par l'article 14 du RGPD en cas de collecte indirecte.

Pour qu'elle soit la plus lisible et pertinente possible, l'information peut être délivrée en plusieurs niveaux. Cela consiste à prioriser les informations essentielles que sont l'identité du responsable du traitement, la finalité de mise à disposition du public à des fins de consultation ou de réutilisation, et les droits des personnes.

Un premier niveau d'information directement sur le formulaire de collecte, renvoyant vers une notice d'information complète pour les autres éléments d'information, est un bon moyen de fournir une information concise, compréhensible et aisément accessible.

La bonne information des personnes concernées est essentielle pour leur permettre de comprendre leurs droits et comment les exercer. Ces droits portent d'une part sur les traitements de diffusion par les administrations, et d'autre part sur les traitements des réutilisateurs de ces données (dans les conditions détaillées ci-dessous).

Attention : les personnes concernées ne peuvent pas s'opposer à **la diffusion initiale des données d'entreprises par l'INPI ou l'INSEE puisqu'il s'agit d'une obligation légale.**

**En revanche, les textes applicables permettent aux personnes concernées de s'opposer, lors de la réalisation de leurs formalités, à toute réutilisation de leurs données à des fins de prospection commerciale.** Cette opposition peut être formulée directement auprès de l'administration qui doit alors en informer les réutilisateurs potentiels. Il appartient ensuite à chaque réutilisateur de prendre en compte cette opposition.

**Les personnes souhaitant s'opposer à ces réutilisations n'ont pas à se justifier, et peuvent le faire à tout moment,** c'est-à-dire soit directement lors des formalités de création ou de modification de l'entreprise, soit par courriel ou par le biais d'un formulaire accessibles sur le site web de l'administration.

Droits des personnes concernées	Auprès de l'administration qui diffuse	Auprès d'un réutilisateur
Accès	✓	✓
Rectification	✓	✓
Limitation du traitement	✓	✓
Opposition à la diffusion	✗	✓
Opposition à la réutilisation commerciale	✓	✓
Opposition aux autres réutilisations	✗	✓

De manière plus générale, en tant que responsable du traitement de la publication des données personnelles, **l'administration doit faciliter l'exercice des droits des personnes concernées.** La CNIL a formulé différentes recommandations s'agissant du recours à des API, particulièrement pertinentes en la matière. Un

autre exemple de bonne pratique serait de proposer un point de contact unique aux personnes concernées pour leur permettre d'exercer leurs droits vis-à-vis de ces différentes bases, en un même endroit.

### **Exemple d'opposition à la prospection commerciale directement auprès de l'INPI**

Face au succès grandissant de ses tutoriels de pâtisserie sur les réseaux sociaux, et aux multiples demandes de ses abonnés, Madame X envisage de créer son entreprise pour vendre ses créations aux plus fins gastronomes de sa région.

Elle se connecte sur le guichet unique des entreprises et entame les démarches pour créer son autoentreprise. Dans ce cadre, elle fournit à l'administration les données la concernant, telles que son nom, prénom, ses coordonnées, l'adresse postale ainsi que les documents justificatifs (comme une pièce d'identité). Pour ne pas avoir à donner l'adresse de son domicile, elle a recours à un service de domiciliation d'entreprises.

Sur le formulaire, Madame X est informée que certaines des données qu'elle fournit à l'administration pour créer sa société seront diffusées en open data par l'INPI, et éventuellement réutilisées par des sociétés tierces.

Il est proposé à Madame X de s'opposer à ce que ses données soient utilisées à des fins de prospection commerciale<sup>7</sup>.

Elle ne peut en revanche pas s'opposer à la mise en open data de certaines de ses données qui le sont en application d'une obligation légale<sup>8</sup>.

## **La réutilisation des données d'entreprises**

### **Le principe de libre réutilisation**

Les données d'entreprises peuvent être réutilisées par toute personne qui le souhaite à d'autres fins que celles pour lesquelles elles ont été collectées par l'administration.

**Cette réutilisation d'informations publiques reste toutefois soumise au respect de la réglementation relative à la protection des données personnelles.**

### **L'exigence d'une base légale valable**

Les réutilisateurs des données d'entreprises sont considérés par la réglementation « Informatique et Libertés » comme responsables de leurs traitements de réutilisation. **Ils doivent donc mobiliser l'une des six bases légales du RGPD pour pouvoir réutiliser licitement les données publiées par l'administration**, et ne traiter que les données nécessaires pour chaque réutilisation envisagée.

Les réutilisateurs des données entreprises publiées par l'administration mobilisent généralement la base légale de l'intérêt légitime. Celui-ci devra systématiquement être mis en balance avec les intérêts, droits et libertés des personnes concernées, en tenant compte de leurs attentes raisonnables.

Le réutilisateur doit également s'assurer du principe d'exactitude des données, ce qui peut le conduire à vérifier régulièrement que les données collectées en source ouverte n'ont pas fait l'objet d'une modification. Comme la CNIL a déjà pu le souligner dans sa recommandation sur les API, cette solution technique permet aux réutilisateurs de s'assurer qu'ils ne traitent que des données pertinentes et à jour, outre la facilitation de l'exercice de leurs droits.

Pour en savoir plus : [voir la fiche « principe » n°2 sur les bases légales](#)

## **L'information et les droits des personnes sur la réutilisation de leurs données**

**Les personnes concernées doivent être informées de ces réutilisations dans les conditions prévues à l'article 14 du RGPD, et notamment de l'identité du responsable du traitement de réutilisation, de la finalité et de la base légale de cette réutilisation, des catégories de données traitées, et de la source de ces données.** Ces informations sont essentielles pour permettre aux personnes

<sup>7</sup> Conformément à l'article R123-320 du code de commerce.

<sup>8</sup> Conformément à l'article L. 123-52 du code de commerce.

concernées d'exercer leurs droits auprès des réutilisateurs, mais aussi auprès du responsable de la diffusion (par exemple en cas de diffusion de données non autorisées).

En principe, tout réutilisateur doit s'assurer qu'une information complète soit fournie à chaque personne concernée, dans des délais raisonnables et au plus tard un mois après la collecte. En cas de première communication avec la personne concernée moins d'un mois après la collecte des données, l'information doit être fournie au plus tard à cette occasion.

**Par exception, notamment lorsqu'une information individuelle se révèle impossible ou exigerait des efforts disproportionnés, une information générale pourra s'avérer suffisante (par exemple, à travers la publication de mentions d'information sur le site web du réutilisateur).**

Si les efforts d'une information individuelle ne sont généralement pas considérés comme disproportionnés lorsque le réutilisateur détient des coordonnées électroniques des personnes concernées, la dérogation pourra plus facilement être envisagée lorsqu'il n'a accès qu'aux coordonnées postales des personnes concernées ou de leur entreprise. Une analyse au cas par cas est donc nécessaire pour déterminer si la fourniture d'une information individuelle exigerait des efforts disproportionnés, en prenant notamment en compte le nombre de personnes à informer et la faiblesse ou l'importance des risques que le traitement peut faire encourir aux personnes, ainsi que le fait que les personnes s'attendent à la réutilisation de leurs données.

**Par exemple :**

- Les simples rediffusions à l'identique de tout ou partie de la base diffusée en open data par l'INPI, avec des mises à jour quotidiennes, ne présenteront pas un degré d'intrusivité exigeant la fourniture d'une information individuelle.
- À l'inverse, les rediffusions en ligne de données d'entreprises contenues dans le RNE et/ou la base SIRENE recoupées avec des données provenant d'autres sources exigeront le plus souvent une information individuelle.
  - **Pour en savoir plus : voir la fiche pratique n°2 dédiée à la réutilisation de données publiquement accessibles pour la diffusion d'annuaires enrichis de professionnels**
- L'intrusivité limitée de traitements ponctuels d'éléments isolés de la base n'exigera pas, le plus souvent, la fourniture d'une information individuelle (il en irait ainsi des réutilisations de données pour la production de notes internes à une entreprise).
- L'utilisation de ces bases de données pour fiabiliser des données d'entreprise qu'un réutilisateur a déjà légitimement obtenu par ailleurs (par exemple dans le cadre d'une relation d'affaires) ne présentent pas, en principe, un niveau d'intrusivité exigeant la fourniture d'une information individuelle.
- La collecte et l'anonymisation d'une partie de ces données afin de constituer une base de données de test (par exemple pour tester le bon fonctionnement d'API dans un environnement de préproduction), n'imposera pas non plus la fourniture d'une information individuelle.

Cette information est essentielle pour que **les personnes puissent exercer leurs droits auprès du réutilisateur. Cela comprend, leur droit d'accès qui permet notamment d'obtenir une information précise sur la source comme sur les destinataires de leurs données.**

**Cela comprend aussi en principe, leur droit d'opposition à la réutilisation de leurs données, conformément à l'article 21 du RGPD.** Ce dernier peut s'exercer à tout moment (c'est-à-dire pendant le traitement mais aussi en amont). Il doit être motivé par des raisons tenant à la situation particulière des personnes, sauf pour les prospections commerciales.

#### **Exemple d'opposition à la rediffusion de ses données d'entreprises sur un site privé**

Alors que Madame X souhaite créer un site internet où elle pourrait poster des actualités sur son activité et ses nouvelles recettes pour fidéliser ses clients et en appâter d'autres, elle s'aperçoit que plusieurs sites apparaissent lors d'une recherche effectuée avec ses nom et prénom. Elle se rend notamment sur le site d'une société privée qui diffuse des informations relatives à sa société, comprenant des données à caractère personnelles la concernant.

Si Madame X a bien conscience que les données de sa société ont été mises à disposition du public à des fins de consultation ou de réutilisation, pour des raisons personnelles, elle trouve embarrassant que ses nom et prénom ainsi que ses mois et année de naissance soient diffusés en sa qualité de dirigeant.

Comme l'indique le site en question dans sa politique de confidentialité, les données figurant au RNE ou au SIRENE peuvent être réutilisées à des fins de rediffusion sur la base de son intérêt légitime.

Néanmoins, Madame X est en droit de s'opposer au traitement de ses données personnelles auprès de l'éditeur du site, en application de l'article 21 du RGPD, en justifiant de raisons tenant à sa situation particulière. Le site devra alors effacer les données sauf s'il démontre qu'il existe des motifs impérieux et légitimes prévalant sur les droits et libertés de la personne.

Madame X peut également formuler une [demande de déréférencement](#) auprès des sociétés exploitant un moteur de recherche (Google, Bing, etc.) afin que la page concernée n'apparaisse plus dans la liste des résultats obtenus lors d'une recherche effectuée à partir de ses nom et prénom.

Outre le recours aux API par les réutilisateurs, **prévoir un mécanisme d'exercice des droits via un formulaire dédié facilement accessible est également un bon moyen pour les réutilisateurs de satisfaire leur obligation de faciliter l'exercice des droits des personnes**, en particulier en cas de rediffusion en ligne des données d'entreprises.

## Annexe à la fiche cas d'usage sur les données d'entreprises : quelques dispositions juridiques relatives à leur ouverture

<p><b>Ouverture de certaines données de la base SIRENE</b></p> <p>L'article R. 123-232 du code de commerce</p>	<p>L'INSEE peut mettre à la disposition du public, selon les modalités définies aux articles R. 321-5 à R. 321-7 du code des relations entre le public et l'administration, les renseignements énumérés aux articles R. 123-222 et R. 123-223 du présent code, à l'exception de ceux concernant les représentants légaux de personnes morales et de ceux indiquant les nationalité, sexe, date et lieu de naissance, date de décès, adresse électronique et numéro de téléphone des personnes physiques, et de ceux indiquant que l'adresse légale ou le siège de l'unité légale et l'adresse d'un établissement correspondent au domicile personnel de la personne physique ou d'un dirigeant de la personne morale ou du groupement.</p>
<p><b>Ouverture de certaines données du RNE</b></p> <p>Article L. 123-52 du code de commerce</p>	<p>L'intégralité des informations inscrites et pièces annexées au registre national des entreprises, à l'exception des documents comptables couverts par une déclaration de confidentialité, fait l'objet d'une mise à la disposition du public gratuite et sous forme électronique, à des fins de consultation ou réutilisation. La mise à disposition des informations inscrites relatives à l'identité et au domicile des personnes physiques mentionnées dans le registre est limitée aux nom, nom d'usage, pseudonyme, prénoms, mois, année de naissance et commune de résidence.</p>
<p><b>Opposition à la prospection commerciale sur la base du RNE</b></p> <p>Article R. 123-320 du code de commerce</p>	<p>Si le déclarant s'oppose à la mise à disposition de ses données à des fins de prospection en application du paragraphe 2 de l'article 21 du règlement (UE) 2016/679 du Parlement européen et du Conseil du 27 avril 2016 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, son opposition est portée à la connaissance des administrations et du public.</p>
<p><b>Opposition à la prospection commerciale sur la base SIRENE</b></p> <p>Article A123-96 du code de commerce</p>	<p>Toute personne physique peut demander soit directement lors de ses formalités de création ou de modification, soit par courriel ou téléprocédure accessibles sur le site insee.fr, que les informations du répertoire la concernant ne puissent être utilisées par des tiers autres que les organismes habilités au titre de l'article R. 123-224 ou les administrations définies au 1° de l'article L. 100-3 du code des relations entre le public et l'administration, à des fins de prospection.</p>
<p><b>Ouverture de la base SIRENE en tant que données de référence</b></p> <p>Articles R. 321-5 et R. 321-6 du CRPA</p>	<p>Sont mises à disposition, sous une forme électronique, des données de référence contenues dans le répertoire des entreprises et de leurs établissements (répertoire SIRENE), produit par l'Institut national de la statistique et des études économiques, sous réserve des exceptions prévues le code de commerce.</p>
<p><b>Ouverture du registre du commerce et des sociétés (RCS)</b></p> <p>Article L. 123-1 du code de commerce</p>	<p>Figurent au registre, pour être portés à la connaissance du public, les inscriptions et actes ou pièces déposés prévus par décret en Conseil d'Etat</p>

<p><b>Ouverture du Bulletin des annonces civiles et officielles</b></p> <p>Article R. 123-209 du code de commerce</p>	<p>Le Bulletin officiel des annonces civiles et commerciales, qui assure la publicité des actes enregistrés au registre du commerce et des sociétés, est publié sous forme électronique, dans des conditions de nature à garantir son authenticité et son accessibilité permanente et gratuite.</p>
<p><b>Ouverture du Bulletin des annonces civiles et officielles</b></p> <p>Article R. 123-156 du code de commerce</p>	<p>L'avis au Bulletin officiel des annonces civiles et commerciales contient pour les personnes physiques : les références de l'immatriculation, les nom, nom d'usage, pseudonyme et prénoms de la personne immatriculée, la ou les activités effectivement exercées, le lieu d'exercice, la date du commencement d'exploitation, le nom commercial.</p>
<p><b>Ouverture des données d'intérêt général</b></p> <p>Article L. 312-1-1 du CRPA</p>	<p>Sont diffusées les bases de données et tous documents dont la publication présente un intérêt économique, social, sanitaire ou environnemental.</p>
<p><b>Ouverture des données de référence</b></p> <p>Article L. 321-4 du CRPA</p>	<p>Sont mises à disposition des données de référence en vue de faciliter leur réutilisation constitue une mission de service public relevant de l'État. Toutes les administrations concourent à cette mission.</p>
<p><b>Ouverture sans anonymisation</b></p> <p>Article D. 312-1-3 du CRPA</p>	<p>Cette catégorie comprend notamment, les documents nécessaires à l'information du public relatifs aux conditions d'organisation de la vie économique, associative et culturelle tel que le répertoire national des associations et le répertoire des entreprises et de leurs établissements ou encore les documents nécessaires à l'information du public relatifs aux conditions d'organisation et d'exercice des professions réglementées et des activités professionnelles soumises à la réglementation, notamment celles relatives à l'exercice des professions de notaire, avocat, huissier de justice et architecte</p>
<p><b>Diffusion de l'annuaire des avocats</b></p> <p>Article 21-1 de la loi n° 71-1130 du 31 décembre 1971, modifiée par la loi du 18 novembre 2016</p>	<p>Sur la base des informations communiquées par les conseils de l'ordre en application du 1° bis de l'article 17, le Conseil national des barreaux établit, met à jour et met à disposition en ligne un annuaire national des avocats inscrits au tableau d'un barreau.</p>

## Fiche cas d'usage n°2 : la réutilisation de données publiquement accessibles aux fins de diffusion d'annuaires « enrichis » de professionnels

Cette fiche s'adresse aux organismes (associations, sociétés, etc.) qui publient en ligne, sous la forme d'**annuaires thématiques ou généralistes intégrant généralement des fiches / encarts / profils individuels**, des données professionnelles relatives à des personnes exerçant une activité commerciale, artisanale ou libérale.

