Consultation sur le projet de recommandation technique relatif à l'utilisation des API pour le partage de données personnelles

Synthèse des contributions

Le 20 septembre 2022, la CNIL a lancé une consultation publique sur son projet de recommandation technique relative à l'utilisation des interfaces de programmation applicatives (API) pour le partage sécurisé de données personnelles (ou « recommandation API ») afin de recueillir les avis des professionnels concernés.

Les contributions ont nourri les travaux de la CNIL en vue de la publication de la recommandation.

CNIL.
COMMISSION NATIONALE
INFORMATIQUE & LIBERTÉS

Synthèse des contributions

Cette consultation publique a majoritairement fait l'objet de contributions de la part de professionnels de la sécurité des données et des systèmes d'information (DPO/DPD, et RSSI) provenant d'organismes publics et privés en proportions similaires. Ces organismes ont déclaré utiliser ou fournir des API et sont ainsi directement impactés par cette recommandation.

À propos de la consultation publique de la CNIL sur le projet de recommandation API

La CNIL a pu observer depuis plusieurs années une augmentation des dispositifs visant à partager des données entre administrations, organismes privés ou encore directement avec des particuliers. Dans l'optique d'une approche de protection des données dès la conception, l'utilisation des interfaces de programmation applicatives, ou API pour ***application programming interface*** en anglais, pour réaliser ces partages peut être recommandée dans certains cas. Elle ne doit toutefois pas se faire sans prendre en compte certaines bonnes pratiques.

Afin de les promouvoir, la CNIL a prévu l'adoption d'une recommandation technique relative à la mise en œuvre et à l'utilisation des API à l'intention de tous les acteurs de la chaîne du partage. Tous les types de partages de données personnelles par voie d'API et tous les types d'organismes sont visés par cette recommandation.

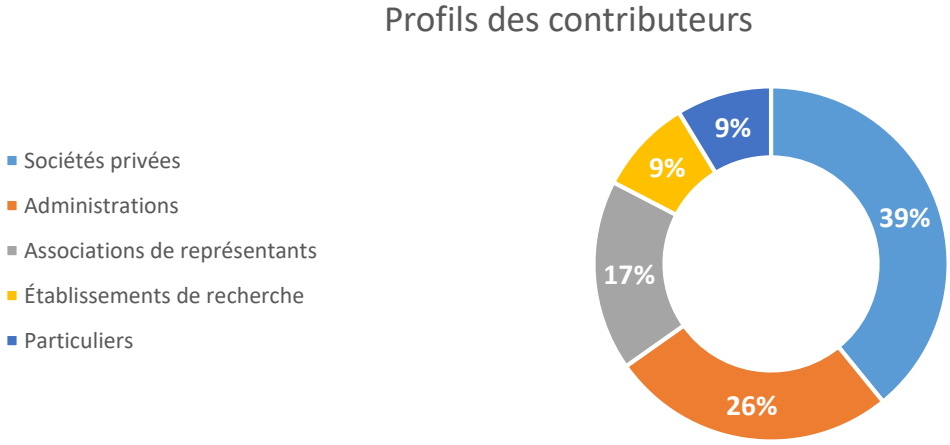
Un projet de recommandation a été publié sur le site de la CNIL afin de recueillir les retours du public du 20 septembre au 1er novembre 2022, en vue de la préparation de la version définitive de la recommandation.

Quelques chiffres sur les participants

24 contributions provenant de contributeurs issus de divers secteurs d'activité ont été reçues suite à la consultation : 9 provenaient de sociétés privées, 6 d'administrations, 4 d'associations de représentants, 2 d'établissements de recherche et 2 de particuliers. Parmi les participants ayant précisé leur fonction, 7 étaient DPO et 8 tenaient un rôle lié à la gestion des données ou à leur sécurité. Enfin, les trois catégories introduites dans la recommandation (détenteur de données, gestionnaire d'API et réutilisateur), sont représentées :

- 12 possèdent les trois;
- 2 ont deux catégories parmi les trois ;
- 1 n'est que détenteur de données ;
- 3 ne sont que gestionnaires d'API ;
- 1 est seulement réutilisateur ;
- 5 n'ont aucune des trois catégories.

Ces chiffres montrent que les profils des participants sont particulièrement diversifiés, à l'image des cas d'usages visés par la recommandation.



CNIL.

Une recommandation ambitieuse

Dans le détail, les réponses aux questions fermées semblent indiquer que la recommandation vise à promouvoir un niveau particulièrement élevé de sécurité, sur un large périmètre. Parmi les participants,

- **13 sur 24 ont répondu non** à la question « Les mesures recommandées sont-elles suffisamment claires et concrètes pour être mises en pratique ? » ; et
- **13 sur 24 ont répondu non** à la question « Les mesures recommandées vous semblent-elles réalisables ? ».

Ces résultats, confirmés par les réponses aux questions ouvertes, montrent que **le caractère non prescriptif des recommandations, les cas dans lesquels celles-ci sont préconisées et la charge de leur mise en œuvre** sont insuffisamment compris à la lecture de la recommandation. Toutefois, les précisions apportées dans les questions ouvertes, et détaillées dans la suite du document, viennent clarifier les raisons de cette incompréhension, en distinguant différents cas :

- certaines recommandations (comme le cloisonnement des bases de données ou la restriction d'accès par filtrage des adresses IP) sont considérées comme trop exigeantes et les répondants estiment qu'il conviendrait de restreindre à certains cas bien précis ;
- d'autres recommandations ne semblent pas réalisables en pratique car les cas où celles-ci devraient être mises en œuvre ne sont pas suffisamment délimités.

