Recommandations

Diffuseurs de données ouvertes (*open data*)

Publié le 12/06/2024

CNIL.
COMMISSION NATIONALE
INFORMATIQUE & LIBERTÉS

Table des matières

Introduction

À qui s'adressent ces recommandations ?

Ces recommandations s'adressent à toute personne – physique ou morale, publique ou privée et ci-après dénommée « *diffuseur de données* » – qui met en ligne à disposition du public, dans un format ouvert, aisément réutilisable et exploitable par machine, des données personnelles (données se rapportant à des personnes physiques identifiées ou potentiellement identifiables).

Cette **opération d'ouverture des données, souvent désignée sous le terme anglais de mise en « *open data* »**, est essentiellement réalisée par des personnes publiques, 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.

Quelle est leur vocation ?

Ces recommandations complètent, tout en élargissant le champ d'étude, le guide co-édité en 2019 avec la CADA sur l'ouverture et la réutilisation des données publiques.

Au moyen de « fiches principes » à caractère pratique et opérationnel, elles fournissent une grille d'analyse générale permettant à tout diffuseur de données personnelles de cheminer le plus rapidement et efficacement possible sur les questions Informatique et Libertés structurantes, ainsi que sur les éléments de réponse pertinents.

À noter

Ces fiches, adoptées à la suite d'une consultation publique, constituent un cadre qui permet d'accompagner les organismes dans leur mise en conformité. Elles rappellent les obligations posées par la règlementation et formulent des recommandations pour s'y conformer. **Ces recommandations ne sont pas contraignantes : les responsables de traitement peuvent s'en écarter, à condition de pouvoir justifier leur choix et sous leur responsabilité**. Certaines recommandations sont également formulées à titre de bonnes pratiques et permettent d'aller plus loin que le respect de la règlementation.

Fiche n°1 : Quelle qualification juridique pour les diffuseurs 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.

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.**

S'agissant de la diffusion publique d'une base de données, le responsable de traitement, qu'il soit public ou privé, 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.

Exemple

L'INPI est le responsable du traitement de mise à disposition du public des données figurant dans le registre national des entreprises, tel que prévu à l'article L. 123-52 du code de commerce.

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.

Exemple

Plusieurs collectivités territoriales décident de développer ensemble une plateforme commune et locale d'ouverture de leurs données, destinée à favoriser, par la mise à disposition d'un point d'accès unique aux informations publiques qu'elles détiennent, la réutilisation de celles-ci et la création corrélative de nouvelles offres de services sur le territoire.

Focus

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 pour qualifier une responsabilité conjointe si les moyens du traitement sont définis distinctement.

À noter, d'une manière générale, que le recours à des licences/CGU est une pratique à encourager de la part des diffuseurs. Si 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 (voir à ce sujet la dernière page de la fiche n°6).

En cas de responsabilité conjointe, chaque responsable de traitement doit s'assurer de la licéité du traitement, notamment en définissant, dans le cadre d'un « accord » et de manière opérationnelle et transparente, leurs obligations respectives (article 26 du RGPD).

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 par exemple, sur « qui fait quoi » pour que soient respectées les règles relatives à la protection des données personnelles. Ainsi, cet accord a notamment vocation à préciser les modalités d'exercice et de prise en compte des droits « informatique et libertés ».

À cet égard, il est à noter que si leurs demandes ont vocation à être prises en charge par le ou les responsables qui est/sont désigné(s), les personnes concernées gardent la liberté d'exercer leurs droits auprès de n'importe lequel des responsables de traitement conjoints.

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[1]) et ne doit les traiter que sur instructions documentées du responsable de traitement.

Exemple

Peut être qualifié de sous-traitant le fournisseur d'une plateforme de partage de données en ligne, tel 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.

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**).

[1] « Sous-traitants : la réutilisation de données confiées par un responsable de traitement », publié le 11 janvier 2022 sur cnil.fr

Fiche 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 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.).

En matière de diffusion publique d'une base de données, les principales bases légales envisageables[2] 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

Questions à se poser pour identifier sa base légale

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. *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.

[2] Aucune des bases légales prévues à l'article 6 du RGPD ne doit être proscrite par principe.

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).