Constitués de données personnelles au sens du RGPD<sup>9</sup>, ces annuaires de professionnels **se distinguent des « annuaires universels »** (listes d'abonnés aux services de téléphonie fixe et mobile) régis par la directive « e-Privacy » et des dispositions spécifiques du code des postes et communications électroniques<sup>10</sup>. Ils sont **élaborés à partir de données publiquement accessibles et, notamment, de celles figurant dans les registres officiels** obligatoirement diffusés sur Internet dans un format ouvert.

**Exemples de données publiquement accessibles :** données de la [base SIRENE des entreprises et de leurs établissements](#) et du [registre national des entreprises](#) respectivement diffusés par l'Institut national de la statistique et des études économiques (INSEE) et l'Institut national de la propriété industrielle (INPI)<sup>11</sup> ; [annuaires d'avocats publiés par le Conseil national des barreaux](#) et les ordres des avocats des différents barreaux français<sup>12</sup> ; « [Annuaire Santé](#) » diffusé par l'Agence du numérique en santé et intégrant notamment le répertoire partagé des professionnels intervenant dans le système de santé (RPPS)<sup>13</sup>.

Ces informations peuvent être **collectées par différents moyens** et, notamment, via des interfaces de programmation applicative (API) mises à disposition du public et/ou le lancement de requêtes automatisées sur les moteurs de recherche (avec indexation ou extraction de contenus). En plus de données d'identité et de contact professionnels, ces annuaires peuvent ainsi recenser des **données très diverses**, telles que la photographie des intéressés, leurs spécialités / domaines d'expertise, leurs diplômes, des références aux dossiers / sujets dont ils ont eu à connaître, à leurs travaux de recherche, des publications à leur nom ou les citant dans des magazines spécialisés, sur des réseaux sociaux, etc.

Par ailleurs, ces annuaires s'accompagnent généralement d'une **offre de services pour les professionnels référencés** (p. ex. : possibilité de « revendiquer » leur fiche et de souscrire un abonnement pour l'enrichir d'éléments les concernant et/ou d'être mis en relation avec des internautes via la plateforme). **Les personnes ayant fait appel aux professionnels référencés disposent parfois de la faculté de publier des commentaires et de noter** la qualité de leurs prestations.

Si de tels annuaires sont susceptibles d'avoir des effets bénéfiques (visibilité pour les professionnels, information des consommateurs, etc.), ils comportent à l'évidence des **risques pour les droits et libertés des personnes référencées. L'exploitation de leurs données par des annuaires, combinées à un enrichissement potentiel de celles-ci (notamment par les notes et commentaires d'internautes), peuvent emporter une atteinte à leur vie privée ou à leur réputation. Ces risques se trouvent**

<sup>9</sup> Les informations d'ordre professionnel restent des données personnelles, dont l'utilisation est soumise à la réglementation relative à la protection des données, comme souligné par la CJUE dans un arrêt du 9 mars 2017 (C-398/15). La CNIL l'avait auparavant rappelé dans sa [délibération n° 2014-041](#) du 29 janvier 2014, confirmée par l'arrêt du Conseil d'État, 10ème SSJS, 30 décembre 2015, n° 376845 : « *Le nom et les coordonnées des personnes physiques, telles que leurs adresses et leurs numéros de téléphone, constituent des informations relatives à une personne physique identifiée et, par suite, des données à caractère personnel au sens des dispositions de la loi du 6 janvier 1978. Dès lors, que ces données soient des coordonnées professionnelles des personnes physiques en cause, et qu'elles soient le cas échéant par ailleurs rendues publiques, est sans incidence à cet égard* ». A noter toutefois que le RGPD ne s'applique pas aux noms, formes juridiques et coordonnées des sociétés, même lorsqu'elles incluent le nom d'une personne physique (ex : « Garage Michel Durand » SAS).

<sup>10</sup> Articles L34 et R10 et suivants du code des postes et communications électroniques (CPCE).

<sup>11</sup> Voir à ce sujet la fiche pratique « Publication et réutilisation de données entreprises ».

<sup>12</sup> L'article 3 de la loi n° 2016-1547 du 18 novembre 2016 de modernisation de la justice du XXI<sup>e</sup> siècle prévoit que les huissiers de justice, les notaires, les commissaires-priseurs judiciaires, les avocats, les avocats au Conseil d'Etat et à la Cour de cassation, les commissaires aux comptes et les experts-comptables « *rendent librement accessibles les données figurant dans leurs annuaires et tables nationales, notamment au moyen d'un standard ouvert et réutilisable, exploitable par un traitement automatisé* » ; l'article 21-1 de la loi du 31 décembre 1971 portant réforme de certaines professions judiciaires et juridiques énonce que « *le Conseil national des barreaux établit, met à jour et met à disposition en ligne un annuaire national des avocats inscrits au tableau d'un barreau* ».

<sup>13</sup> Arrêté du 23 septembre 2022 relatif à la mise en œuvre du « Répertoire partagé des professionnels intervenant dans le système de santé » (RPPS).

**renforcés par le fait que ces données peuvent être inexactes et faire l'objet d'indexations** par différents moteurs de recherche externes.

Ces traitements de données font ainsi l'objet de **nombreuses plaintes auprès de la CNIL, en particulier de la part de médecins et d'avocats** qui, après saisie de leurs nom et prénom sur un moteur de recherche, découvrent l'existence de ces annuaires « parallèles ». Bien souvent en effet, les personnes concernées n'ont pas autorisé leur mise en œuvre, n'ont pas été informées de celle-ci et ne sont pas mises en mesure d'exercer leurs droits « informatique et libertés ».

Dans ce contexte, les développements qui suivent visent à **rappeler les principaux points de vigilance à considérer pour assurer la conformité au RGPD des traitements de données intervenant dans le cadre du référencement et, le cas échéant, de la notation des professionnels concernés**. Leur bonne prise en compte sera source de sécurité juridique et permettra le développement de relations de confiance, aussi bien avec les intéressés qu'avec les internautes exploitant le contenu de ces sites web.

## Identifier la base légale des traitements

Lorsqu'ils n'interviennent pas exclusivement en vertu d'un [contrat](#) entre l'éditeur et les professionnels concernés (ce qui est généralement le cas dès lors que ces annuaires préexistent la plupart du temps à d'éventuelles relations contractuelles), **deux bases légales sont susceptibles de fonder ces traitements : l'« intérêt légitime » de la société éditrice (art. 6.1.f du RGPD) et le « consentement » des personnes concernées (art. 6.1.a du RGPD)**.

De façon générale, la possibilité de recourir à la première est subordonnée au respect de certaines conditions : le traitement doit poursuivre un intérêt légitime de l'organisme ou d'un ou plusieurs tiers, être nécessaire à sa satisfaction et ne pas porter une atteinte excessive aux droits et intérêts des personnes concernées.

Pour savoir si la base légale de l'intérêt légitime peut valablement justifier un projet de collecte et d'exploitation de données à des fins de publication d'un annuaire en ligne, son responsable doit procéder à une mise en balance des intérêts en présence, **en appréciant tout particulièrement :**

- **les bénéfices attendus**, pour lui et les tiers, de la rediffusion de données publiquement accessibles et de leur éventuel enrichissement consécutif ;
- **les attentes raisonnables des personnes concernées à l'égard de tels traitements**, compte tenu notamment du contexte de la publication initiale de leurs données ;
- **les risques qui en résultent pour elles**, en particulier au regard de la nature des données en cause ; et
- **les garanties envisagées** pour écarter ou réduire les impacts potentiels.

**Schématiquement, deux types de traitement peuvent être distingués :**

- 1. Les annuaires constitués uniquement à partir de données diffusées en sources ouvertes (*open data*) à des fins de réutilisation : ces traitements peuvent se fonder sur la base légale de l'intérêt légitime.**

**Lorsqu'ils consistent uniquement à référencer des professionnels au sein d'un annuaire et se limitent, par défaut (c'est-à-dire sauf intervention directe de ces derniers), à rediffuser les données « élémentaires » sur leur activité** (données d'identité, spécialités / domaines d'expertise, coordonnées du lieu d'exercice de la profession, etc.) qui se trouvent publiées dans un format ouvert en vertu d'un cadre légal spécifique (voir les exemples fournis en introduction), **leur licéité n'est pas subordonnée au recueil d'un consentement préalable du professionnel concerné**.

En effet, en favorisant utilement l'accès du public aux informations administratives et pratiques relatives à des prestations de services susceptibles de l'intéresser, ils s'inscrivent dans la continuité des objectifs poursuivis par les textes prévoyant la diffusion en sources ouvertes (*open data*) des registres / répertoires officiels de professionnels, au regard de la valeur qu'ils présentent pour la « communauté ».

**Le fait qu'à la finalité informative de ces annuaires puisse s'ajouter une finalité commerciale** (p. ex. : présence de bandeaux publicitaires à destination des internautes, dispositifs d'abonnements payants permettant aux professionnels de procéder à une présentation personnalisée de leur entreprise, d'être mis en relation avec les internautes via la plateforme, etc.), **est en principe sans incidence** sur la possibilité de recourir à la base légale de l'intérêt légitime : d'un côté, les sociétés éditrices sont autorisées à poursuivre un tel intérêt dans le cadre de leur activité à but lucratif ; de l'autre, les personnes concernées, en principe préalablement informées de l'ouverture de leurs données par les administrations, pouvaient raisonnablement s'attendre au traitement en cause (principe de libre réutilisation des données publiques), et le risque d'atteinte à leurs droits et libertés est très limité compte tenu de la nature des informations traitées.

**À noter : l'éditeur de l'annuaire devra prendre soin de mentionner la source des données qu'il diffuse et la date de leur dernière mise à jour (art. L. 322-1 du code des relations entre le public et l'administration - CRPA). Conformément au principe d'exactitude des données personnelles, il devra également veiller, autant que nécessaire, à leur actualisation (art. 5.1.d du RGPD), en utilisant le cas échéant l'API fournie par le diffuseur initial des données, ou en tenant compte de la fréquence de mise à jour du fichier à disposition du public en téléchargement.**

## **2. Les annuaires qui viennent enrichir les données professionnelles en *open data* avec d'autres données publiquement accessibles ou fournissant un service de « notation » / commentaires : la possibilité de recourir à la base légale de l'intérêt légitime doit faire l'objet d'une appréciation au cas par cas.**

De tels traitements sont par nature plus intrusifs et, suivant les conditions concrètes de leur mise en œuvre, présentent davantage de risques au regard de la protection des données.

**L'analyse d'impact sur la protection des données, qui devra souvent être menée préalablement à leur mise en œuvre (AIPD, requise en cas de cumul de différents critères dégagés par le Comité européen de la protection des données : collecte à large échelle, surveillance systématique, croisement de données, profilage, évaluation / « scoring », etc.), permettra notamment de déterminer s'ils sont propres à générer un niveau de risque pour les droits et intérêts des personnes qui ne saurait être considéré comme acceptable en l'absence de leur consentement préalable.**

*S'agissant des traitements consistant à agréger des données venant de différentes sources*

En vue d'assurer une plus-value à leurs annuaires, les éditeurs peuvent vouloir y intégrer d'autres informations que celles figurant dans les registres et répertoires officiels, en particulier lorsqu'il s'agit de les enrichir d'éléments en libre accès sur Internet (p. ex. : coordonnées électroniques, travaux de recherche, analyses publiquement exprimées sur un sujet, etc.).

Cet enrichissement sans le consentement préalable des personnes concernées est possible **si les collectes et traitements de données envisagés n'emportent pas une atteinte disproportionnée à leurs droits et intérêts, compte tenu notamment de leurs attentes raisonnables**. Pour apprécier l'existence d'un tel équilibre, doivent en particulier être considérés le contexte de la publication initiale des données, la nature de celles-ci et les conditions de leur collecte / enregistrement.

Par exemple, s'il s'agit de données dont la publication résulte d'une action positive des intéressés, pouvant être considérée comme témoignant, de façon manifeste, de leur souhait de voir ces informations associées à leur « identité numérique professionnelle » (p. ex. : leur numéro de téléphone mobile figurant à leur demande dans les annuaires publics d'abonnés à la téléphonie ; les informations concernant les expertises, expériences et compétences qu'ils mentionnent sur leur site web ou dans leur profil public d'un réseau social professionnel), leur rediffusion dans un annuaire généraliste ou thématique de professionnels présente moins le risque de heurter leurs intérêts que s'il s'agissait de données les concernant publiées par des tiers. Les éditeurs y procédant pourront ainsi se fonder plus aisément sur leur intérêt légitime, sous réserve de respecter les autres droits des personnes concernées (p. ex. : protection de certains contenus par le droit d'auteur ; pas d'exploitation de la photographie d'une personne sans recueil préalable de son consentement).

**À noter :** il sera nécessaire de contrôler régulièrement, au moyen de mesures techniques et organisationnelles appropriées, que les données extraites des profils publics des réseaux sociaux professionnels n'ont pas fait l'objet de suppressions ou au contraire d'enrichissements par les personnes concernées.

D'autres traitements seront davantage de nature à créer un déséquilibre au détriment des droits et intérêts des professionnels référencés, en particulier ceux consistant à alimenter, au moyen de recoupements de données, la fiche de ces professionnels en éléments inférés ou dérivés (p. ex. : déduction de l'adresse électronique professionnelle à partir de celles de leurs collègues, établissement de corrélations entre elles et des personnes avec lesquelles elles ont collaboré). Ainsi, le consentement des personnes concernées devrait être recherché.

Enfin, certains traitements très intrusifs ne satisferont pas, à l'évidence, le test de la balance des intérêts. Ce sera par exemple le cas de l'enrichissement des fiches de l'annuaire en données sans lien direct avec l'activité professionnelle des personnes concernées, collectées notamment sur des plateformes de réseaux sociaux dédiées pour l'essentiel à des échanges d'ordre personnel. Une telle agrégation de données changerait la nature même de l'annuaire, en conduisant de fait à l'établissement de profils de *particuliers* au potentiel d'atteinte à la vie privée particulièrement élevé (voir à ce sujet [la délibération de la formation restreinte de la CNIL n° SAN-2011-203 du 21 septembre 2011](#), confirmée par [l'arrêt « Pages Jaunes Groupe » du Conseil d'État n° 353193](#)).

**À noter :** lorsque le consentement des personnes concernées est requis pour garantir la licéité d'un traitement, il devra, pour être valable, satisfaire aux conditions prévues par le RGPD, telles que rappelées dans [cette fiche pratique](#). Concernant les modalités envisageables pour son recueil, voir la fiche relative aux bases légales du présent guide.

*S'agissant des traitements intégrant la collecte et la publication de notes et commentaires d'internautes*

La CNIL considère que, **sous réserve que leurs conditions de mise en œuvre ne portent pas une atteinte disproportionnée aux droits ou intérêts des artisans, commerçants et/ou professionnels libéraux référencés, de tels traitements peuvent se fonder sur la base légale de l'intérêt légitime** dès lors qu'ils participent à la liberté d'expression et d'information de l'éditeur et des contributeurs, concernent des professionnels vis-à-vis desquels le public dispose d'une « liberté de choix » (par opposition à d'autres catégories de professionnels, tels les enseignants de l'Éducation nationale)<sup>14</sup> et constituent une pratique expressément admise par le législateur (la loi n°2016-1321 pour une République numérique lui a donné un cadre juridique spécifique).

