



## DONNÉES PERSONNELLES

# Quels faisceaux d'indices pour un entrepôt de données de santé ?

Dans une récente affaire, la Cnil a appliqué un faisceau d'indices pour qualifier une base de données d'entrepôt de données de santé. Ce type de décision (publiée) est suffisamment rare pour être souligné.

**L**a création d'entrepôts de données de santé ne fait que croître, aussi bien dans le secteur privé que dans le secteur public. Les entrepôts de données de santé, eu égard à la nature des données qu'ils traitent (données sensibles) et à leur volume, sont soumis à un cadre juridique strict. Pour autant, la notion d'entrepôt de données de santé ne fait pas l'objet d'une définition légale. La qualification d'une base de données en entrepôt de données de santé n'est pas, en conséquence, toujours aisée.

### Les entrepôts de données de santé soumis à un cadre juridique strict...

Pour mémoire, la Cnil a publié un référentiel relatif « *aux traitements de données à caractère personnel mis en œuvre à des fins de création d'entrepôts de données de santé dans le domaine de la santé* »<sup>1</sup> (le « référentiel »). Ce référentiel concerne uniquement les entrepôts de données de santé « *nécessaire à l'exécution d'une mission d'intérêt public ou relevant de l'exercice de l'autorité publique dont est investi le responsable de traitement* ».

Ainsi, un organisme, responsable de traitement, souhaitant mettre en œuvre un entrepôt de données de santé dans le cadre de l'exercice d'une mission d'intérêt public, doit, par principe, s'assurer de la conformité de son projet au référentiel. Dans le cas où l'organisme considère que son projet est en stricte conformité avec le référentiel, il peut alors se contenter d'une déclaration de conformité auprès de la Cnil.

Dans l'hypothèse, en revanche, où il existe des écarts avec les exigences prévues au référentiel, il appartient à l'organisme de saisir la Cnil d'une demande d'autorisation spécifique préalable<sup>2</sup>. Tel est le cas notamment pour un organisme privé qui réalise un entrepôt de données de santé dans le cadre de son intérêt légitime.

### L'absence de définition légale

Si les entrepôts de données de santé ne font pas l'objet d'une définition légale, plusieurs autorités administratives/publiques indépendantes et groupement d'intérêt public ont eu l'occasion d'apporter des éclairages sur cette notion.

La Cnil a publié plusieurs documents dans lesquels elle explique ce qu'il faut entendre par « *entrepôt de données de santé* ». Ainsi, par exemple, dans une publication « *Traitements de données de santé : comment faire la distinction entre un entrepôt et une recherche et quelles conséquences* » du 2 mars 2023, l'autorité de protection des données a indiqué que : « *Lorsque le responsable de traitement envisage la constitution d'une importante base de données et la réutilisation des données contenues (« entrepôt ») dans plusieurs projets de recherche, un raisonnement en deux temps doit être opéré car chaque traitement fait l'objet d'un régime juridique distinct : - la création de l'entrepôt de données en tant que tel (c'est-à-dire la collecte et la conservation des données dans une base unique pendant une plus longue durée) ; - les projets de recherches, d'études ou d'évaluations réalisés à partir des données conservées dans l'entrepôt par le même responsable de traitement ou d'autres organismes* ». Par ailleurs, la Cnil a expliqué, dans les termes suivants, comment opérer la distinction entre un « *entrepôt de données de santé* » et

une « recherche » : « Les entrepôts de données sont créés principalement pour collecter et disposer des données massives (données relatives à la prise en charge médicale du patient, données socio-démographiques, données issues de précédentes recherches, registre de pathologies, etc.). Ces données sont ensuite réutilisées, la plupart du temps partiellement, à des fins d'études, de recherches et d'évaluations dans le domaine de la santé. Ces bases de données sont constituées pour une longue durée (au moins 10 ans en général) et l'objectif est d'obtenir un volume de données important. Elles peuvent être alimentées par de multiples sources (professionnels de santé, patients, pharmacies, établissements de santé, etc. ».