Exemple : la publication par une commune, dans un format ouvert, de l'organigramme de ses services pouvant comporter l'identité de ses responsables.

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 des

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 n°5 et 6 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 ;

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 ces recommandations).

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.

> **À noter**
>
> 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.

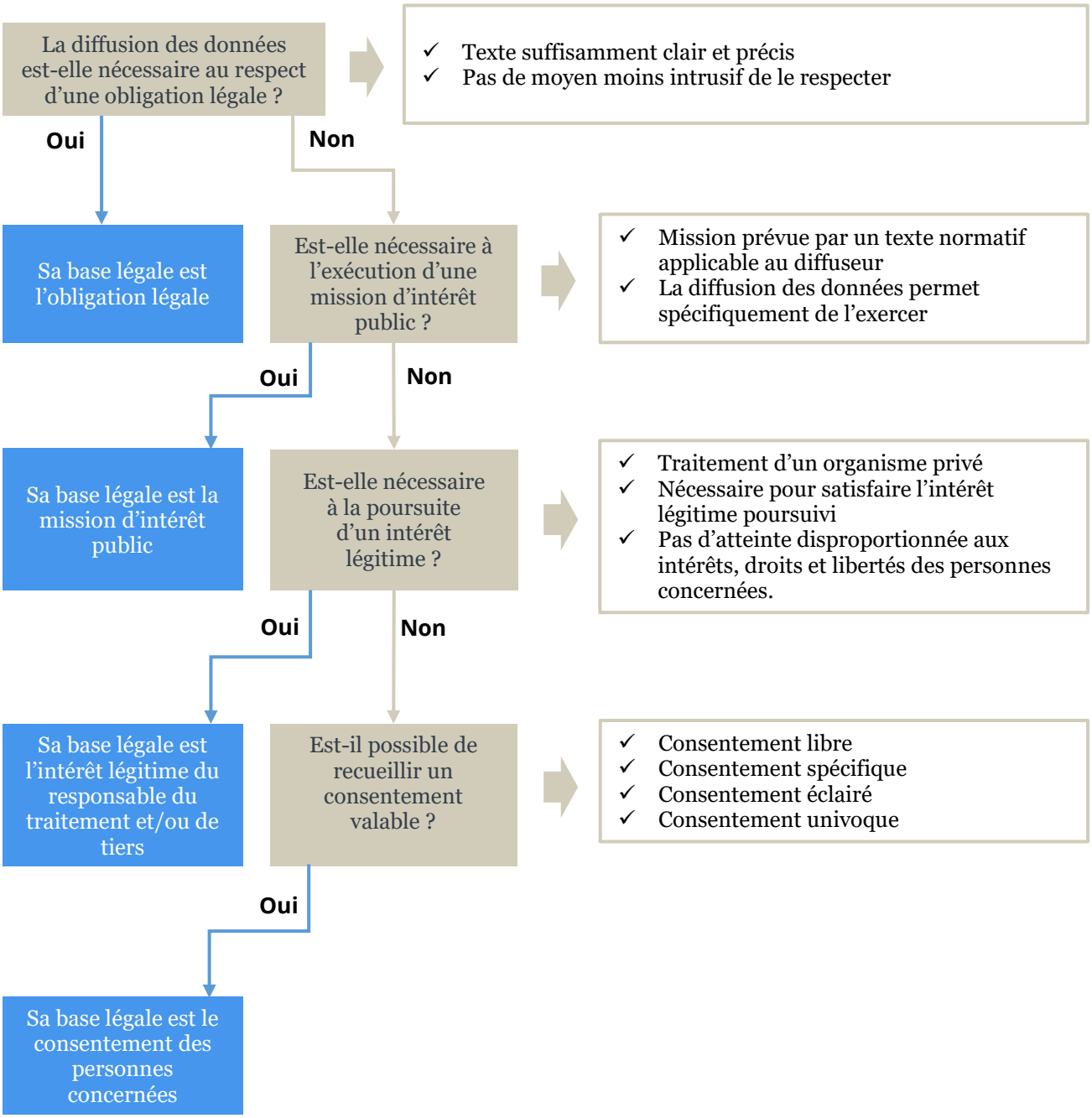
- **le consentement doit être univoque** : c'est-à-dire donné par une déclaration ou tout autre acte positif clair, sans ambiguïté ; la jurisprudence refuse les cases précochées pour l'expression du consentement[3].