**Il appartient toutefois à l'éditeur de l'annuaire de prendre toutes les mesures permettant de garantir l'équilibre de la balance des intérêts en cause et ainsi la licéité de son traitement.**

La CNIL souligne notamment que lorsqu'un professionnel fait, ou est susceptible de faire, l'objet de commentaires négatifs pouvant avoir de lourdes conséquences sur son activité ou sa réputation, il lui sera souvent difficile de faire valoir son point de vue de façon circonstanciée et efficiente, en particulier du fait de la difficulté à identifier l'auteur du commentaire, de la lourdeur des procédures, parfois juridictionnelles, à mettre en œuvre pour obtenir la suppression de contenus abusifs, et de la subjectivité qui s'attache à l'appréciation de certains « savoir-être » et « savoir-faire ».

**La CNIL estime qu'une attention particulière doit être apportée au traitement des demandes d'opposition des professionnels :** en principe, il doit être fait droit à la demande d'un professionnel qui demande à sortir d'un annuaire professionnel privé. **C'est en prenant en compte, notamment, cet élément, qu'on peut admettre l'intérêt légitime de l'éditeur à publier un tel annuaire commenté.**

En outre, l'éditeur pourra utilement se référer aux exigences posées par la norme *ISO 20488:2018* s'adressant aux gestionnaires d'avis en ligne de consommateurs (p. ex. : identification des auteurs d'avis et publication en premier des avis les plus récents).

En tout état de cause, il devra impérativement :

- 1. Faire preuve d'une rigueur particulière pour informer, dans les conditions prévues par le RGPD et développées au point suivant, les personnes évaluées de la mise en œuvre du**

---

<sup>14</sup> En 2008, l'illégalité d'un site web (« Note2be ») organisant la notation de professeurs a été pointée par des juridictions (TGI Paris, ord. Réf. 3 mars 2008, n° 08/51650 et CA Paris, 14ème ch., sect. A, 25 juin 2008), ainsi que par la CNIL. Celle-ci a considéré que la société éditrice ne pouvait se prévaloir de son intérêt légitime compte tenu du fait, en particulier, que le site était susceptible de créer une confusion dans l'esprit du public avec un système officiel de notation. À l'inverse, dans un arrêt du 11 mai 2017, la Cour de cassation (pourvoi n° 16-13669, CNB c/ Jurisystem) a jugé que les tiers à la profession d'avocat ne sont pas tenus par les règles déontologiques de celle-ci, visant à garantir son indépendance, sa dignité et son intégrité, et qu'il leur appartient seulement, dans leurs activités propres, de délivrer au consommateur une information loyale, claire et transparente.

**traitement, ainsi que pour garantir l'effectivité de leurs droits « informatique et libertés ».**

**À noter :** bien que l'éditeur de l'annuaire puisse n'avoir, s'agissant des commentaires d'internautes, qu'un rôle passif (absence de contrôle avant publication) et ainsi le statut d'hébergeur de contenus au sens de l'article 6 précité de la LCEN, il doit être qualifié de responsable du traitement de ces commentaires au sens de l'article 4.7 du RGPD dès lors qu'il est à l'initiative de ce traitement et détermine les conditions de mise en œuvre de ce dernier (p. ex. : critères de notation et de classement des avis, durée de conservation et de publication). Par conséquent, il lui appartient, en particulier, d'examiner les demandes des personnes concernées exerçant leurs droits d'opposition et d'effacement.

À noter que si le traitement est mis en œuvre à des fins de prospection, le droit d'opposition sera obligatoirement discrétionnaire (art. 21.3 du RGPD).

- 2. Respecter les règles spécifiques du code de la consommation encadrant une telle pratique (articles L.111-7-2 et D.111-17 et suivants), en particulier la délivrance aux utilisateurs du site d'une information loyale, claire et transparente sur les modalités de publication, modération et de traitement des avis (p. ex. : existence ou non d'une procédure de contrôle, critères de classement, date de « l'expérience de consommation » et durée de publication), ainsi que la mise en place d'une fonctionnalité gratuite permettant aux responsables des produits/services de signaler un doute sur leur authenticité ;**

**À noter :** bien qu'elle ne soit pas exigée mais seulement encadrée par le législateur, la CNIL recommande la mise en œuvre d'une procédure de contrôle des avis intégrant, dans le respect des règles de protection des données personnelles, des mesures permettant d'identifier la source de l'information et de vérifier sa fiabilité : elle responsabilisera les utilisateurs du service et participera à la légitimité de l'éditeur de l'annuaire à se prévaloir du droit à l'information des internautes, comme au caractère adéquat, pertinent et non excessif des avis publiés.

Dans cette même perspective, la CNIL recommande également la mise à disposition :

- de critères de notation objectifs et appropriés au regard du type de prestation fournie par le professionnel (tenant compte par exemple du fait que les résultats obtenus par un médecin ou avocat ne dépendent pas nécessairement de la qualité de leurs prestations) ;
  - des mentions de sensibilisation à l'intention des contributeurs, les alertant notamment sur la nécessité de ne pas faire un usage illicite ou abusif de leur liberté d'expression (p. ex. : injure, diffamation, dénigrement), ainsi que sur les risques encourus en pareil cas (sanctions civiles et pénales) ;
  - d'un dispositif de signalement des propos répréhensibles, aux fins de suppression de ceux-ci par l'éditeur ou l'hébergeur.
- 3. Enfin, la CNIL recommande que la durée de conservation et de publication des avis n'excède pas cinq ans. Une durée raisonnable devrait être déterminée au cas par cas, en considération de la catégorie de professionnels concernés (ex. : les restaurateurs sont plus souvent notés que les notaires, et la qualité de leurs prestations est davantage susceptible d'évolutions dans le temps ; ces différences justifient qu'un délai plus court soit retenu pour les premiers que pour les seconds). À l'issue de ce délai, des données agrégées (par exemple, une agrégation des notes) pourraient être conservées et publiées selon les mêmes règles, avant d'être définitivement supprimées.**
  - 4. Tenir compte des dispositions relatives au droit de réponse prévu par la loi n° 2004-575 du 21 juin 2004 pour la confiance dans l'économie numérique (LCEN – article 6.IV) et précisé par le décret n° 2007-1527 ;**

**À noter :** si l'éditeur de l'annuaire dispose des coordonnées électroniques des professionnels référencés, une bonne pratique consiste en l'information systématique des personnes concernées par la publication d'un nouvel avis.

**Attention :**

**Les textes permettent le prononcé de sanctions aussi bien administratives que judiciaires, de nature civile comme pénale,** pour non-respect des dispositions du RGPD, du code de la consommation et/ou de la LCEN, ainsi que pour les « délits de presse » prévus par la loi de 1881 (diffamation et injure, en particulier).

## Informer les personnes concernées

Comme n'importe quel responsable de traitement, les éditeurs d'annuaires de professionnels sont concernés par le **principe de transparence des traitements de données personnelles** (voir la fiche du présent guide dédiée à ce sujet) : ils doivent informer les personnes dont ils traitent les données des objectifs qu'ils poursuivent et des conditions de mise en œuvre de leurs traitements. **Une telle information permet aux personnes de conserver la maîtrise des usages qui sont faits de leurs données**, en les mettant notamment en mesure d'exercer leurs droits.

**Compte tenu des risques qu'une exploitation insuffisamment contrôlée des annuaires « enrichis » de professionnels fait peser sur les personnes (traitement de données non pertinentes, excessives ou inexactes ayant pour effet de porter atteinte à leur vie privée ou à leur réputation), les enjeux qui s'attachent au caractère effectif de leur information sont particulièrement importants.**

**Ainsi, ce n'est que dans des hypothèses très limitées que les éditeurs devraient pouvoir se prévaloir des dispositions de l'article 14.5.b du RGPD**, autorisant les organismes ne collectant pas les données directement auprès des personnes concernées à ne pas informer celles-ci lorsqu'une telle information exigerait des « efforts disproportionnés ».

### Deux situations peuvent être schématiquement distinguées.

*Lorsque l'éditeur dispose des coordonnées électroniques des personnes, il devra, par principe, procéder à une information individuelle et complète de chaque professionnel concerné*

En effet, dès lors qu'une information individuelle **peut être effectuée au moyen de l'envoi automatisé de courriels à chacune des adresses présentes dans la base de données, il est naturel de considérer que l'effort à fournir pour y procéder n'est pas « disproportionné » et ce, quel que soit le niveau de risque associé à la mise en œuvre du traitement.**

Une telle information devrait d'ailleurs poser d'autant moins de difficultés que l'éditeur cherchera souvent à contacter les professionnels pour les inviter, dans le cadre d'une action de prospection commerciale, à revendiquer leur fiche et à souscrire à des offres payantes (p. ex. : service de présentation personnalisée de leur entreprise, de mise en relation avec les internautes, etc.). En pareil cas, l'information fournie devra faire état de cette autre finalité ayant présidé à la collecte de leurs données.

*Lorsque l'éditeur de l'annuaire ne dispose pas des coordonnées électroniques des professionnels, une analyse plus fine devrait être menée.*

De façon générale, **plusieurs éléments permettent de présumer le caractère proportionné des efforts à fournir pour informer individuellement les personnes concernées :**

- **les annuaires peuvent faire peser des risques sur les droits et intérêts des professionnels référencés** (vie privée, réputation), qui pourront être élevés compte tenu notamment de la volumétrie, de la nature des données traitées, de l'audience du site et des fonctionnalités proposées ;
- **en pratique, les éditeurs disposent toujours de la faculté de les contacter** (leur adresse administrative figure dans les registres officiels diffusés en sources ouvertes) ;
- **l'information prévue par le RGPD pouvant être délivrée par tout moyen, ils n'ont pas forcément à supporter le coût financier d'un envoi postal** (p. ex. : transmission aux ordres professionnels concernés d'une mention d'information destinée à être relayée auprès de leurs membres au moyen de courriels individuels) ;
- **l'information constitue un élément clef de l'équilibre de la balance des différents intérêts en présence**, et le fait qu'elle soit fournie directement aux personnes concernées garantit qu'elles seront effectivement mises en mesure d'exercer leurs droits.

Néanmoins, **certains éditeurs pourront valablement invoquer l'exception** prévue par le RGPD, en particulier lorsque leur traitement est censé présenter un **caractère véritablement prévisible pour les personnes concernées**, ou n'occasionne qu'un **niveau de risque faible pour leurs droits et intérêts** compte tenu de ses caractéristiques (p. ex. : absence d'agrégation de données issue de sources diverses, ou d'outil de collecte et de publication d'avis en ligne sur la qualité des prestations fournies).

À titre illustratif, une association qui, dans le cadre d'actions visant à valoriser la composition du tissu artisanal de sa région, souhaiterait procéder à une « simple » rediffusion des données des professionnels concernés

figurant dans le registre national des entreprises, serait légitime à arguer du caractère disproportionné de l'effort à fournir.

**À noter :** dans ce cas, une information « RGPD » générale, complète et aisément accessible (distincte des conditions générales d'utilisation du site, en particulier) devra tout de même être publiée sur le site web de l'éditeur.

**En toute hypothèse, l'information devra être délivrée aussi tôt que possible, au plus tard lors de la première prise de contact avec les intéressés ou de la communication des données à des tiers et, dans tous les cas, dans un délai ne dépassant pas un mois après la collecte. Elle devra l'être de façon « concise, transparente, compréhensible et aisément accessible, en des termes clairs et simples » (art. 12 du RGPD), et intégrer l'ensemble des éléments requis en cas de collecte indirecte des données (art. 14 du RGPD).**

**Pour plus d'informations à ce sujet, [voir la fiche du présent guide dédiée à l'information des personnes](#).**

### **Exemple de mention d'information à l'intention des avocats référencés dans un annuaire « enrichi » publié sur Internet**

#### **Pourquoi le site « TousAvocats.fr » traite-t-il de mes données ?**

Le site « TousAvocats.fr », placé sous la responsabilité de notre société X (située ...), procède :

- **au référencement automatique de l'ensemble des avocats inscrits auprès d'un barreau français (création de fiches personnelles), sur la base d'une réutilisation d'informations publiques ;**
- **à la collecte et la publication d'avis relatifs à la qualité de leurs prestations, telle que ressentie par les personnes ayant fait appel à leurs services (notes sous forme d'étoiles assorties de commentaires).**

En tant qu'avocat en exercice, vous disposez d'une fiche dédiée, que nous vous invitons à consulter régulièrement, notamment pour vérifier l'exactitude des données y figurant. Vous pouvez y accéder suivant un critère matériel (spécialité), alphabétique (nom de famille) ou géographique (commune d'implantation), étant précisé qu'aucun classement « qualitatif » n'est effectué par notre société sur la base du calcul de notes moyennes.

Les traitements de données personnelles précédemment évoqués permettent, dans le cadre de la constitution et de la diffusion d'un annuaire « enrichi », la communication et l'obtention de nombreuses informations utiles et pratiques sur les modalités d'organisation et d'exercice par ses membres de la profession d'avocat. Ils se fondent juridiquement sur des « intérêts légitimes » au sens de l'article 6.1.f du règlement général sur la protection des données (RGPD), en ce qu'ils participent à la liberté d'expression et d'information de l'éditeur et des contributeurs, et sont réalisés dans des conditions respectueuses des intérêts, droits et libertés des personnes concernées.

#### **Quelles sont les conditions de mise en œuvre de ces traitements ?**

Votre fiche individuelle comporte des données relatives à votre identité, vos coordonnées professionnelles, votre ou vos spécialités et expériences professionnelles, ainsi qu'à tout élément relatif à la pratique de votre métier que vous avez souhaité publiquement valoriser (p. ex. : travaux de recherche, affaires traitées). Ces données proviennent :

- des sources publiques suivantes :

\* base SIRENE de l'INSEE et annuaires officiels d'avocats publiés par le Conseil national des barreaux et les ordres des avocats des différents barreaux, diffusés en open data à des fins de réutilisation par des tiers ;

\* sites web des personnes concernées et profils publics de réseaux sociaux professionnels (X et XX) ;

- des éléments que vous aurez souhaité apporter en complément, après revendication gratuite de votre fiche (procédure accessible ici : ...).

Ces informations sont complétées de celles (notes sous forme d'étoiles et commentaires) qui résultent de l'exercice par les internautes de leur faculté de publier un avis en ligne, dans le respect notamment des conditions légales encadrant une telle pratique, telles que rappelées par notre charte relative à la construction, la vérification et la modération des avis (charte accessible ici : ...).

Enfin, ces données, régulièrement actualisées (voir ici les conditions de leur mise à jour : ...), sont conservées et publiées aussi longtemps que vous serez inscrit auprès d'un barreau français, à l'exception des avis en ligne dont le traitement n'excèdera pas cinq ans. Néanmoins, en vertu de vos droits d'opposition et d'effacement, les données vous concernant peuvent faire l'objet d'une suppression en amont.

### **Quels sont mes droits à l'égard de ces traitements et comment les exercer ?**

Conformément au RGPD, vous disposez de droits d'accès, de rectification, d'opposition, d'effacement et de limitation (consulter ici la page dédiée aux droits : ...), que vous pouvez exercer gratuitement et facilement à partir de votre compte en ligne, ou en contactant par courriel, courrier ou téléphone le délégué à la protection des données de notre société : ... .

À noter que vous disposez également du droit d'introduire une réclamation auprès de la CNIL.

## **Respecter les droits des personnes**

Les éditeurs d'annuaires devront enfin **garantir l'effectivité des droits « informatique et libertés » dont bénéficient les professionnels** concernés par leurs traitements (voir à ce sujet [la sanction de la CNIL n° SAN-2021-014 du 15 septembre 2021](#)). Ces droits constituent un ensemble de leviers d'action leur permettant de veiller à la préservation de leurs intérêts.