Dans un communiqué du 26 octobre 2021 annonçant la publication du référentiel, la Cnil a défini les entrepôts de données de santé comme des « bases de données destinées à être utilisées notamment à des fins de recherche, d'études ou d'évaluation dans le domaine de la santé ». Le référentiel, quant à lui, ne contient aucune définition de l'entrepôt de données de santé. Il précise toutefois qu'il s'applique aux responsables du traitement qui souhaitent « réunir des données en vue de leur réutilisation, pour les finalités mentionnées au point 3.1 [telles que la production d'indicateurs et le pilotage stratégique de l'activité, l'amélioration de la qualité de l'information médicale, l'optimisation du codage dans le cadre du programme de médicalisation des systèmes d'information, le fonctionnement d'outils d'aide au diagnostic médical ou à la prise en charge, la réalisation d'études de faisabilité, ou plus généralement, la recherche ou l'étude dans le domaine de la santé] ». À noter que plusieurs CHU ont eu l'occasion de solliciter et d'obtenir de la Cnil une autorisation dans le cadre de la constitution d'un entrepôt de données de santé regroupant des données produites lors de la prise en charge des patients du CHU et ce, à des fins de recherches<sup>5</sup>. Ces délibérations – tout comme celles qui ont été rendues dans d'autres cas de figure – n'apportent

toutefois aucun éclairage supplémentaire sur la notion d'entrepôt de données de santé. En l'occurrence, il s'agissait, dans chacun de ces cas, d'un organisme qui décide – seul ou conjointement avec un partenaire – de regrouper des données de santé à des fins de recherches.

Selon le Health Data Hub, les entrepôts de données de santé visent à collecter et mettre à disposition des données « qui peuvent être massives et issues de sources diverses (établissements de santé et notamment les hôpitaux, professionnels de santé, patients...) », pour être ensuite « réutilisées à des fins de recherche ou encore d'évaluation dans le domaine de la santé. ». Le Health Data Hub précise que les établissements de santé, qui ont déployé des entrepôts de données de santé, visent principalement à « industrialiser le recueil et le rassemblement de données ». Une telle initiative leur permet (i) « à moyen terme l'interconnexion des multiples sources de données de [l'établissement] ; (ii) « davantage de visibilité aux potentiels utilisateurs sur l'existant en matière de données (...) utilisables pour la recherche, l'innovation ou le pilotage » et (iii) des opportunités de collaboration externe, y compris avec d'autres secteurs et à l'étranger<sup>4</sup>.

Pour la Haute Autorité de Santé (HAS), les entrepôts de données de santé doivent être définis comme « la mise en commun des données d'un ou plusieurs systèmes d'information médicaux, sous un format homogène pour des réutilisations à des fins de pilotage, de recherche ou dans le cadre des soins. ». Ces réutilisations sont qualifiées de « réutilisations secondaires ». La HAS considère que la création d'entrepôts de données de santé repose sur plusieurs étapes. Il y a d'abord la collecte des données « depuis les différentes sources constituant le SIH » afin de « centraliser ces données initialement cloisonnées dans chacun des SI » dans un entrepôt de données de santé. Ensuite, vient l'étape de la transformation et de l'agrégation des données « rarement homogènes » et ce, « afin d'aboutir à

un entrepôt exploitable ». La HAS précise par ailleurs que la notion d'entrepôt de données de santé intègre aussi « la plateforme technologique utilisée pour transformer ces données ». La troisième étape réside dans « la mise à disposition « des jeux de données spécifiques (parfois nommés datamarts) à chaque usage secondaire de la donnée »<sup>5</sup>.

Pour le Comité consultatif national éthique, un entrepôt de données de santé est « une infrastructure informatique qui rassemble en un lieu physique des données, pouvant être exprimées dans des formats variés, provenant de plusieurs sources souvent hétérogènes, et parfois de natures très différentes ». On retrouve également la notion d'organisation / transformation des données (i) « selon un modèle unifié de manière à faciliter pour l'utilisateur leur exploitation » ou (ii) sans un tel modèle, « dès lors que le lien entre les différentes données des bases de données rassemblées dans l'entrepôt peut être établi ». Par ailleurs, le Comité indique qu'il y a de plus en plus d'appariements entre des entrepôts de données de santé et d'autres plateformes de données « pour pouvoir permettre in situ des traitements sur leurs données. »<sup>6</sup>.

Au regard de tous ces éléments, la qualification d'entrepôt de données de santé pourrait ainsi notamment être retenue en présence des critères suivants : (i) collecte et stockage, sous forme homogène, de données de santé massives issues de sources diverses ; (ii) réutilisation de ces données à des fins de recherches (usage secondaire) dans le secteur de la santé.