Illustration

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.

[3] Cour de justice de l'Union européenne, grande chambre, 1er octobre 2019, Planet49 GmbH, C-673/17, curia.europa.eu

Schéma récapitulatif : identifier la base légale du traitement



Fiche n°3 : Comment informer les personnes concernées

Le RGPD impose d'informer les personnes concernées du traitement de leurs données et notamment de leur diffusion 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, 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 ?

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** (voir la fiche n°2) **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) ; doit également être précisée, si des données « sensibles » visées à l'article 9 du RGPD sont concernées, l'exception prévue par cet article (et, le cas échéant, du droit de l'Union ou de l'État membre en vertu duquel les données sont traitées) permettant de les diffuser (voir la fiche n°5) ;

- 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** (voir la fiche n°4), 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) ;

- 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).

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 par exemple 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 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

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).

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

Dans certains cas prévus par les textes, et en particulier dans ceux prévus ci-dessous, les responsables de traitements peuvent ne pas procéder à une information individuelle des personnes concernées.

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

Exemple

Cette hypothèse correspond au cas où les personnes auraient été préalablement informées, par l'organisme ayant transmis leurs données à celui qui les diffusera, que les informations les concernant feraient l'objet d'un partage sur internet par ce destinataire dans telles et telles conditions.

Attention

Pour que cette exception soit mobilisable, il faudra que l'ensemble des éléments d'information prévus à l'article 14 du RGPD (ou article 13, en cas de collecte directe des données) aient déjà été portés à la connaissance des personnes concernées.

Cas n°2 : les données n'ont pas été directement collectées auprès des personnes concernées et leur information se révèle impossible ou exigerait des efforts disproportionnés

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 diffuseur 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** ;

> **Exemple**
>
> Le fait, pour une administration, de diffuser des données publiques dans le respect des dispositions du code des relations entre le public et l'administration (limitant les possibilités de publication sous une forme non anonymisée et imposant, sauf exceptions légales, l'occultation des mentions dont la divulgation génèrerait une telle atteinte) engendre, *a priori*, moins de risques pour les personnes que le fait, pour un organisme privé, de partager sur internet des données personnelles sous une forme nominative ou pseudonymisée avec un risque fort de réidentification

- **d'autre part, la difficulté et le coût d'une information individuelle**.
    - Lorsqu'une information par courriel est possible, elle est généralement requise.
    - Dans les autres hypothèses, 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 celles-ci 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, ou que la diffusion est permise voire imposée par des dispositions légales, que le responsable de traitement est systématiquement dispensé de l'obligation d'informer. Il lui faut prendre en compte les autres paramètres et notamment l'intrusivité du traitement.

Le diffuseur 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'imposent pas (p. ex. : la diffusion obligatoire en *open data*, par la direction générale des finances publiques, de la base de données « demandes de valeurs foncières / DVF », qui comprend les transactions immobilières des cinq dernières années, à l'adresse mais 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).

À retenir

Tout diffuseur 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 : réalisation d'une AIPD, anonymisation ou pseudonymisation des données, suppression ou occultations de certaines d'entre elles, 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 communication lorsque les personnes concernées sont nombreuses (publications sur le site web où sont diffusées les données, sur ses comptes de réseaux sociaux, dans un journal local, etc.).

Cas n°3 : 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 à des fins journalistiques.

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.

À noter

D'autres cas peuvent être prévus par des dispositions légales spécifiques, constituant une mesure nécessaire et proportionnée pour garantir des objectifs d'intérêt public particulièrement importants.

Les textes instaurant les traitements concernés et autorisant une telle dérogation doivent contenir des dispositions **particulières**, 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

Schéma récapitulatif : informer les personnes concernées

Puis-je être dispensé de la délivrance d'une information individuelle ?

La personne a déjà obtenu les informations ?

→ **L'ensemble des éléments d'information prévus à l'article 14 du RGPD** (ou article 13, en cas de collecte directe des données) doivent avoir déjà été portés à la connaissance des personnes concernées.

Ou, l'information se révèle impossible ou exigerait des efforts disproportionnés ?