### **Le droit, pour les personnes référencées, de s'opposer à leur présence dans l'annuaire ou d'obtenir l'effacement de certaines informations les concernant**

Comme indiqué précédemment, en présence d'une personne s'opposant de façon globale, pour des raisons tenant à sa situation particulière, au référencement de ses données dans l'annuaire, avec les notes et commentaires associés, la CNIL estime qu'il n'y a pas, sauf cas exceptionnel, de « motif impérieux et légitime » à conserver un professionnel dans un annuaire privé pour lequel il y a une situation litigieuse entre le professionnel et les commentateurs. L'éditeur devrait donc en principe accueillir favorablement la demande (interruption du traitement de ses données dans son intégralité) de tout professionnel souhaitant ne plus figurer dans son annuaire.

**Exemple :** un médecin ou avocat dont la fiche comprend de mauvais commentaires, auxquels le secret professionnel lui interdit de répondre de façon circonstanciée, pourra obtenir la suppression de sa fiche au sein de l'annuaire au regard de son impact sur son image et sa réputation.

À noter que si le traitement est mis en œuvre à des fins de prospection, le droit d'opposition sera obligatoirement discrétionnaire (art. 21.3 du RGPD).

S'agissant des demandes de suppression portant sur certaines des données diffusées (p.ex. : notes et commentaires), elles devraient faire l'objet d'une appréciation au cas par cas par le responsable du traitement, à la lumière des dispositions de [l'article 17 du RGPD](#) relatives au droit à l'effacement : analyse *in concreto* des motifs particuliers invoqués en cas d'opposition à leur traitement (pas de refus systématique et standardisé), du caractère illicite du traitement<sup>15</sup>, etc.

<sup>15</sup> Dans un arrêt du 4 mai 2023 (affaire C 60/22), la Cour de justice de l'Union européenne (CJUE) souligne que le « traitement illicite » visé à l'article 17 du RGPD est le traitement mis en œuvre en « violation par le responsable du traitement du principe de « responsabilité » tel qu'énoncé à l'article 5, paragraphe 2, dudit règlement, lu conjointement avec l'article 5, paragraphe 1, sous a), et l'article 6, paragraphe 1, premier alinéa, de ce dernier » ; le traitement « doit satisfaire aux conditions de licéité du traitement énumérées à l'article 6 dudit règlement. Par ailleurs, dans la mesure où les articles 7 à 11 du RGPD (...) ont pour objet de préciser la portée des obligations incombant au responsable du traitement en vertu de l'article 5, paragraphe 1, sous a), et de l'article 6, paragraphe 1, dudit règlement, le traitement de données à

**Exemple :** toute personne dont les coordonnées postales professionnelles correspondent à sa résidence personnelle devrait pouvoir obtenir leur suppression de l'annuaire (le fait que cette information soit par ailleurs obligatoirement diffusée dans un registre officiel apparaît sans incidence, compte tenu de l'exposition supplémentaire que sa duplication emporte) ; de même, si la liberté d'expression et d'information peut mettre en échec une demande d'effacement, un professionnel pourra solliciter la suppression des commentaires d'internautes présentant un caractère diffamatoire, injurieux ou simplement excessif (sans lien avec leur pratique professionnelle ou ne satisfaisant pas, manifestement, l'objectif poursuivi par leur traitement, à savoir l'information du public).

Enfin, tout éditeur d'annuaires doit prendre des mesures raisonnables pour informer les principaux moteurs de recherche des demandes d'effacement auxquelles il donne une suite positive (art. 17.2 du RGPD).

## **Le droit d'accéder aux données les concernant et d'obtenir des informations sur les conditions de mise en œuvre de leur traitement**

Tout professionnel référencé dans l'annuaire doit, en particulier, pouvoir obtenir, dans les conditions prévues à [l'article 15 du RGPD](#), une copie des données le concernant.

## **Le droit de demander la rectification, la suppression ou l'enrichissement de toutes les données les concernant qui sont inexactes, obsolètes ou incomplètes**

**Exemple :** un professionnel en reconversion pourra exiger la suppression des données le concernant figurant dans un annuaire thématique relatif à son ancien secteur d'activité ; toute personne dont la carrière a évolué (nouvelles expériences professionnelles) pourra demander la prise en compte dans sa fiche des dernières évolutions.

## **Le droit de demander la limitation du traitement (cessation temporaire de la diffusion des données sur Internet)**

**Exemples :** le temps que l'éditeur de l'annuaire vérifie l'inexactitude de données mises en cause, ou, lorsque la personne s'est opposée au traitement de ses données, le temps qu'il analyse le point de savoir s'il dispose de motifs légitimes et impérieux prévalant sur ceux de la personne concernée.

La bonne prise en compte de ces droits par les éditeurs d'annuaires suppose qu'ils aient mis en place :

- des **dispositifs visant à faciliter leur exercice** (p. ex. : fourniture de différents points de contact dédiés à la soumission des demandes ; faculté laissée aux personnes concernées d'agir directement sur la base de données, après vérification de leur légitimité à le faire) ;
- ainsi que des **procédures garantissant leur traitement effectif dans le délai maximal d'un mois** (art. 12 du RGPD).

Pour plus d'informations au sujet des droits des personnes concernées, [voir la fiche du présent guide dédiée à ce sujet](#).

---

*caractère personnel, afin d'être licite, doit également respecter, ainsi qu'il ressort de la jurisprudence de la Cour, ces autres dispositions dudit chapitre qui concernent, en substance, le consentement, le traitement de catégories particulières de données personnelles à caractère sensible et le traitement de données personnelles relatives aux condamnations pénales et aux infractions ».*

## Fiche cas d'usage n°3 : la réutilisation de données publiquement accessibles à des fins de constitution ou d'enrichissement de fichiers destinés à la prospection commerciale

La CNIL reçoit régulièrement des plaintes relatives aux pratiques de sociétés récupérant sur des sites web des données personnelles à des fins de prospection commerciale, alors même que certains des intéressés ont expressément indiqué s'opposer à une telle réutilisation de leurs informations.

Les sociétés visées ont généralement recours à des logiciels de moissonnage automatisé de données (ou *web scraping* en anglais) et sont notamment :

- des entreprises qui collectent pour leur propre compte et sur les annuaires en ligne l'ensemble des données personnelles se rapportant à leur secteur géographique, et qui démarchent ensuite les personnes concernées sur leur offre de biens ou de services ;
- des entreprises spécialisées dans la « pige immobilière », qui constituent des bases de prospects en moissonnant les coordonnées des personnes les ayant diffusées sur des sites de « petites annonces », dans le but de les louer / revendre auprès d'agences immobilières qui les utiliseront dans le cadre de leur recherche de nouveaux mandats de commercialisation.

D'autres sociétés exploitent des données publiquement accessibles pour enrichir leurs propres bases de prospects d'informations supplémentaires les concernant (p. ex. : ajout d'adresses courriel ou de données tirées des réseaux sociaux professionnels). Elles commercialisent également, à des fins de prospection ciblée, des profils de consommateurs sur la base de critères tels que la profession, la catégorie socio-professionnelle, la localisation, l'âge, le sexe ou encore les centres d'intérêt.

Bien que publiquement accessibles, les données en question sont des données personnelles. Dès lors, elles doivent être réutilisées dans le respect des dispositions du RGPD et, le cas échéant, des dispositions spéciales applicables, en particulier celles du code des postes et communications électroniques (CPCE).

Cette fiche pratique aborde les principaux points de vigilance et garanties que ces organismes doivent prendre en compte : les conditions de licéité des traitements, l'information des personnes concernées, le respect de leurs droits, l'encadrement de la sous-traitance et la nécessité potentielle d'une analyse d'impact.

Axée sur la réutilisation de données *publiées* sur Internet, elle ne traite pas de la question des *cookies* et autres traceurs à des fins de ciblage publicitaire (voir à ce sujet : <https://www.cnil.fr/fr/cookies-et-autres-traceurs>). De plus, les conditions de mise en œuvre des actions de prospection commerciale faisant par ailleurs l'objet de [nombreux contenus disponibles sur le site web de la CNIL](#), elle s'intéresse plus particulièrement à la phase de constitution et d'enrichissement des bases de données qui seront utilisées à de telles fins.

### Garantir la licéité des traitements

Deux bases légales prévues par le RGPD sont susceptibles de fonder les opérations de collecte et d'exploitation de données en ligne à des fins de prospection commerciale (transmission des données à des partenaires commerciaux ou prospection par l'organisme ayant collecté les données) : [le consentement des personnes concernées](#) (art. 6.1.a) et [l'intérêt légitime du responsable du traitement](#) (art. 6.1.f).

Pour déterminer la base légale des traitements qu'ils mettent en œuvre, **les organismes doivent prendre en compte les éléments suivants.**

- **Les traitements pourront se fonder sur la base légale du consentement lorsque les personnes auront donné préalablement leur accord à la collecte et à l'utilisation envisagée de leurs données, de manière libre, spécifique, éclairée et univoque.**

Ce sera le cas, en particulier, lorsque :

- **les personnes auront répondu positivement à une demande directe et précise de l'organisme, les ayant contactées via des coordonnées dont il disposait déjà licitement dans le cadre de la gestion de ses activités commerciales** ; le message suivant aura, par exemple, pu leur être adressé : « *Pour améliorer la connaissance de nos prospects et cibler davantage les offres commerciales qui leur sont adressées, nous souhaiterions collecter et exploiter les informations liées à leur situation géographique,*

profession et centres d'intérêts qu'ils ont choisi de diffuser, le cas échéant, sur leurs profils publics des réseaux sociaux suivants : .... Acceptez-vous que nous procédions à un tel enrichissement des données vous concernant ? » ;

- **elles auront coché, sur le site où sont publiées leurs données, une case « J'accepte » (non pré-cochée) assortie de précisions sur les traitements en cause et l'identité des sociétés qui s'en prévalent** (nature des données exploitées, finalité de l'exploitation et noms des réutilisateurs autorisés) ; une telle situation suppose qu'un accord préalable ait été passé entre l'éditeur du site diffusant les données et les réutilisateurs de celles-ci.

**Attention :**

**L'acceptation de CGU informant l'utilisateur d'un site web de la possible utilisation de ses données à des fins commerciales, ou la validation sur le site d'une case à cocher par laquelle la personne concernée accepte, de façon générale et indifférenciée, la réutilisation de ses données à des fins de prospection commerciale, ne permettra pas à l'organisme de fonder son traitement sur la base légale du consentement.**

En effet, la validité du consentement suppose notamment que l'identité des responsables de traitement qui s'en prévalent, pour une ou plusieurs finalités particulières, soit portée à la connaissance de l'utilisateur au moment de consentir.

Pour en savoir plus sur les conditions de recours à la base légale du consentement, voir la [fiche pratique sur le site web de la CNIL](#).

- **L'utilisation licite de coordonnées publiées en ligne, à des fins de prospection par voie électronique et sans lien direct avec la profession de la personne concernée, est systématiquement soumise au recueil d'un consentement préalable.**

**Le consentement préalable des personnes concernées est requis** lorsque l'exploitation envisagée par l'organisme consiste, au moins pour partie :

- **dans l'envoi par voie électronique (courriels, SMS/MMS, automates d'appels et fax) de messages publicitaires<sup>16</sup> ; ou**
- **dans la commercialisation, pour une utilisation à de telles fins, de coordonnées de particuliers.**

En effet, l'article L. 34-5 du CPCE interdit la mise en œuvre d'opérations de prospection par voie électronique en l'absence de consentement préalable des personnes concernées.

Si cet article ne s'applique pas aux traitements mis en œuvre par les courtiers en données (« *data brokers* » en anglais) lorsqu'ils ne prospectent pas eux-mêmes les personnes concernées, la CNIL estime que les transmissions de coordonnées auxquelles ils procèdent, et qui sont destinées à permettre de telles actions de prospection par leurs clients, doivent également se fonder sur le consentement préalable<sup>17</sup>.

**Pour autant, lorsqu'il apparaît possible de considérer, au regard de différents critères développés au point suivant (point 3), que les personnes concernées pouvaient raisonnablement s'attendre à l'utilisation de leurs coordonnées électroniques à des fins de prospection commerciale, la CNIL admet que les organismes, sur la base de leur intérêt légitime, procèdent à leur collecte en ligne en vue de recueillir le consentement requis de la part des intéressés.**

<sup>16</sup> Ce point est souligné dans la décision de sanction pécuniaire prononcée par la CNIL à l'encontre d'une société exploitant les données de réseaux sociaux professionnels : n° [SAN-2020-018 du 8 décembre 2020](#).

<sup>17</sup> [Délibération n° 2021-131 du 23 septembre 2021 portant adoption d'un référentiel relatif aux traitements de données à caractère personnel mis en œuvre aux fins de gestion des activités commerciales, Légifrance](#).

Ainsi :

- une entreprise souhaitant prospector des particuliers par voie électronique sur son offre de services, et ayant pu, à cette fin, collecter en ligne des adresses électroniques et numéros de téléphone mobile sur la base de son intérêt légitime, devra obligatoirement adresser un premier message aux personnes concernées dépourvu de tout caractère promotionnel et exclusivement destiné à recueillir leur accord à la réception de sollicitations commerciales par courriels et/ou SMS ; les données recueillies à cette fin devront être détruites en l'absence de réponse positive ;
- de même, un courtier en données devra, avant de transmettre des adresses électroniques à ses clients aux fins de prospection par ceux-ci, recueillir l'accord préalable des personnes concernées en leur indiquant l'objectif de cette transmission et les catégories de destinataires ; il est recommandé de préciser, dans le message de demande, leur nombre approximatif et secteur d'activité ;
- enfin, les clients d'un courtier en données collectées en ligne devront également recueillir le consentement des personnes concernées avant toute opération de prospection par courriel ou SMS/MMS ; ce consentement pourra être recueilli soit par eux-mêmes dans le cadre de l'envoi d'un premier message « neutre » aux intéressés, soit par l'intermédiaire du courtier en données, dans le cadre de la fourniture aux personnes concernées, au stade du recueil de leur consentement, d'une liste précise et actualisée de ses clients.

Pour en savoir plus sur les modalités de recueil des consentements suivant les cas de figure, voir le [référentiel de la CNIL « Gestion des activités commerciales »](#)<sup>18</sup>.

**À noter :**

**Sous réserve qu'elle entre bien dans le champ des attentes raisonnables des personnes concernées (voir le point suivant), la collecte comme l'exploitation des coordonnées postales et téléphoniques à des fins de prospection par voie non électronique (envois de courriers postaux et appels téléphoniques hors automates d'appel), y compris la transmission de ces données à de telles fins, pourra se fonder sur la base légale de l'intérêt légitime du responsable du traitement.**

**Il en va de même des opérations de prospection par voie électronique visant exclusivement des personnes sollicitées au titre de leur activité professionnelle spécifique (prospection dite « entreprises vers entreprises »).**

**Les droits des personnes concernées devront toutefois être impérativement respectés, comme précisé plus bas.**

Voir à ce sujet la [fiche pratique sur le site web de la CNIL](#).

- **En l'absence de consentement préalable des personnes concernées, leurs données pourront être collectées sur la base légale de l'intérêt légitime uniquement s'il apparaît que cette collecte ne porte pas d'atteinte disproportionnée à leurs droits et intérêts.**

**Selon le RGPD, le traitement ne peut être mis en œuvre sur la base légale de l'intérêt légitime lorsque les droits et intérêts des personnes concernées prévalent sur les intérêts de l'organisme. Celui-ci doit donc effectuer une pondération entre les différents intérêts en présence, tenant compte, tout particulièrement, des attentes raisonnables de ces dernières.**

Cette prise en compte est essentielle : en l'absence d'un acte positif et explicite de la part des personnes concernées, l'intérêt légitime requiert de ne pas les surprendre dans les modalités de mise en œuvre comme dans les conséquences du traitement.

<sup>18</sup> Délibération précitée n° 2021-131.

**En amont de toute collecte de données en ligne, il appartient ainsi à l'organisme de considérer différents critères (méthode du faisceau d'indices) pour évaluer, au cas par cas, le caractère *a priori* acceptable et prévisible pour les individus de l'exploitation envisagée de leurs données.**

Les critères à considérer portent notamment sur les points suivants.