### La Cnil et le faisceau d'indices

À l'occasion d'une affaire contentieuse, la formation restreinte de la Cnil a récemment qualifié une base de données d'entrepôt de données de santé<sup>7</sup>. L'affaire concernait un éditeur équipant des cabinets médicaux et des centres de santé d'un logiciel de gestion d'agenda, de dossiers patients et de prescriptions. Dans ce cadre, l'éditeur proposait à certains de ses utilisateurs d'adhérer

à un observatoire pour collecter des données de dossiers patients traités par le logiciel pour qu'elles soient ensuite transmises et utilisées par d'autres clients de l'éditeur (dont des sociétés appartenant au groupe de l'éditeur), notamment pour réaliser des études en santé. À la suite d'un contrôle sur place, le rapporteur a considéré que, du fait de ces traitements, l'éditeur s'était constitué un entrepôt de données et aurait dû, pour ce faire, solliciter une autorisation de la Cnil, ce qu'il n'avait pas fait.

Dans son analyse, la formation restreinte a souligné que la notion d'entrepôt de données de santé n'est pas définie par la loi (en l'occurrence la loi Informatique et libertés), et qu'il s'agit d'« une construction doctrinale de la Cnil (...) ». L'autorité de protection des données, qui estime que la qualification d'un entrepôt de données de santé s'apprécie « à l'aide d'un faisceau d'indices », a listé les trois éléments qui lui semblent être « déterminants » pour une telle qualification : (i) « la réutilisation des données dans des traitements ultérieurs », (ii) « l'alimentation au fil de l'eau de la base », (iii) « ainsi que [les] finalités du traitement ». Elle a également précisé que ce faisceau d'indices doit tenir compte « notamment, mais pas uniquement » de la durée de conservation des données de santé à caractère personnel.

Dans le cas d'espèce, la Cnil a d'abord retenu que l'éditeur avait collecté « massivement des données de

santé de patients et de médecins », et ce sur une période de plusieurs mois. Si la collecte de données massives ne fait pas partie des « éléments déterminants » retenus par la Cnil, ce critère a toutefois été utilisé par la Cnil. Ensuite, l'autorité a relevé que l'éditeur avait alimenté sa base « au fil de l'eau, afin d'obtenir un volume important de données (remontée journalière des données depuis les postes médecins) ». Enfin, la Cnil a constaté que l'éditeur mettait les données à disposition de ses clients à des fins « d'études et des statistiques dans le domaine de la santé ». La Cnil en a donc tiré la conclusion que l'éditeur s'était bien constitué un entrepôt de données de santé.

Cette décision nous livre un éclairage pratique bienvenu quant à la qualification d'une base de données en entrepôt de données de santé. Nous ne sommes toutefois pas à l'heure d'une pleine clarification. Certes, dans cette décision, la Cnil a souligné des éléments à prendre en compte dans le travail de qualification d'un entrepôt de données de santé, pour autant, ces éléments ne sont pas limitatifs. La prudence reste donc de mise...

**Alexandre FIEVEE**

Avocat associé

**Alice ROBERT**

Avocate of Counsel

DERRIENNIC ASSOCIES

## Notes

- (1) [https://www.cnil.fr/sites/cnil.fr/files/atoms/files/referentiel\\_entrepot.pdf](https://www.cnil.fr/sites/cnil.fr/files/atoms/files/referentiel_entrepot.pdf)
- (2) À noter que l'organisme pourrait aussi se dispenser de formalité Cnil dans l'hypothèse théorique d'un recueil du consentement explicite des personnes concernées par les traitements de l'entrepôt de données de santé (collecte, enregistrement, conservation, etc.), sous réserve de pouvoir démontrer que les traitements mis en œuvre sont conformes au RGPD.
- (3) Délibération 2020-028 du 27 février 2020 (eHop Rennes – CHU Rennes) ; Délibération n°2019-124 du 10 octobre 2019 (CHUGA-EDS – CHU Grenoble) ; Délibération n°2019-103 (Include – CHU Lille).
- (4) Health Data Hub, « Kit de création d'un entrepôt de données de santé », janvier 2024.
- (5) HAS, rapport « Entrepôts de données de santé hospitaliers en France. Quel potentiel pour la Haute Autorité de santé ? », 20 octobre 2022.
- (6) Avis commun « Plateforme de données de santé : enjeux d'éthique », février 2023.
- (7) Délibération SAN-2024-013 du 5 septembre 2024.



Vous avez envie de vous exprimer sur un sujet qui vous tient à cœur, de partager votre analyse avec la communauté des lecteurs d'Expertises, d'exposer un point de vue différent sur un article déjà publié, de lancer un débat sur un thème émergent, ou simplement de commenter l'actualité du droit du numérique ?

Contactez la rédactrice en chef d'Expertises Sylvie Rozenfeld [sr@expertises.info](mailto:sr@expertises.info)