→ **Pas le cas en principe si adresses mail à disposition ou facilement et légalement accessibles.**
Si seules des adresses postales seraient potentiellement exploitables, mettre en rapport :

- l'atteinte portée à la vie privée, les risques pour les personnes concernées et
  leurs attentes raisonnables (ex. : nature des données, contexte de publication),
- la difficulté d'une information individuelle (ex. : données pseudonymisées),
- le coût de sa délivrance.

Ou, le traitement est mis en œuvre aux fins d'expression universitaire, artistique ou littéraire, ou de journalisme

→ **Le texte doit obligatoirement contenir certains éléments** (article 23 du RGPD) : finalités et responsable du traitement, catégories de données traitées et durée de conservation, risques pour les droits et libertés des personnes concernées et garanties destinées à prévenir les abus.

Ou, le traitement est prévu par un texte écartant ce droit ?

→
- ✓ La **dérogation** doit être **nécessaire** pour concilier la protection des données personnelles et la liberté d'expression et d'information.
- ✓ **L'activité de journaliste** doit être exercée **à titre professionnel,** au sens de l'article L7111-3 du code du travail.

SI OUI : Prévoir une information publique générale et complète sur l'identité et les coordonnées du responsable de traitement, les finalités et base légale de celui-ci, les catégories de données traitées, sources, destinataires et durée de conservation, et les droits des personnes concernées

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

Les personnes dont les données sont partagées sur Internet 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), 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 n°2) **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 / Base légale	Accès	Rectification	Opposition	Effacement	Limitation	Portabilité
Consentement	✓	✓	✗ (1)	✓	✓	✓
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) à noter que les personnes concernées devraient toutefois pouvoir obtenir l'effacement de leurs données dans certains cas, en particulier si leur traitement est illicite ou n'est pas/plus strictement nécessaire.

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

- **le droit d'accès : les personnes peuvent demander à l'organisme une copie des données les concernant et solliciter certaines informations relatives aux caractéristiques du traitement ;**

Exemple : une personne dont des données se trouveraient mises à disposition en *open data* par un organisme ne les ayant pas directement collectées auprès d'elle pourra demander à cet organisme diffuseur toute information quant à la source des données la concernant.

- **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 avocat ou médecin, dont le nom serait mal orthographié dans un annuaire officiel de la profession diffusé en ligne, pourrait demander sa correction au responsable du traitement en lui présentant sa carte d'identité.

- **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 : une personne ayant consenti à la mise à disposition, en *open data,* des données qu'elle a fournies à une entreprise, notamment dans une logique d'altruisme, peut exiger quand elle le souhaite le retrait des informations la concernant de la base de données partagées.

- **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, les personnes peuvent s'opposer à tout moment, et pour des raisons tenant à leur situation particulière[4], 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 : un agent public ou usager du service public, dont les données sont susceptibles, sans que cela ne constitue une obligation légale, d'être diffusées en *open data* en application de l'article D312-1-3 du code des relations entre le public et l'administration, peut demander à l'organisme de ne pas procéder à ce traitement en invoquant les risques (ex. : harcèlement) s'attachant à l'exposition sur internet de ses données compte tenu de circonstances particulières liées à sa situation personnelle.

- **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, ou encore quand les personnes concernées 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 ;**

  Exemple : des personnalités publiques, dont les données clients seraient diffusées, nominativement et sans leur consentement préalable, par une entreprise privée à des fins de réalisation d'études sur leurs habitudes de consommation pourraient exiger l'effacement des informations les concernant en raison de l'absence de base légale fondant le traitement.

- **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) ;**

---

[4] Le Conseil d'Etat a souligné dans un arrêt du 18 mars 2019 (n° 406313) qu'une personne se prévalant de son droit d'opposition ne pouvait « *se borner à invoquer des craintes d'ordre général concernant notamment la sécurité du fonctionnement de la base, sans faire état de considérations qui lui seraient propres* ».