#### *La finalité poursuivie et la nature des données exploitées*

Le caractère plus ou moins intrusif du traitement envisagé doit impérativement être pris en compte quand il s'agit d'analyser s'il n'heurte pas les droits et intérêts des personnes concernées et, en particulier, si celles-ci peuvent raisonnablement s'attendre à la collecte de leurs données dans l'objectif poursuivi par l'organisme.

En effet, celles-ci s'attendront plus facilement à ce que soient collectées leurs coordonnées postales et/ou électroniques, éventuellement assorties de données d'identité et d'informations relatives à leur situation géographique et professionnelle, que d'autres catégories de données permettant d'établir des profils plus détaillés : par exemple, des informations relatives à leurs centres d'intérêts, préférences, ressentis, conditions / habitudes de vie et interactions sociales.

#### **À retenir :**

**Les traitements intrusifs visant à profiler des individus, suivant différents critères et sur la base de données collectées en ligne, aux fins d'établissement de profils et de ciblage publicitaire sont par nature porteurs de risques importants pour leurs droits et intérêts.**

**En effet, les processus d'analyse des données publiquement accessibles sur les réseaux sociaux, blogs et forums, en vue de construire des profils permettant de mieux cerner la personnalité, les habitudes de vie, la situation sociale et/ou financière, les ressentis, relations, préférences et centres d'intérêts des personnes concernées, sont susceptibles d'aboutir à des conclusions et prédictions inexactes, voire à des décisions défavorables aux personnes, ainsi que de perpétuer des stéréotypes et d'enfermer les individus dans leurs choix (voir à ce sujet la [fiche pratique sur le site web de la CNIL](#)).**

**Ainsi, leur licéité sera en principe subordonnée au recueil d'un consentement préalable compte tenu du caractère particulièrement intrusif du processus de profilage : agrégation de données concernant une même personne tirées de différents sites web, recoupements et inférences de données, etc.**

**Il faut également noter que la licéité d'opérations tendant à collecter ou à inférer des données sensibles (p. ex. : données révélant les convictions religieuses ou concernant la santé ou la vie sexuelle) requiert le recueil d'un consentement explicite, sauf s'il est possible de mobiliser une autre exception à l'interdiction de traiter des données sensibles (art. 9 du RGPD), ce qui apparaît peu probable. Pour plus de précisions à ce sujet (notamment concernant les éléments à prendre en compte), voir la fiche du présent guide relative à la minimisation des données traitées.**

#### *La possibilité qui est ou non laissée aux personnes par le site source de s'opposer à la réutilisation de leurs données à des fins de prospection commerciale*

Le fait que l'éditeur du site web source ait prévu, sur le modèle de ce qui existe en matière d'annuaires universels (« liste orange »), un mécanisme d'opposition à ce type de réutilisations (mécanisme que l'on trouve par exemple sur certains sites de petites annonces de particuliers et professionnels) constitue un élément important à cet égard.

Il peut, par exemple, permettre à des sociétés spécialisées dans l'édition de pages immobilières de collecter les coordonnées des vendeurs de biens immobiliers ne s'étant pas expressément opposés à une telle exploitation de leurs données, compte tenu également du lien existant entre le contexte de la diffusion des données et l'objet des opérations de prospection envisagées.

**Attention : la collecte des données se rapportant aux personnes ayant exprimé leur opposition à ce type de traitement, en cochant la case prévue à cet effet par le site éditeur ou en ne décochant pas une case qui permet une opposition par défaut, constitue un traitement déloyal de données personnelles interdit par le RGPD ([article 5.1.a](#)).**

Il en va de même, en l'absence de mécanisme d'opposition mis à leur disposition par l'éditeur, lorsque les personnes ont manifesté d'elles-mêmes leur refus à être démarchées par des sociétés (p. ex. : insertion d'une mention dans leur petite annonce).

#### *La nature du site web et les conditions d'accès aux données*

Le contexte de la collecte et la nature du site web qui diffuse les données doivent également être prise en compte.

Par exemple, il sera plus aisé de reconnaître le caractère prévisible de la collecte des données à des fins de prospection lorsqu'elles sont publiées sur un annuaire universel que sur un site de petites annonces spécifiquement dédié à la commercialisation, entre particuliers, de biens ou de services.

De même, devrait être considéré le fait que les données figurent ou non dans des contenus d'ordre essentiellement personnel d'ordre essentiellement personnel (ce n'est pas *a priori* pas le cas de ceux publiés sur un réseau social professionnel), ou encore que leur accès soit ou non subordonné à la création préalable d'un compte, devrait être considéré.

#### *Le contenu de la politique de confidentialité ou des CGU du site éditeur*

La collecte des données publiées pour réaliser des opérations de démarchage est susceptible de ne pas entrer dans les attentes raisonnables des personnes lorsque la politique de confidentialité ou les CGU de l'éditeur :

- interdisent le moissonnage automatisé ou la récupération manuelle des données à des fins de prospection commerciale ; ou
- prévoient l'absence de transmission de données à des partenaires commerciaux ; ou
- subordonnent la transmission de données au consentement préalable des personnes concernées.

#### *L'objet de la prospection*

La collecte des données à des fins de prospection commerciale peut davantage entrer dans les attentes raisonnables des personnes si elle concerne des biens ou services en lien direct avec :

- l'activité professionnelle des personnes concernées (p. ex. : prospection par un organisme de formation en cybersécurité des délégués à la protection des données dont les coordonnées sont publiées en tant que données ouvertes (*open data*) par la CNIL ; prospection par une entreprise pharmaceutique des médecins référencés dans l'Annuaire Santé diffusé par l'ANS) ; ou
- l'objet de leurs petites annonces (p. ex. : prospection des vendeurs de biens immobiliers par une agence immobilière, prospection des vendeurs de voitures par des garages).

À l'inverse, si elle concerne des biens ou services sans aucun lien avec ces derniers (p. ex. : prospection des avocats, dont les coordonnées figurent dans l'annuaire en ligne du Conseil national des barreaux, par une entreprise de cosmétiques), il sera plus difficile de considérer que les personnes peuvent raisonnablement s'y attendre.

**Lorsqu'il apparaît, au regard des critères exposés ci-dessus, que l'équilibre de la balance des intérêts n'est pas satisfait, notamment du fait que le traitement envisagé ne s'inscrit pas dans le cadre des attentes raisonnables des personnes concernées, l'organisme doit s'abstenir de procéder à la collecte des données publiées en ligne, y compris aux fins d'envoyer un premier courriel pour solliciter un consentement.**

**Dans le cas contraire, et comme vu précédemment, l'organisme pourra :**

- **les collecter aux seules fins de recueillir le consentement des personnes concernées, lorsque leur exploitation le requiert** (cas de la prospection « entreprises vers consommateurs » par voie électronique obligatoirement fondée sur le consentement) ;
- **les collecter et les exploiter sur la base légale de son intérêt légitime**, lorsque la licéité de leur utilisation n'est pas assujettie au recueil d'un consentement préalable (cas de la prospection par voie non électronique ou effectuée en lien direct avec la profession des personnes concernées).

Dans tous les cas, ces opérations devront être effectuées dans le respect des autres dispositions du RGPD, en particulier de celles relatives à l'information et aux droits des personnes concernées.

## Assurer la transparence des traitements et veiller à l'effectivité des droits des personnes concernées

### Les exigences liées à l'information des personnes concernées

Tous les organismes doivent respecter le **principe de transparence des traitements de données personnelles** : même dans les hypothèses où la collecte des données en ligne puis leur exploitation pour de la prospection commerciale peuvent se fonder sur leur intérêt légitime, ils doivent informer les personnes concernées des objectifs qu'ils poursuivent et des conditions de leur mise en œuvre.

Cette information permet aux personnes concernées de **conserver la maîtrise des usages qui sont faits de leurs données**, en les mettant en capacité de consentir de façon éclairée au traitement envisagé ou, *a minima*, d'exercer leurs droits, notamment les droits d'opposition et d'effacement.

Cette information devra être délivrée dans les conditions prévues aux articles [12 et 14 du RGPD](#), à savoir :

- **en principe de manière individuelle** ;
- **aussi tôt que possible**, au plus tard lors de la première prise de contact avec les intéressés ou, le cas échéant, de la communication des données à des tiers, et, dans tous les cas, **dans un délai ne dépassant pas un mois** après la collecte ;
- **de façon concise, transparente, compréhensible et aisément accessible, en des termes clairs et simples**.

Son contenu devra **intégrer l'ensemble des éléments en principe requis en cas de collecte indirecte des données, et notamment** :

- les finalités et la base légale du traitement, comportant des précisions sur l'intérêt légitime mobilisé, le cas échéant ;
- les catégories de données collectées ;
- les destinataires ou catégories de destinataires de données ;
- la ou les sources d'où elles proviennent (lorsque plusieurs sources ont été exploitées et qu'il n'est pas possible d'attribuer les données d'une personne à une source en particulier, des informations générales sur ces sources doivent être fournies ; ces informations générales doivent toutefois préciser, lorsqu'ils sont connus, les noms des différents sites web sur lesquels les données auront été collectées) ;
- le cas échéant, l'existence d'un profilage, et des informations utiles concernant sa logique sous-jacente, ainsi que l'importance et les conséquences prévues pour la personne concernée ;
- les droits dont disposent les personnes concernées.

### Les exigences liées aux droits des personnes concernées

Tout organisme collectant des données sur Internet aux fins de prospection commerciale, pour son propre compte ou par des tiers partenaires, doit :

- **tenir compte des oppositions « absolues » (indépendantes de tel ou tel traitement) exprimées par des personnes à une telle utilisation de leurs données** (personnes inscrites sur des [listes anti-prospection](#), tel le dispositif [BLOCTEL](#)) ;
- **ne pas démarcher les personnes s'étant déjà opposées / ayant retiré leur consentement** à recevoir des sollicitations commerciales de sa part ;
- traiter les demandes d'exercice des **droits d'accès, de rectification, d'effacement et de limitation** ([art. 15 à 18 du RGPD](#)), dans les délais et conditions prévus par les textes ;

- **notifier aux partenaires commerciaux toute demande de rectification, effacement de données ou limitation du traitement** effectué ([art. 19 du RGPD](#)) lorsque les données leur ont été transmises ;
- **notifier à ces derniers toute demande de retrait du consentement effectuée, le cas échéant, auprès de lui** (article 7.3 du RGPD) lorsqu'il a recueilli le consentement pour les opérations de prospection réalisées par ses partenaires.

Pour plus d'informations, voir [le principe de transparence des traitements de données personnelles](#) et [les droits des personnes concernées par les traitements](#) dans ce guide.

## Réaliser si nécessaire une analyse d'impact sur la protection des données

Si le traitement projeté est susceptible d'engendrer des risques pour les droits et libertés des personnes concernées, au regard des critères identifiés par le Comité européen de la protection des données (p. ex. : traitement à large échelle, profilage, traitement de données sensibles, etc.), une analyse d'impact relative à la protection des données (AIPD) devra être obligatoirement réalisée avant la mise en œuvre du traitement (voir la [rubrique dédiée du site de la CNIL](#)).

Dans tous les cas, il s'agit d'une bonne pratique permettant à l'organisme de s'assurer, en liaison avec le délégué à la protection des données lorsqu'une telle personne a été désignée, que le traitement sera respectueux du RGPD.

## Fiche cas d'usage n°4 : La réutilisation de données publiquement accessibles à des fins de recherche scientifique (hors santé)

*Le web est source de quantités de données offrant des capacités inédites dans différents domaines de la recherche. Toutefois, le traitement de données personnelles publiquement accessibles, même pour une finalité légitime de recherche scientifique, reste soumis aux obligations du RGPD et de la loi « informatique et libertés ».*

En complément [des fiches déjà publiées par la CNIL](#) sur la conformité des traitements de données à des fins de recherche (hors santé), cette fiche se concentre sur les spécificités des cas de réutilisation de données publiquement accessibles.

Elle n'aborde pas [les traitements mis en œuvre à des fins de recherche, d'étude ou d'évaluation dans le domaine de la santé](#) qui font l'objet d'un encadrement juridique spécifique, résultant des articles 66 et suivants de la loi Informatique et Libertés. Cette fiche reste toutefois applicable aux traitements de données de santé dans le cadre d'études ne relevant pas du domaine de la santé.

### Qu'est-ce qu'une « recherche scientifique » au sens de la réglementation en matière de protection des données ?

La notion de « recherche scientifique » bénéficie d'une acception large dans le RGPD. Son considérant (159) énonce : « 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 ».

[Le Comité européen à la protection des données \(CEPD\)](#), dans [ses lignes directrices sur le consentement](#), donne un éclairage supplémentaire en indiquant que la recherche scientifique se caractérise par « un projet de recherche établi conformément aux normes méthodologiques et éthiques du secteur en question, conformément aux bonnes pratiques ».

La recherche a ainsi pour objet de produire des connaissances nouvelles dans tous les domaines dans lesquels la méthode scientifique est applicable.

Afin d'aider les responsables de traitement à déterminer s'ils peuvent bénéficier des dispositions relatives à la recherche scientifique, **la CNIL propose un faisceau de critères permettant d'aider le responsable de traitement à déterminer si le traitement qui poursuit une finalité de recherche, relève de la recherche scientifique :**

- si le responsable de traitement remplit **un ou les deux critères organiques**, il peut bénéficier d'une présomption que son traitement relève de la « **recherche scientifique** » ;
- en revanche, s'il ne peut remplir **aucun des critères organiques**, il devra alors **examiner sa situation au regard de l'ensemble des critères de fond énumérés ci-dessous**.

Ces orientations d'interprétation ont pour vocation d'éclairer les responsables de traitement et ne présentent pas de caractère réglementaire.

### Les critères organiques

Ces critères relatifs à la nature du responsable de traitement et celle du financement du projet constituent des indices, qui pourront être considérés comme suffisants en fonction des situations.

- **Critère n° 1 : la nature du responsable de traitement**

**Par exemple :** une université, un organisme de recherche tel que le CNRS, un centre de recherches comme l'Institut de recherche criminelle de la Gendarmerie nationale (IRCGN) ou une entreprise dédiant une grande

partie de son activité à la « R&D » (dont les entreprises bénéficiant du statut de jeune entreprise innovante – JEI) ont vocation à mettre en œuvre ce type de traitement.

**À noter :** La qualification de responsable de traitement doit s'apprécier au cas par cas et en fonction de l'organisme s'il y en a un (v. [les lignes directrices du CEPD](#), en anglais).

En principe, en dehors du cas où le chercheur agit exclusivement pour son propre compte (chercheur « indépendant », c'est-à-dire sans rattachement à un organisme particulier et/ou sans mandat confié pour la réalisation de la recherche), le responsable du traitement mis en œuvre est l'organisme public ou privé au sein duquel il effectue sa recherche (université, entreprise privée, etc.). En effet, y compris lorsque le chercheur dispose d'une grande autonomie, l'organisme au sein duquel il travaille met à sa disposition des moyens de traitement pour réaliser des recherches, doit veiller à ce que les règles de protection des données personnelles soient respectées par ses chercheurs et peut pour cela donner des instructions aux chercheurs et en assurer le respect. Cela justifie que l'organisme assume la responsabilité des traitements mis en œuvre.

#### • Critère n° 2 : le mode de financement

Les projets de recherche peuvent être financés soit par des fonds publics, soit par des fonds privés, soit par des fonds mixtes.

**Par exemple :** dans le cas d'une entreprise privée responsable de traitement, le fait que le projet fasse l'objet d'une mesure d'incitation fiscale spécifique au soutien à la recherche scientifique (en particulier, le crédit impôt recherche – CIR) ou d'une aide publique à la recherche et au développement (par exemple, un financement de l'agence nationale de la recherche – ANR), l'indice pourrait être considéré comme suffisant.

#### **Focus sur la présomption de qualification de chercheur au sens de l'article 40-8 du règlement sur les services numériques**

Un chercheur qui remplit les conditions fixées à l'article 40-8 du règlement sur les services numériques s'agissant de l'accès aux données des très grandes plateformes ou très grands moteurs de recherche aux chercheurs agréés, est présumé mettre en œuvre un traitement qui s'inscrit dans le cadre d'une finalité de recherche scientifique au sens du RGPD.

