Réutilisation de bases de données : les vérifications nécessaires pour respecter la loi

23 janvier 2025

Les responsables de traitement qui utilisent des bases de données librement mises à disposition sur Internet ou fournies par un tiers doivent vérifier que leur constitution ou leur partage n'est pas manifestement illicite. La CNIL rappelle les vérifications à réaliser.

Dans le cadre de leurs activités (recherche scientifique, développement de systèmes d'intelligence artificielle, prospection commerciale), certains organismes souhaitent réutiliser des bases de données personnelles :

- **librement mises à disposition sur Internet** en dehors du cadre légal relatif à l'ouverture des données (« open data ») ;
- **détenues par un tiers** (par exemple, un courtier en données ou data broker).

L'organisme ou la personne qui constitue, met en ligne ou partage la base de données doit respecter la loi (par exemple, il est interdit de voler ou de diffuser des données volées). De la même manière, **tout réutilisateur d'une base de données doit s'assurer que sa constitution ou son partage n'est pas manifestement illicite** (par exemple, il est interdit de réutiliser des données provenant d'une fuite de données).

Le réutilisateur ne peut pas réutiliser une base de données constituée ou mise en ligne **si elle enfreint manifestement** le règlement général sur la protection des données (article 5.1.a du RGPD) ou d'autres règles, telles que celles interdisant les atteintes à la sécurité des systèmes d'information ou les atteintes à des droits de propriété intellectuelle.

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

Vérifier que la constitution ou le partage de la base de données n'est pas manifestement illicite

Le réutilisateur de bases de données doit procéder à des vérifications, mais sans nécessairement aller jusqu'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.) par le tiers qui a constitué ou mis à disposition la base de données.

Pour l'aider dans cette démarche, **la CNIL recommande en particulier aux réutilisateurs** de vérifier les points suivants :

1 La description de la base de données mentionne leur source.

Exemples :

- Une base de données dont la description expliquerait qu'elle a été constituée à partir de publications sur un réseau social professionnel dont le nom est précisé.
- À l'inverse, une base de données contenant des images de vidéosurveillance sans précision de la source ne devrait pas être réutilisée avant d'avoir obtenu davantage de précisions permettant de lever les doutes quant à la conformité de sa constitution et de sa diffusion.

2 La constitution ou la diffusion de la base de données ne résulte pas manifestement d'un crime ou d'un délit. Il n'y a pas eu de condamnation ou de sanction publique de la part d'une autorité compétente entraînant une suppression ou une interdiction d'exploiter ultérieurement les données.

Exemples :

- En cas d'achat d'une base de données sur le *dark web* provenant par exemple d'une fuite ou d'un vol de données (au sens de l'article 323-1 du code pénal), vous ne pouvez pas ignorer son origine délictuelle : la base de données est donc manifestement illicite.
- Il en irait de même si vous souhaitez réutiliser une base de données pour laquelle une décision de justice a retenu une atteinte à un droit de propriété intellectuelle comme celui, particulier, des producteurs de bases de données (au sens de l'article L. 342-1 du code de la propriété intellectuelle).

3 L'origine des données est suffisamment documentée pour qu'il n'y ait pas de doute flagrant sur le fait que la base de données est licite.

Il s'agit notamment de vérifier que le traitement source repose sur une base légale, en particulier si les données sont tellement intrusives qu'elles ne pourraient être traitées sans le consentement des personnes.

Exemples :

- Sur une plateforme d'échange de bases de données, vous repérez 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, la diffusion d'une telle base de données en l'absence de consentement des personnes **peut être illicite.**
- À l'inverse, il serait envisageable de constituer un jeu de données à partir d'une base de données dont la description ne laisse pas de doute flagrant quant à sa licéité. Par exemple, une base de données pseudonymisées, initialement rendues publiques par les personnes concernées sur un site web identifié et qui ne contiendrait pas de données sensibles.

4 La base de données ne contient pas de données sensibles ou de données d'infraction.

Si elle en contient, il est recommandé de mener des vérifications supplémentaires pour s'assurer que la constitution de la base de données ou sa mise à disposition est licite :

- pour les données sensibles (données de santé ou révélant des opinions politiques par exemple, cf. article 9 du RGPD), il s'agit principalement de s'assurer du recueil d'un consentement explicite des personnes concernées, ou que les données ont été manifestement rendues publiques par ces dernières (par ex. : une publication que la personne a délibérément rendue visible à tous sur les réseaux sociaux) ;

- pour les données relatives à des infractions (article 10 du RGPD), une telle utilisation doit être autorisée par la loi Informatique et Libertés.

Attention : le responsable du traitement qui a constitué ou partagé la base de données aura pu commettre d'autres manquements au RGPD. Il s'agit, par exemple :

- de mentions d'informations incomplètes lors de la constitution ou de la diffusion de la base de données ; ou

- un 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).

De tels manquements ne signifient pas nécessairement qu'une réutilisation est illicite, **à condition que celle-ci soit elle-même conforme au RGPD (voir ci-dessous).**

S'assurer de la conformité de son propre traitement de données

Ces vérifications préalables n'exonèrent pas l'organisme qui réutilise la base de données d'une analyse complète de la conformité de ses propres traitements au RGPD et à la loi (identifier une base légale, recueillir le consentement explicite pour le traitement, le cas échéant, de données sensibles ou identifier une autre exception au principe d'interdition posé par l'article 9 du RGPD, etc.).

Intégrer ces vérification dans le cadre d'**une analyse d'impact relative à la protection des données (AIPD) peut constituer une bonne démarche.**

À toutes fins utiles, la CNIL propose une méthode et des outils pour engager ou affiner une démarche de conformité au RGPD (définir une finalité et une base légale, informer les personnes, minimiser les données, etc.).

Encadrer les relations avec le détenteur des données

La CNIL recommande **la conclusion d'un accord** entre le détenteur initial des données afin de permettre de s'assurer que la réutilisation des données est licite.

À cet égard, la CNIL recommande de **fournir un certain nombre d'indications** dans le contrat telles que :

- la source, le contexte de la collecte des données, la base légale du traitement et l'analyse d'impact relative la protection des données si nécessaire, afin d'écarter les risques d'avoir une base de données illicite ;

- les mentions d'information des personnes portées à la connaissance des personnes (en particulier s'agissant de la finalité et des destinataires) ;

- d'éventuelles garanties sur le fait que le partage de données par le détenteur initial des données est licite (par exemple : la finalité pour la réutilisation doit être compatible avec la finalité initiale, le partage doit être licite, etc.).

Pour approfondir

- Ouverture et partage des données : la CNIL publie ses recommandations
- IA : assurer que le traitement est licite – En cas de réutilisation des données, effectuer les tests et vérifications nécessaires