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, sous une forme directement exploitable 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 : toute personne ayant activement consenti, auprès d'une entreprise, à la mise à disposition en *open data* de données la concernant peut 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

- Permettre aux personnes concernées de s'opposer « d'emblée », au stade de leur information sur la mise en œuvre du traitement de diffusion, à certaines réutilisations de leurs données et communiquer sur ces oppositions auprès des éventuels réutilisateurs (p. ex. : publication d'un fichier d'opposition à la prospection commerciale, ou encore d'un « étiquetage » des informations en cause directement dans la base de données) ;
- Fournir, sur la plateforme de mise à disposition des données, 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 concernées de les exercer à partir de leur espace personnel.

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.

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[5], 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[6] :

- **au respect d'une obligation légale ou à l'exécution d'une mission d'intérêt public ;**

[5] Par exemple, l'arrêté du 16 mars 2021 relatif au traitement automatisé, par le ministère de la transition écologique, des données d'urbanisme énumérées à l'article R. 423-76 du code de l'urbanisme (traitement « SITADEL »), exclut la possibilité pour les personnes concernées de s'opposer à la diffusion au public du numéro d'enregistrement de leur demande d'autorisation d'urbanisme, ainsi que du lieu des travaux (adresse et référence cadastrale).

[6] La condition de nécessité doit être strictement appréciée.

Exemple

Un délégué à la protection des données ne peut demander l'effacement de ses coordonnées publiques qui sont diffusées en open data par la CNIL conformément à l'article 83 du décret d'application de la loi Informatique et Libertés.

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

Exemple

Dans certains cas, un journaliste peut refuser de supprimer des données ouvertes figurant dans ses publications en ligne au nom de la liberté dont il dispose en la matière.

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 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 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, 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 du public à connaître des données détenues par une entreprise, notamment à des fins de recherche scientifique ;
- **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 du CRPA relatives à la diffusion des informations publiques détenues par les administrations, ou celles du RGPD 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 ?

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 associée 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, 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[7] :** 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

[7] À noter qu'entrera en application, en mai 2024, le règlement d'exécution (UE) 2023/138 de la Commission européenne (21 décembre 2022) « *établissant une liste d'ensembles de données de forte valeur spécifiques et les modalités de leur publication et de leur réutilisation* ». Dans le prolongement de la directive européenne « Open data » du 20 juin 2019, cet acte dresse ainsi une liste des données devant être impérativement diffusées. La Commission européenne ayant la possibilité d'adopter des actes délégués pour étendre les catégories de jeux de données à haute valeur, il convient de suivre attentivement ces évolutions.

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[8].

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 le code du patrimoine, 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 les textes prévoient 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*