En effet, le [règlement sur les services numériques](#) exige plusieurs critères (parmi lesquels qu'ils soient affiliés à un organisme de recherche au sens de l'article 2-1 de la directive (UE) 2019/790, indépendants de tous intérêts commerciaux et que leur demande indique la source de financement des recherches, etc.).

### Les autres critères de fond

À défaut de disposer d'indices suffisants au regard des critères n°s 1 et 2, il convient d'examiner conjointement les critères suivants (fondés sur le Manuel de Frascati et sur la définition de la R&D au sens du Crédit Impôt Recherche). Ces critères étant cumulatifs, le responsable de traitement devra en principe démontrer qu'ils sont tous remplis pour que le traitement puisse être considéré comme relevant de la recherche scientifique au sens du RGPD. Lorsque ce n'est pas le cas, une analyse au cas par cas est nécessaire pour qualifier le traitement.

- **Critère n° 3.1 : la nouveauté :** Le traitement doit viser à obtenir des résultats nouveaux : l'objet de la recherche peut aider à la qualification de la recherche scientifique.

À cet égard, la publication d'articles dans une revue à comité de lecture ou l'octroi d'un brevet permet de qualifier le critère de nouveauté.

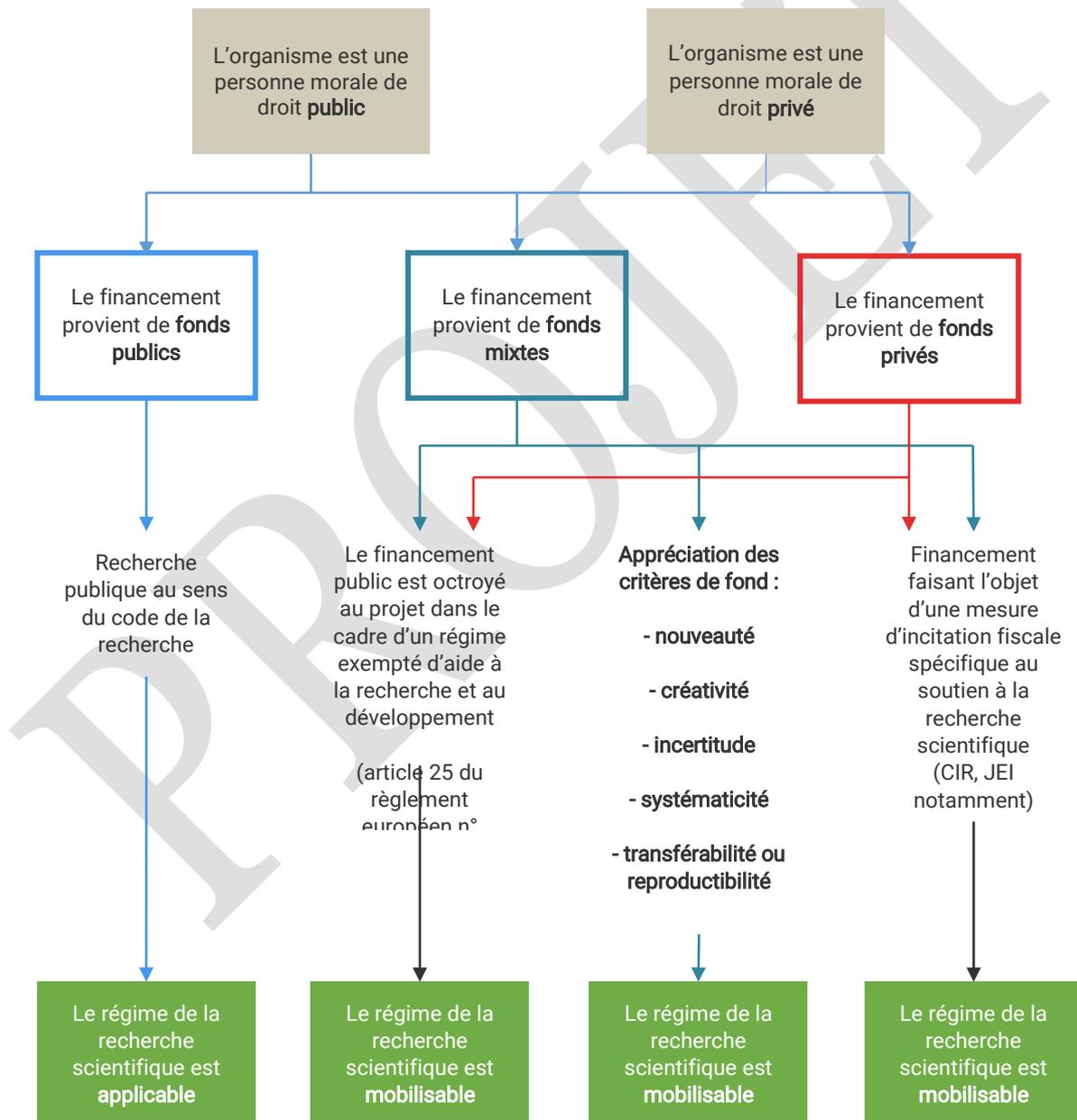
- **Critère n° 3.2 : la créativité :** Ce critère repose sur des notions et hypothèses originales et non évidentes – l'apport des travaux à la connaissance scientifique ou à l'état de la technique. Le développement d'un savoir collectif qui ne profite pas seulement à l'entité morale porteuse du projet de recherche est un indice fort pour qualifier celle-ci de scientifique.

**Critère n° 3.3 : l'incertitude :** Le traitement doit revêtir un caractère incertain quant au résultat final.

- **Critère n° 3.4 : la systématique** : le traitement doit s'inscrire dans une planification et une budgétisation, mettre en œuvre une méthodologie scientifique. Le respect de normes sectorielles pertinentes de méthodologie et d'éthique est un indice fort pour qualifier la recherche de scientifique.
- **Critère n° 3.5 : la transférabilité-reproductibilité** : le traitement doit déboucher sur des résultats qu'il est possible de reproduire ou de transférer dans un champ plus large que celui de la recherche.

À titre d'exemple, la publication de l'étude réalisée et la présentation de la méthodologie de recherche adoptée est un indice fort permettant de souligner la volonté de partage du ou des porteurs de projet.

## Schéma récapitulatif des critères de qualification de la « recherche scientifique » au sens de la réglementation en matière de protection des données



## La licéité de la réutilisation des données

Le fait que le traitement poursuive un objectif de « recherche scientifique » au sens du RGPD ne suffit pas à garantir sa licéité.

Schématiquement, un réutilisateur de données personnelles doit d'une part se fonder sur une **base légale** au sens de l'article 6 du RGPD et, d'autre part sur une dérogation prévue par l'article 9 du RGPD s'il entend traiter des données sensibles.

### L'exigence d'une base légale : généralités

En matière de recherche scientifique, les bases légales les plus pertinentes sont :

- **La mission d'intérêt public**, à condition que celle-ci soit prévue par une base juridique dans le droit auquel l'organisme est soumis (tel qu'une réglementation européenne, une loi ou un acte réglementaire), et que les traitements à des fins de recherche scientifique s'inscrivent dans cette mission.
- Si le responsable du traitement n'est pas investi d'une telle mission, **la base de l'intérêt légitime**, à condition qu'il prévale sur les intérêts, droits et libertés des personnes concernées. Sauf cas particulier, les personnes morales de droit public ne peuvent pas invoquer cette base légale.
- **Le consentement des personnes concernées au traitement de leurs données**. Bien qu'il soit difficile à recueillir s'agissant de la réutilisation de données publiquement accessibles, il sera souvent la meilleure base légale pour les traitements particulièrement intrusifs pour les personnes concernées.

Lorsque cela sera possible (par exemple au moyen d'un outil de messagerie accessible par toute personne), le chercheur pourra solliciter directement auprès des personnes concernées le recueil de leur consentement pour traiter leurs données à des fins de recherche, en s'assurant que ce [consentement est libre, spécifique et éclairé, qu'il peut être retiré à tout moment](#). Sinon, il devra solliciter le responsable du traitement de la publication initiale des données pour que ce dernier collecte un consentement valable en son nom (bien que cela suppose une coopération avec ce dernier qui n'est pas toujours possible en pratique).

### L'exigence d'une base légale : cas d'application

Lorsqu'il s'agit de réutiliser des données publiquement accessibles à des fins de recherche scientifique, trois cas de figure sont à envisager :

- **Lorsque le chercheur entend réutiliser des données mises à disposition du public en vertu de la législation française ou européenne sur l'ouverture des données (« open data »)**, qui prévoit leur libre réutilisation.

Dans ce cas, **le consentement ne sera presque jamais nécessaire**.

**Par exemple**, tel serait le cas si des chercheurs, salariés d'une ONG, menaient une recherche sur la liberté d'expression en France et envisageaient l'analyse d'un nombre important de décisions de justice mises à la disposition du public sur Légifrance (étant précisé que la recherche n'aurait ni pour objet ni pour effet de permettre la réidentification des personnes concernées, conformément aux articles 44.5 et 46.5 de la loi Informatique et Libertés). Ce traitement peut être fondé sur l'intérêt légitime poursuivi par l'ONG qui serait alors le responsable du traitement.

- **Lorsque les données n'ont pas été publiées dans une perspective d'open data** mais pour une finalité particulière (par exemple les données figurant sur les réseaux sociaux ou des forums de discussions spécialisés). Dans ce cas, **une analyse au cas par cas sera requise pour s'assurer que le traitement ne porte pas une atteinte disproportionnée aux intérêts, droits et libertés des personnes concernées**.

En particulier, il s'agira de prendre en compte :

- le caractère plus ou moins privé des informations recueillies ;

- les catégories de personnes (p. ex. : personnalités publiques, mineurs) et de données concernées, plus ou moins porteuses de risques (p. ex. : données professionnelles ou données relatives à la vie privée),
- les mesures susceptibles d'être adoptées pour limiter l'impact du traitement sur les personnes décrites ci-dessous (p. ex. : anonymisation à bref délai, pseudonymisation, droit d'opposition inconditionnel).

**Par exemple :**

- Des chercheurs au sein d'un laboratoire public étudient l'évolution de la langue française sur Internet. Ils décident pour cela de collecter et d'analyser des commentaires publiés librement sur différents réseaux sociaux qu'ils anonymisent à bref délai pour suivre le nombre d'occurrences de certaines expressions ou formes orthographiques. Dans la mesure où le responsable de traitement est un laboratoire public, la base légale de la mission d'intérêt public pourrait être mobilisée.
- Un institut de recherche privé mène une recherche scientifique pour évaluer l'impact sur l'emploi de l'IA générative pour la génération de code informatique, et envisage de collecter et d'analyser des données publiées sur un réseau social professionnel (par exemple, des profils publics, des commentaires et publications librement accessibles). L'intérêt légitime peut être envisagé selon les cas comme une base légale valable à condition que des mesures suffisantes soient prises (telles que l'anonymisation des données lorsque cela est possible, leur pseudonymisation à défaut, ou encore telles que des mesures de sécurité des données).
- **Enfin, il arrive que des jeux de données contenant des données personnelles soient librement mis à disposition sur Internet en dehors de toute législation.** Le plus souvent, il s'agit de données qui étaient déjà publiquement accessibles et qui constituent une base de données ou un corpus diffusé sur le site web d'une université ou d'une plateforme dédiée au partage de jeux de données, pour faciliter leur réutilisation.

Dans ce cas spécifiquement, et de façon plus générale, le contrôle de la licéité de la mise en ligne de la base de données relève en premier lieu du responsable de traitement qui opère cette mise en ligne. Cependant, l'opérateur qui réutilise les données doit s'assurer qu'il n'est pas en train de réutiliser un jeu de données dont la constitution était manifestement illicite (par exemple provenant d'une fuite de données).

En effet, la CNIL estime qu'il résulte du principe général de licéité des traitements de l'article 5.1.a du RGPD que le réutilisateur ne saurait réutiliser un jeu de données dont il ne peut ignorer qu'il est constitué ou mis en ligne en méconnaissance du RGPD ou de règles plus générales (telles que celles interdisant les atteintes à la sécurité des systèmes d'information ou des atteintes à des droits de propriété intellectuelle).

En outre, la personne qui télécharge ou réutilise un jeu de données manifestement illégal risque de se rendre coupable du délit de recel (article 321-1 du code pénal).

Si la possibilité de réutiliser un jeu de données librement mis à disposition sur Internet n'est pas subordonnée à des vérifications approfondies sur le respect de l'ensemble des règles du RGPD ou d'autres règles juridiques applicables (droit d'auteur, données couvertes par le secret des affaires, etc.), vérifications qui relèvent en premier lieu de l'opérateur qui met en ligne les données, la CNIL recommande aux réutilisateurs de s'assurer :

- **que la description du jeu de données mentionne leur source ;**

**Par exemple :** sur un blog librement accessible sur Internet, un chercheur identifie un jeu de données comportant un ensemble de discussions entre particuliers. Selon sa description, les données sont pseudonymisées mais aucune source n'est mentionnée. Le chercheur devrait alors s'abstenir de le réutiliser sans obtenir davantage de précision qui lui permettrait de lever ses doutes quant à la conformité de sa constitution et de sa diffusion.

- que la constitution ou la diffusion du jeu de données **ne résulte pas manifestement d'un crime ou d'un délit ou a fait l'objet d'une condamnation ou d'une sanction publique** de la part d'une autorité compétente ;

**Par exemple :** en cas d'achat d'un jeu de données sur le *dark web* provenant par exemple d'une atteinte à un système de traitement automatisé punie par la loi (au sens de l'article 323-1 du code pénal), un chercheur ne saurait en ignorer l'origine délictuelle et dont l'illicéité serait alors manifeste.

Il en irait de même pour un chercheur souhaitant réutiliser un jeu de données pour lequel une décision de justice a retenu une atteinte à un droit de propriété intellectuelle comme le droit *sui generis* des producteurs de bases de données (au sens de l'article L. 343 du code de la propriété intellectuelle).

- que l'origine des données soit suffisamment documentée pour qu'il n'y ait **pas de doutes flagrants sur la licéité** du jeu de données (notamment que le traitement source ne soit pas manifestement dépourvu de base légale lorsque les données sont tellement intrusives qu'elles ne sauraient être traitées sans le consentement des personnes) ;

**Par exemple :** sur une plateforme d'échange de jeux de données, un chercheur repère un ensemble compilant les trajets domicile-travail de milliers de personnes. Sa description explique qu'il s'agit de données de géolocalisation précises, non anonymes. Dans cette hypothèse, le chercheur ne saurait ignorer qu'il existe un doute sérieux quant à la licéité de la diffusion d'un tel jeu de données en l'absence de consentement des personnes.

À l'inverse, un chercheur pourrait réutiliser un jeu de données dont la description ne laisse pas de doute flagrant quant à sa licéité. Il en irait ainsi d'un jeu de données pseudonymisées, initialement rendues publiques par les personnes concernées sur un site web identifié et qui ne contiendrait pas de données sensibles.

Il en irait de même pour la réutilisation d'un jeu de données dont le diffuseur les présenterait comme anonymes, par exemple s'il s'agit de données agrégées provenant des utilisateurs de son service.

- que le jeu de données ne contienne **pas de données sensibles ou de données d'infraction** (au sens des articles 9 et 10 du RGPD), ou de mener des vérifications supplémentaires pour s'assurer que ce traitement était licite (il s'agirait principalement de s'assurer du recueil d'un consentement explicite des personnes concernées, ou que les données ont été manifestement rendu publiques par ces dernières comme cela est précisé ci-dessous).

**Par exemple :** sur un forum en ligne, un chercheur découvre l'accessibilité d'un jeu de données non anonymes qui contiendrait, selon sa description, les parcours de soin d'une centaine de patients atteints d'une pathologie particulière qui proviendrait d'hôpitaux français. Dans ce cas, le chercheur devrait sérieusement douter de la licéité de la diffusion de ce jeu de données compte tenu de l'encadrement des données sensibles prévu par le RGPD et la loi « informatique et libertés ».