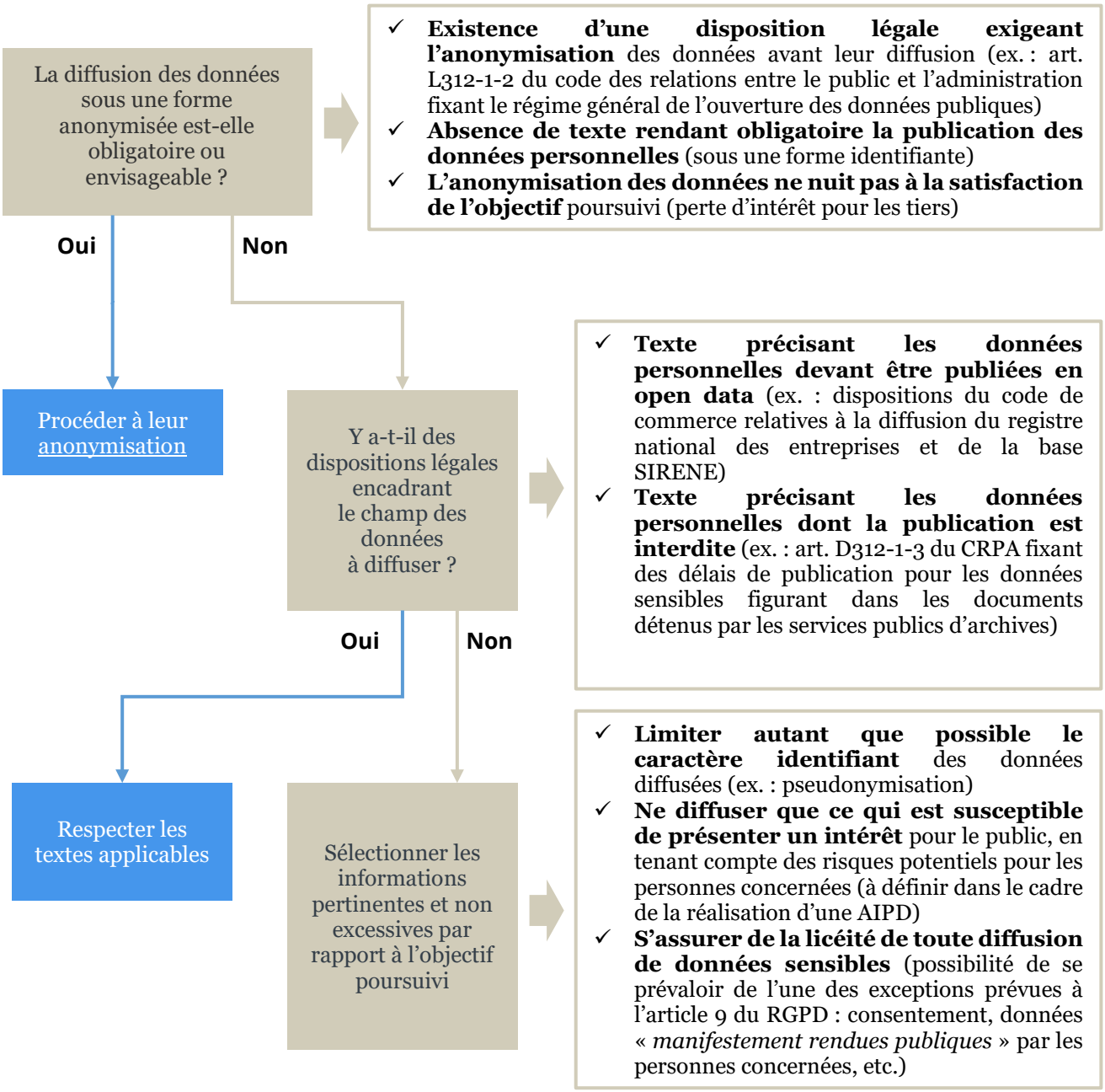
La pseudonymisation consiste à traiter les données de telle façon que celles-ci ne puissent plus directement être attribuées à une personne, en protégeant les informations permettant d'établir indirectement ce lien. Le plus souvent, il s'agit de remplacer les informations les plus identifiantes d'un 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).

Ainsi par exemple, **en cas de diffusion de données pseudonymisées relevant du champ des données sensibles** visées à l'article 9 du RGPD (p. ex. : données se rapportant à la santé, à la vie sexuelle, aux opinions politiques, convictions religieuses, etc.), pour lesquelles il existe un principe d'interdiction de traitement, le diffuseur doit au préalable **avoir vérifié et être en capacité de démontrer, en particulier dans son analyse d'impact relative à la protection des données (AIPD), qu'il peut bien se prévaloir de l'une ou l'autre des exceptions prévues par les textes** (article 9 du RGPD et 46 de la loi Informatique et Libertés) : recueil de consentements ou données déjà « *manifestement rendues publiques* », en particulier.

[8] En cas de doute sur l'étendue des données à diffuser en *open data*, les administrations peuvent saisir la CADA pour avis (article R342-4-1 du CRPA).

Schéma récapitulatif : garantir la minimisation des données diffusées



Fiche n°6 : Comment garantir l'exactitude et la sécurité 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 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 publient 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 identifiante 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.

Sauf à ce qu'il existe des dispositions légales s'y opposant (obligation d'archivage prévue par un texte), 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.

À noter

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 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 ?

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**

  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**

  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.

Exemples

- 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 moissonnages 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.

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.

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.**

Toutes les administrations recourant à la « licence ouverte de réutilisation d'informations publiques » (celle de référence établie par le Gouvernement)[9], qui se limite à pointer l'éventuelle nécessité de respecter le cadre juridique de la protection des données personnelles, **peuvent ainsi utilement préciser les obligations que celui-ci prévoit**.

Exemple

Sur la plateforme « data.gouv.fr » où elle diffuse son jeu de données sur les désignations des délégués à la protection des données (« 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 aux recommandations de la CNIL concernant les conditions de réutilisation de données publiées sur Internet peut, également, être opportunément effectué.