Ces vérifications préalables pourraient utilement figurer dans l'analyse d'impact relative à la protection des données ([AIPD](#)).

Il est à noter que certains manquements commis par le responsable des traitements de constitution et de diffusion d'un jeu de données n'impactent pas directement et irrémédiablement la licéité des traitements mis en œuvre par le réutilisateur, à condition que la réutilisation soit elle-même conforme au RGPD.

**Par exemple :** tel serait notamment le cas de la fourniture de mentions d'informations incomplètes lors de la constitution ou de la diffusion du jeu de données<sup>19</sup>, ou du défaut de documentation adaptée de la conformité de ces traitements (par ailleurs souvent difficile à vérifier sans interaction avec le diffuseur ou l'éditeur du jeu de données).

En tout état de cause, ces vérifications préalables n'exonèrent pas les réutilisateurs d'une analyse complète de la conformité de leurs propres traitements, y compris lorsqu'ils réutilisent des jeux de données dont la constitution et la diffusion ne relèvent pas du droit français ou européen (contrairement à leur réutilisation par une entité établie sur le territoire français ou européen qui est soumise au RGPD).

---

<sup>19</sup> Le manquement à l'obligation d'information fait l'objet d'une contravention (article R. 625-10 du code pénal), alors que le délit de recel ne porte que sur les crimes et délits (article 321-1 du code pénal).

## Les dérogations permettant de traiter des données sensibles

Pour traiter des données sensibles, outre une base légale valide, des conditions supplémentaires doivent être remplies.

Les données sensibles (au sens de l'article 9 du RGPD) sont celles relatives à l'origine ethnique ou prétendument raciale, les opinions politiques, les convictions religieuses ou philosophiques ou l'appartenance syndicale, ainsi que le traitement des données génétiques, des données biométriques<sup>20</sup> 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.

En dehors du cas particulier des recherches dans le domaine de la santé, pour lesquelles des [dispositions spécifiques sont prévues](#) (article 44.3 de la loi Informatique et Libertés), de telles données ne pourront être traitées que si l'une des conditions suivantes est remplie :

- si elles ont été **rendues manifestement publiques par la personne concernée** ;
- si le traitement est nécessaire à la [recherche publique au sens du code de la recherche](#), et que ces utilisations sont rendues nécessaires « pour des motifs d'intérêt public important », après avis motivé et publié de la CNIL qu'il leur appartient donc de solliciter au titre de l'article 44.6 de la loi Informatique et Libertés.  
**La recherche privée, qui ne rentre pas dans les critères définis dans le code de la recherche, ne peut donc pas mobiliser cette exception prévue par la loi « informatique et libertés ».**
- si un texte de droit national ou européen (par exemple un décret en Conseil d'État pris après [avis de la CNIL](#)) autorise ce traitement au titre d'un **motif d'intérêt public important**.
- **ou si les personnes concernées y ont expressément consenti.**

### Comment apprécier si des données ont été manifestement rendues publiques : le cas des réseaux sociaux

[Selon le Comité européen de la protection des données \(CEPD\)](#), les éléments suivants peuvent être pertinents pour aider à évaluer si les données ont été manifestement rendues publiques par la personne concernée :

- le paramétrage par défaut de la plate-forme de médias sociaux (c'est-à-dire si la personne concernée a effectué une action pour changer ces paramètres privés par défaut en paramètres publics) ; ou
- la nature de la plate-forme de médias sociaux (c'est-à-dire si cette plate-forme a vocation à connecter des connaissances proches avec la personne concernée ou à nouer des relations intimes (comme les sites de rencontres), ou bien s'il est destiné à fournir un plus large éventail de relations interpersonnelles, telles que les relations professionnelles, ou le microblogging, le partage de médias, les réseaux sociaux plateformes de partage d'avis en ligne, etc. ; ou
- l'accessibilité de la page où les données sensibles sont publiées (c'est-à-dire si les informations sont accessibles au public ou si, par exemple, la création d'un compte est nécessaire avant d'accéder aux informations) ; ou
- la visibilité de l'information sur le caractère public des contenus publiés par les personnes concernées (c'est-à-dire si, par exemple, il existe une bannière continue sur la page ou si le bouton de publication informe la personne concernée que les informations seront rendues publiques) ; ou
- si la personne concernée a elle-même publié les données sensibles, ou si, à l'inverse, les données ont été publiées par un tiers (par exemple, une photo publiée par un ami qui révèle des données sensibles) ou déduites.

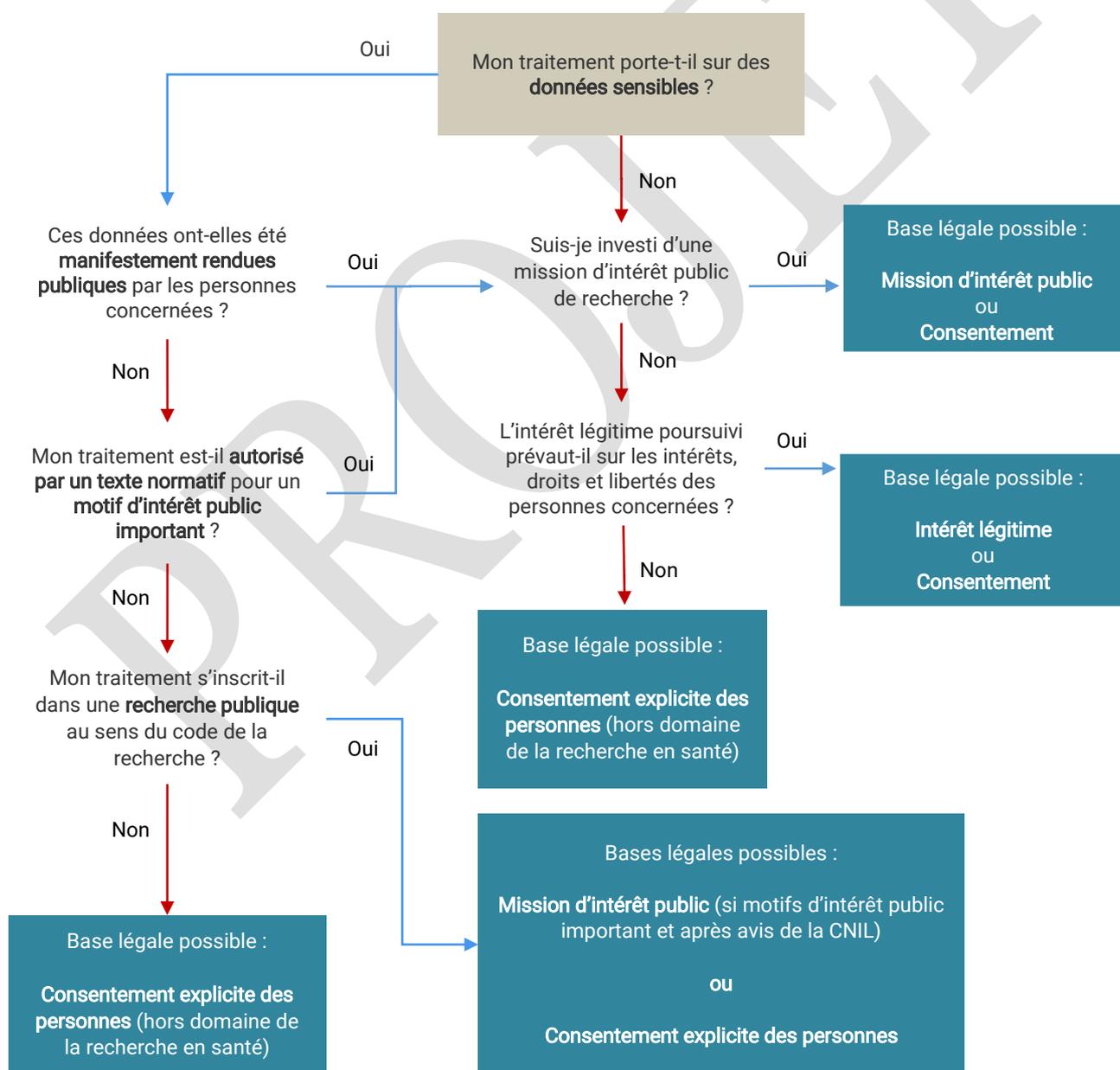
<sup>20</sup> La CNIL rappelle qu'une photographie ou un vidéogramme ne constituent pas en eux-mêmes des données biométriques. En revanche, constituerait un tel traitement la création de gabarits biométriques (c'est-à-dire de représentations numériques des caractéristiques physiques des personnes comme des visages), à partir de photographies collectées sur Internet pour fournir un moteur de recherche permettant d'identifier un individu à partir d'une image (voir la délibération de la formation restreinte de la CNIL SAN-2022-019 du 17 octobre 2022)

Le CEPD note que **la présence d'un seul élément peut ne pas toujours suffire à établir que les données ont été « manifestement » rendues publiques par la personne concernée**. En pratique, une combinaison de ces éléments ou d'autres éléments peut devoir être prise en compte pour pouvoir démontrer que la personne concernée a manifesté clairement l'intention de rendre les données publiques.

**Par ailleurs, les données manifestement rendues publiques ne se limitent pas aux données accessibles sur les réseaux sociaux.**

**Par exemple :** des chercheurs au sein d'une université publique entreprennent une recherche dans le domaine des sciences politiques. A cet égard, ils traitent des données publiquement disponibles sur des sites institutionnels afin d'étudier les conditions d'exécution par les élus de leur mandat politique (taux de présence, nombre de rapports, sens de leurs votes sur différentes thématiques, etc.). Dans la mesure où le responsable de traitement est une université publique et que les orientations politiques sont manifestement rendues publiques par les personnes concernées dans le cadre de leur activité, le traitement de ces données sensibles est licite.

### Schéma récapitulatif des conditions de traitement



**À noter :** en dehors des données sensibles visées par l'article 9 du RGPD, [d'autres catégories de données font l'objet de mesures de protection spécifiques](#) comme :

- **les données relatives à des infractions et condamnations**, qui ne peuvent être légalement traitées que par certaines catégories d'acteurs (autorités publiques et les personnes morales gérant un service public, agissant dans le cadre de leurs attributions légales, associations d'aide aux victimes conventionnées par le ministère de la justice, etc.) ;
- **ou encore celles relatives à des mineurs**, massivement présents en ligne et auxquels le cadre juridique de la protection des données accorde une attention particulière, s'agissant notamment des conditions de prise en compte de [leurs droits](#).

## Les mesures de minimisation des données et la proportionnalité des réutilisations

S'agissant des traitements à des fins de recherche scientifique, l'article 89 du RGPD insiste sur l'importance de la mise en œuvre de mesures permettant d'assurer le principe de minimisation des données.

### Ne collecter que les données nécessaires à l'objectif poursuivi

Tout d'abord, les réutilisateurs de données doivent **mener une réflexion sur le champ des données « strictement nécessaires » à leur recherche en veillant à ne pas collecter pas davantage de données que nécessaires.**

Lorsque la collecte de données n'est pas réalisée manuellement mais au moyen de procédés automatisés, cela implique :

- **de privilégier autant que possible l'utilisation d'interfaces de programmation applicative (« API ») permettant de cibler les données recueillies** (cf. recommandation de la CNIL n° 2023-050 du 25 mai 2023),
- **de ne recourir à des techniques de moissonnage de données (*scraping*) que de manière ciblée et pertinente**, c'est-à-dire :
  - en définissant **des critères précis et pertinents de collecte** afin d'opérer une distinction quant à la nature des données collectées ;
  - **en limitant le risque de recueil de données sensibles**, lors de la sélection des mots clés ou catégories de contenus, par exemple ;
- **de supprimer toute donnée non pertinente, en particulier s'il s'agit d'une donnée sensible au sens de la réglementation, dès qu'elle est identifiée comme telle.** Cela est valable quelle que soit le mode de collecte, qu'il s'agisse du téléchargement d'un fichier librement accessible, du recours à des API ou à des techniques de moissonnage de données, dès lors qu'un tri exhaustif n'étant pas toujours possible de manière automatisée.

**À noter :** si le recours à des techniques de moissonnage n'est pas en soit incompatible avec les exigences du RGPD, il est susceptible de compromettre le bon fonctionnement des sites « aspirés » (par exemple en occasionnant des dénis de service si trop de requêtes sont adressées simultanément), ou d'être interdits par d'autres réglementations (par exemple par des conditions générales d'utilisation qui s'appuieraient sur le droit des producteurs de bases de données).

### L'anonymisation ou la pseudonymisation des données collectées

**Il est essentiel d'anonymiser ou pseudonymiser les données quand leur exploitation sous une forme identifiante (directement ou non) n'est pas nécessaire.**

Toutes les fois où un [traitement d'anonymisation ou de pseudonymisation](#) des données collectées en ligne peut être mis en œuvre, il convient d'y recourir. Les enjeux et avantages de ces techniques sont présentées dans une [fiche pratique de la CNIL](#).

De tels traitements, qui excluent ou limitent le risque d'identification des personnes concernées, **doivent être envisagés aussi bien au stade de la collecte des données qu'au stade de leur exploitation.**

**Par exemple :**

- **L'anonymisation** peut être très pertinente dans le cadre d'une étude quantitative, lorsque des résultats agrégés suffisent à produire des analyses pertinentes (par exemple : le nombre de personnes s'étant publiquement prononcé sur tel sujet sur un réseau social de type microblogging).
- Lorsque l'anonymisation des données nuirait au déroulement de la recherche, par exemple dans le cadre d'une approche qualitative qui supposerait l'analyse approfondie de contenus publiés par des personnes concernées, la **pseudonymisation** des données et leur conservation sous une forme non directement identifiante, pendant la durée de l'analyse des données collectées peuvent être envisagée (ce fut notamment le cas d'un projet de traitement ayant pour finalité une recherche sur les impacts pour la vie privée des publications d'informations librement accessibles sur les réseaux sociaux, cf. [délibération de la CNIL n° 2018-151](#) du 3 mai 2018).
- A l'inverse, le traitement de données brutes directement identifiantes (telles que des profils publiés sur un réseau social), pendant une longue durée nécessitera une analyse approfondie de la base légale, en particulier s'il s'agit de l'intérêt légitime, dès lors que les risques pour les intérêts, droits et libertés seront plus importants en l'absence de toute mesure de pseudonymisation.

Par ailleurs, si les finalités pour lesquelles des données personnelles sont traitées n'imposent pas ou plus au réutilisateur d'identifier les personnes, celui-ci n'est pas tenu de conserver, d'obtenir ou de traiter des informations supplémentaires pour les identifier à la seule fin de respecter le présent règlement, par exemple pour répondre à des demandes d'exercice des droits. En revanche, dans un tel cas, les personnes en sont si possible informées, et peuvent fournir des informations complémentaires permettant de les identifier afin d'exercer leurs droits (article 11 du RGPD).

## Documentation de la conformité

**La démonstration, documentée par le responsable de traitement du bon respect des principes précédemment évoqués est essentielle.**

Cette évaluation sera d'ailleurs nécessaire dans le cadre de la réalisation de l'[analyse d'impact relative à la protection des données](#) (AIPD), **obligatoire avant la mise en œuvre du traitement si ce dernier est susceptible d'engendrer un risque élevé pour les personnes.**

Ce sera souvent le cas de l'utilisation d'outils de moissonnage de données, mais aussi plus généralement de toute collecte à grande échelle portant sur des personnes vulnérables, ou encore de croisement de données ou de profilage portant sur des données sensibles ou à caractère hautement personnel.

Cette analyse permet de documenter les mesures lors de la collecte et la constitution d'un jeu de données (par exemple, en détaillant les critères de sélection des données, la mise en œuvre de leur anonymisation ou de leur pseudonymisation, etc.). L'AIPD permettra aussi de s'assurer de la pertinence du choix et du [respect d'une durée de conservation limitée des données, de la confidentialité et la sécurité des données](#). Il s'agit notamment de prendre en compte les risques que poseraient d'éventuelles réutilisations de ces données.

## L'information des personnes concernées

Le RGPD prévoit que tout responsable de traitement doit fournir des [mentions d'information](#) aux personnes concernées sur le traitement de leurs données. Cette information est en principe directement fournie à chaque personne dont les données sont traitées.

**Lorsque les données ne sont pas collectées auprès de la personne concernée comme c'est le cas de la réutilisation des données publiquement accessibles, le RGPD prévoit que la fourniture d'une information publique peut s'avérer suffisante** (par exemple sur le site du responsable du traitement), notamment dans les deux cas suivants :

- Plus précisément, **l'article 14.5.b du RGPD prévoit qu'une information générale peut suffire lorsqu'une information individuelle se révèle impossible ou exigerait des efforts disproportionnés, en particulier s'agissant des traitements à des fins de recherche scientifique.**

En pratique, le caractère proportionné s'apprécie en fonction de l'atteinte portée à la vie privée des personnes dont les données sont traitées et de la difficulté et le coût d'une information individuelle :

- Lorsqu'une information par voie électronique est possible, elle est généralement requise ;

- Lorsque les données disponibles sont des données postales, une analyse au cas par cas est nécessaire, en prenant en compte notamment le coût de l'information, la faiblesse ou l'importance des risques que le traitement peut faire encourir aux personnes, le fait que les personnes peuvent ou non raisonnablement s'attendre au traitement de leurs données ;
- Beaucoup de jeux de données pseudonymisées mis à disposition de chercheurs sur Internet ne contiennent pas de données de contact, conformément au principe de minimisation décrit ci-dessus. Lorsque les données de contact ne sont pas ou plus disponibles, il est généralement plus facile de démontrer qu'une information individuelle exigerait des efforts disproportionnés. Une analyse de proportionnalité est cependant nécessaire pour s'assurer que le défaut d'information individuelle ne remet pas en cause la licéité du traitement.
- La fourniture d'une information générale est aussi envisageable lorsqu'une information individuelle est susceptible de rendre impossible ou de compromettre gravement la réalisation des objectifs dudit traitement de chaque personne concernée.

**Par exemple :** ce pourrait être le cas des travaux de recherche scientifique impliquant l'analyse de données diffusées sur les réseaux sociaux, si l'information des personnes concernées risque de les conduire à modifier leur comportement en ligne et, ainsi, de biaiser les travaux en question.

Cette transparence est essentielle pour permettre aux personnes concernées d'exercer leurs droits sur le traitement de leurs données, tels que le droit d'accès, le droit d'opposition, le droit à l'effacement ou le droit à la rectification de leurs données auprès du réutilisateur. Il est à noter que [des dérogations sont permises lorsqu'elles sont nécessaires à l'objectif de recherche](#).

Par ailleurs, la fourniture de mentions d'information complètes (notamment sur la source des données) permet également aux personnes d'exercer leurs droits auprès du diffuseur de leurs données.

## Fiche cas d'usage n°5 : le moissonnage de données publiquement accessibles par des autorités publiques dans le cadre de leurs missions

Certaines autorités publiques ont recours à des techniques de moissonnage de données publiquement accessibles dans le cadre de leurs prérogatives de puissance publique.

Pour ce faire, elles extraient des informations de sites web, tels que les réseaux sociaux, au moyen de procédés informatiques appelés outils de moissonnage, ou de « *web scraping* » en anglais.

Compte tenu du risque de surveillance généralisé inhérent à de telles techniques, et plus généralement des enjeux pour les droits et libertés des citoyens qui ne maîtrisent pas les réutilisations de leurs données accessibles en ligne, ces traitements de données personnelles doivent comprendre un certain nombre de garanties.

### La nécessité d'un encadrement juridique

Ni le fait que des données soient publiquement accessibles, ni la création volontaire de profils sur les plateformes en ligne n'emporte, par principe, la possibilité de leur moissonnage ainsi que de leurs réutilisations. Les acteurs qui souhaitent les exploiter ne sont pas exonérés de l'**obligation de collecter ces données de manière licite et loyale**.

**Pour qu'une autorité publique puisse mettre en œuvre de tels traitements dans le cadre de ses missions, il est le plus souvent nécessaire que le traitement soit autorisé par un texte de droit français ou européen.** Par exemple, cela sera notamment rendu nécessaire par l'[article 31](#) de la loi Informatique et Libertés qui exige une autorisation par arrêté ou par décret (pris après avis motivé et publié de la CNIL) pour les traitements de détection d'infraction mis en œuvre pour le compte de l'Etat par les autorités compétentes (c'était notamment le cas du [traitement mis en œuvre par VIGINUM ayant recours à des outils de moissonnage de données visant à détecter et caractériser les opérations d'ingérence numérique étrangères aux fins de manipulation de l'information sur les plateformes en ligne](#)<sup>21</sup>).

**Cet encadrement spécifique devra faire apparaître de manière explicite les finalités des traitements.**

Lorsque le traitement affecte les droits et libertés fondamentaux dont bénéficie le citoyen au sens de l'article 34 de la constitution (telles que la liberté d'expression et la liberté d'opinion), **seule une loi pourra autoriser sa mise en œuvre en définissant son périmètre et des garanties suffisantes.**

À cet égard, la **limitation de la collecte aux données librement accessibles** (c'est-à-dire aux contenus accessibles à tout utilisateur non inscrit sur le site en question et sans création d'un compte) et **manifestement rendues publiques par la personne concernée** (ce qui exclut notamment les commentaires ou interactions de tiers portant sur la personne concernée) est une garantie essentielle. Elle a notamment été déterminante pour apprécier la proportionnalité de [l'expérimentation de l'exploitation de données rendues publiques sur les sites internet des opérateurs de plateformes en ligne par les administrations fiscale et douanière](#).

Il est possible de citer d'autres garanties ayant déjà été prévues par des textes, telles que :

- la limitation des services sur lesquels le recours aux outils de moissonnage de données est possible (par exemples aux seuls sites internet des opérateurs de plateforme de mise en relation de plusieurs parties en vue de la vente d'un bien, de la fourniture d'un service ou de l'échange ou du partage d'un contenu, d'un bien ou d'un service, dépassant un seuil important de visiteurs) ;
- l'interdiction de prendre une décision entièrement automatisée sur la base d'un traitement de moissonnage de données ;
- l'interdiction de certains procédés, tels que le recours à un système de reconnaissance faciale ou d'identification vocale ;
- l'existence d'un contrôle externe, effectif et adapté aux caractéristiques du traitement envisagé, ou encore
- l'exigence d'une transparence accrue, notamment à travers la production de rapports annuels publics.

<sup>21</sup> Ce traitement est prévu par le [décret n° 2021-1587 du 7 décembre 2021 portant autorisation d'un traitement automatisé de données à caractère personnel dans le but d'identifier les ingérences numériques étrangères](#).

## La minimisation de la collecte

Le respect du principe de minimisation est fondamental compte tenu du volume important de données et plus particulièrement, d'informations non nécessaires ou non pertinentes susceptibles d'être moissonnées.

Pour cette raison, **la CNIL recommande d'utiliser une interface de partage de données (API) ou toute autre fonctionnalité de mise à disposition par le site ou la plateforme lorsque cela est possible, plutôt que de recourir à des outils de moissonnage de données.**

Lorsqu'une autorité décide de recourir à des outils de moissonnage de données, elle doit veiller à **définir, en amont de la mise en œuvre du traitement, des critères précis de collecte afin d'opérer une distinction quant à la nature des données collectées.** Cela suppose en amont d'élaborer des indicateurs et critères de pertinence, permettant d'orienter et cibler la collecte.

Il convient ensuite **de s'assurer de la suppression des données non pertinentes immédiatement après leur collecte, notamment lorsqu'elles sont sensibles** (par exemple des informations sur la santé, la religion ou l'orientation sexuelle des personnes).

Un tri exhaustif n'étant pas toujours possible de manière automatisée, l'autorité publique doit **veiller à supprimer toute donnée non pertinente dès qu'elle est identifiée comme telle** dans le cadre de l'exploitation des données ainsi collectées.

Par exemple, dans [le cas de VIGINUM](#), l'administration doit s'assurer que seules les informations en lien avec les thématiques et acteurs présumés étrangers d'intérêt soient collectées, notamment en utilisant un procédé automatisé de suppression des données non pertinentes (afin de garantir que les données initialement collectées ne seront pas accessibles aux personnes habilitées avant l'opération de suppression automatisée des données non pertinentes).

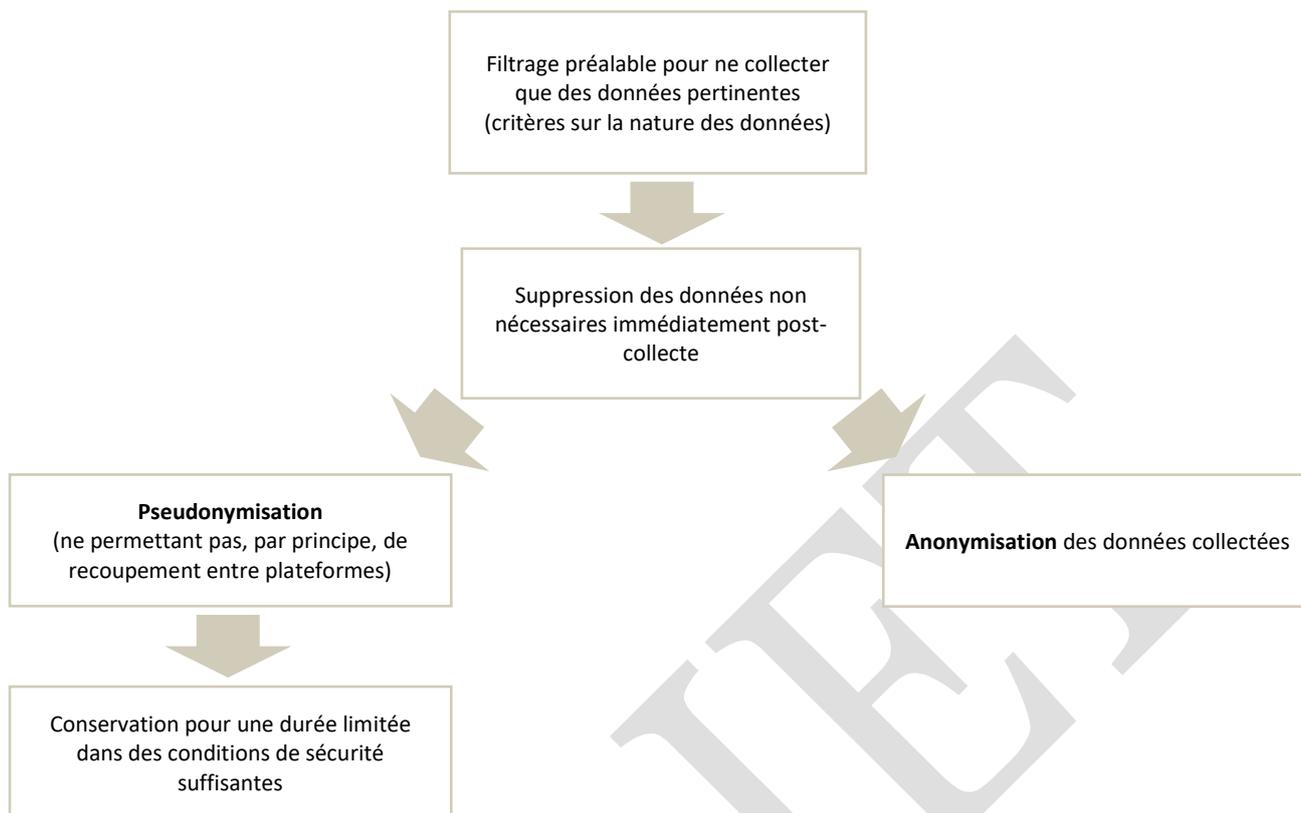
La CNIL recommande qu'une anonymisation ou une pseudonymisation des données soit effectuée juste après la collecte des données et que le recoupement, sur plusieurs plateformes en ligne, à partir des identifiants collectés ne soit possible qu'après avoir conduit une analyse documentée dans l'AIPD afin de mettre en balance l'intérêt de ce recoupement et les risques pour les droits et libertés des personnes concernées.

## La sécurité et la conservation des données

**Les données moissonnées doivent être correctement sécurisées au regard des risques encourus.**

Par exemple, dans le cas des [expérimentations mises en œuvre par le PEReN \(pôle d'expertise de la régulation numérique de traitements\)](#), une habilitation individualisée des accès aux données et une journalisation de ces accès est prévue, La CNIL a recommandé en complément la mise en place d'un mécanisme de double authentification.

**Ces données ne doivent pas être conservées plus longtemps que nécessaire au regard de la finalité de leur collecte.** La mise en œuvre de ces traitements ne doit pas non plus engendrer une collecte permanente de données sur les plateformes concernées. La CNIL recommande de définir au cas par cas la durée de la collecte et la période de récupération des données (profondeur historique), qui devra également être documentée. À titre d'exemple, le [texte encadrant le traitement de VIGINUM](#) a précisé la période au cours de laquelle les données peuvent être collectées, soit une période maximale de sept jours.



Lorsque l'autorité a recours aux services d'un sous-traitant pour réaliser de telles opérations, un contrat doit être conclu, contenant toutes les mentions prévues par l'article 28 du RGPD (telles que celles relatives aux instructions documentées et aux mesures de sécurité). Par ailleurs, le responsable de traitement doit s'assurer que son sous-traitant présente des garanties suffisantes pour la conformité des traitements (pour en savoir plus : voir le [guide du sous-traitant](#)).

## L'information des personnes et l'exercice de leurs droits

En principe, l'autorité doit [informer](#) les personnes concernées du traitement qu'elle met en œuvre.

**Une information générale peut s'avérer suffisante, par exemple sur le site internet de l'autorité en question lorsque la fourniture d'une information individuelle exigerait des efforts disproportionnés pour l'autorité ou qu'elle est écartée par le texte autorisant le traitement.**

Cette information permet notamment aux personnes concernées d'exercer leurs droits sur leurs données auprès de l'autorité. Il peut s'agir du droit d'accès, du droit à la rectification ou à l'effacement des données.

Lorsque l'autorité reçoit une demande d'exercice de l'un de ces droits après avoir procédé à l'anonymisation ou la pseudonymisation des données, il est possible pour l'autorité de démontrer qu'elle n'est pas en mesure d'identifier la personne concernée. Cependant, les personnes devraient toujours pouvoir fournir des informations complémentaires permettant de les identifier.

**Enfin, le droit d'opposition devra parfois être envisagé pour certains traitements. L'effectivité de ce droit d'opposition pourra notamment être garantie au moyen d'une liste repoussoir, y compris en amont du traitement.**

S'agissant de la liste repoussoir mise en œuvre par le PEReN dans le cadre de ses expérimentations, [la CNIL a par exemple recommandé](#) de ne conserver que des empreintes cryptographiques des identifiants des personnes concernées (en supprimant tout lien entre les identifiants d'une même personne) et de supprimer les données brutes immédiatement. Elle a aussi recommandé de permettre aux personnes concernées de choisir la durée de prise en compte de leur opposition.

## La réalisation d'une AIPD

**La démonstration, documentée par le responsable de traitement du bon respect des principes précédemment évoqués est essentielle avant tout déploiement de tels dispositifs.**

Cette évaluation sera d'ailleurs nécessaire dans le cadre de la réalisation de l'[analyse d'impact relative à la protection des données](#) (AIPD), **obligatoire avant la mise en œuvre du traitement si ce dernier est susceptible d'engendrer un risque élevé pour les personnes**. Ce sera souvent le cas de l'utilisation d'outils de moissonnage de données, dès lors qu'il s'agit d'un usage innovant conduisant à une collecte à grande échelle, portant souvent sur des données sensibles ou à caractère hautement personnel et pouvant conduire à une décision automatisée avec effet légal ou similaire.

Si le traitement est mis en œuvre par une autorité compétente à des fins de prévention et de détection des infractions pénales, d'enquêtes et de poursuites en la matière, pour le compte de l'État, cette analyse d'impact doit même être adressée à la CNIL dans le cadre d'une demande d'avis.

PROJET