[9] Le CRPA (article L323-2) prévoit en effet que lorsque la réutilisation à titre gratuit d'informations publiques donne lieu à l'établissement d'une licence, cette licence est choisie parmi celles figurant sur une liste fixée par décret, qui est révisée tous les cinq ans, après concertation avec les collectivités territoriales et leurs groupements. Lorsqu'une administration souhaite recourir à une licence ne figurant pas sur cette liste, cette licence doit être préalablement homologuée par l'État, dans des conditions fixées par décret.

De plus, pour limiter les risques de réutilisation à des fins illégitimes, les diffuseurs de données peuvent, dans certains cas, **apporter par voie contractuelle** (ex. : conditions générales d'utilisation) **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* ».

À noter toutefois que s'agissant des informations publiques diffusées en application du CRPA, **les administrations devront veiller à ne pas se montrer davantage restrictives qu'elles ne le peuvent légalement**[10]. En effet, l'article L323-2 de ce code, relatif aux licences de réutilisation de ces informations, précise que « *cette licence fixe les conditions de la réutilisation des informations publiques. Ces conditions ne peuvent apporter de restrictions à la réutilisation que pour des motifs d'intérêt général et de façon proportionnée. Elles ne peuvent avoir pour objet ou pour effet de restreindre la concurrence* ».

Ainsi, la liberté dont disposent les administrations en la matière se trouve notamment limitées par les critères d'homologation par l'Etat des licences spéciales de réutilisation (qui doivent être établies dans le respect des conditions prévues à l'article D. 323-2-2 du CRPA). Elle peut également l'être par certaines dispositions légales, telles celles du règlement d'exécution (UE) 2023/138 de la Commission européenne (21 décembre 2022) « *établissant une liste d'ensembles de données de forte valeur spécifiques et les modalités de leur publication et de leur réutilisation* ». Celui-ci prévoit en effet, s'agissant des données d'entreprises, « *une mise à disposition « aux conditions de la licence Creative Commons BY 4.0 ou de toute licence ouverte équivalente ou moins restrictive ; sans autres conditions supplémentaires concernant la réutilisation de données à caractère personnel, le cas échéant* ».

[10] À noter que la CADA se montre vigilante à cet égard, comme en témoigne son conseil n°20183496 du 24 janvier 2019 (voir le point 3).

Schéma récapitulatif : garantir l'exactitude et la sécurité des données diffusées

Mesures garantissant l'exactitude des données

- ✓ **Veiller, préalablement à leur ouverture, à ce que les données aient fait l'objet de toutes les corrections** nécessaires
  - Bonne pratique : attirer, si possible, l'attention des personnes concernées 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** (ex. : formulaire de contact)
- ✓ **Dater les données diffusées** (notamment via la fourniture de métadonnées)
- ✓ **Adopter des modalités techniques permettant (si pertinent) de les mettre à jour régulièrement** et de façon automatisée (notamment via le recours à des interfaces de programmation applicative/API)
- ✓ **Tracer les modifications opérées sur les données et en informer l'ensemble des destinataires**, si possible et sauf efforts disproportionnés

Mesures garantissant la sécurité des données

- ✓ **Prévoir des modalités d'accès aux données adaptées à leur sensibilité** (dans le respect, par les administrations, du CRPA)
  - 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
  - Adoption d'un dispositif empêchant des moissonnages 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 »)
  - 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
  - Recours à une API pour les données sensibles (meilleure supervision de leur partage)
- ✓ **Vérifier la robustesse de l'anonymisation/pseudonymisation, le cas échéant, et assurer la sécurité des données qui ne sont pas diffusées**
  - Cloisonnement logique ou physique entre les données ouvertes et celles permettant leur identification
  - Mise à disposition des réutilisateurs d'un point de contact pour signaler tout risque d'atteinte à leur confidentialité
- ✓ **Veiller à l'intégrité et la disponibilité des données partagées**
  - Adoption de mesures adaptées de traçabilité et de vérification de l'intégrité
  - Rédaction d'un plan de reprise et de continuité d'activité informatique
- ✓ **S'assurer que les données ne sont pas diffusées pour une durée excessive** (réduction de la vraisemblance des risques)