Ces contributions ont permis à la CNIL :

- d'améliorer la recommandation en précisant certains des termes utilisés ;
- de clarifier les principaux points incompris dans une page explicative qui sera publiée sur le site web de la CNIL ultérieurement ;
- d'illustrer la recommandation par des cas d'usages ayant vocation à être publiés sur le site web de la CNIL ;
- de faciliter l'application de la recommandation grâce à une grille d'analyse liant les facteurs de risques spécifiques aux API à des objectifs concrets ;
- de détailler davantage certains points tels que les rôles techniques introduits ou encore les critères permettant de déterminer les conditions dans lesquelles une recommandation est préconisée.

La suite de ce document présente les problématiques le plus fréquemment soulevées dans les réponses.

Principales problématiques soulevées

Sur le caractère contraignant des recommandations

Certaines des contributions reçues expriment un manque de clarté dans le caractère obligatoire des mesures préconisées dans la recommandation. Par ailleurs, d'autres réponses, questionnant le coût potentiellement induit par ces mesures ou la difficulté relative à leur mise en œuvre en pratique, semblent partir de l'hypothèse que les mesures préconisées visent à s'appliquer à tous les types de partages de données personnelles par voie d'API.

D'après les réponses, cette ambiguïté pourrait être due à plusieurs causes :

- l'équivocité de l'emploi du conditionnel « devrait » dans le document ;
- l'emploi de formules visant à restreindre le champ d'application des recommandations, sans que celui-ci soit précisé, telles que « lorsque cela s'applique » ;
- la difficulté à utiliser la grille d'analyse proposée pour déterminer de manière opérationnelle le niveau de risques d'un traitement ;
- un manque de clarté plus général sur le périmètre de la recommandation et la charge de la mise en œuvre des recommandations (notamment sur l'application de certaines mesures aux API ouvertes ou à

accès restreint, ou encore sur certains cas de partage où la situation ou la nature des organismes impliqués ne permet pas la flexibilité nécessaire à la mise en œuvre des recommandations[1]) ;

Conformément à ce qui avait été annoncé lors du lancement de la consultation, **la CNIL publiera prochainement une fiche pratique visant à préciser comment les recommandations pourront s'appliquer aux cas concrets et être mises en œuvre en pratique par les acteurs impliqués.**

De plus, **le conditionnel, lorsqu'il est employé, indique ici un caractère non contraignant des mesures recommandées**. Celles-ci sont alors préconisées à titre de bonne pratique par la CNIL, et **il est laissé à l'appréciation des acteurs impliqués de juger si leur mise en œuvre est opportune**. Les critères devant être pris en compte pour juger de cette opportunité sont par ailleurs bien identifiés par les participants, comme détaillé dans la section suivante.

Enfin, des recommandations visent à préconiser **le recours à certaines méthodes dans la mise en œuvre des obligations légales des organismes, sans préjuger des cas où ces obligations s'appliquent**. Ainsi, si la CNIL recommande certaines de ces méthodes dans ce document, **il reste à la charge des organismes de s'informer des obligations relatives à leur traitement**.

Une explication plus précise de la portée des préconisations et de la signification de certains termes comme « lorsque cela s'applique » est introduite en début de la recommandation.

Sur les conditions de mise en œuvre des recommandations

Les participants ont indiqué que certaines des mesures mises en avant ne semblaient pas réalisables en l'état pour plusieurs raisons :

- le coût potentiel lié à leur mise en œuvre ;
- des conditions pratiques ne permettant pas leur mise en œuvre (comme une importante disparité de moyens entre détenteurs de données et réutilisateurs, ou encore l'absence de restrictions d'accès pour une API ouverte) ;
- le manque de visibilité sur les finalités exactes des réutilisations ;
- le risque de contradiction entre différents principes du RGPD, comme par exemple :
    - de manquer au principe de minimisation en collectant trop fréquemment des données pour s'assurer de leur exactitude,
    - d'introduire davantage de risques en détaillant le fonctionnement de l'API dans une documentation par soucis de transparence, ou encore,
    - de risquer une plus grande indisponibilité des données en cloisonnant les données partagées des données collectées) ;

Toutefois, **les mesures recommandées doivent être étudiées sous l'angle de leur faisabilité et être en adéquation avec le niveau de risque déterminé par le responsable de traitement, selon le principe de redevabilité (ou *accountability*) introduit par le RGPD**. Ainsi, les critères précédemment cités devraient être étudiés au cas par cas, et lorsqu'ils constituent un obstacle trop important pour la mise en œuvre de certaines mesures alors que le niveau de risque résiduel reste trop important, d'autres mesures devraient être envisagées. Dans certains cas rares, où les mesures pouvant être mise en œuvre dans le partage de données par API ne suffisent pas à atteindre un niveau de risque résiduel acceptable, une autre solution que l'API devrait être envisagée, ou l'opportunité du traitement questionnée.

Pour faciliter la mise en œuvre des mesures recommandées, des précisions ont été apportées afin de cibler les situations dans lesquelles elles sont préconisées.

Sur le cadre juridique

Plusieurs participants ont indiqué que certains aspects relatifs au cadre juridique applicable au partage de données par voie d'API n'avaient pas été éclaircis par la recommandation. La recommandation API a vocation à indiquer les mesures techniques préconisées par la CNIL, **elle ne vise en aucun cas à préciser le cadre**

[1] Par exemple, une équipe de recherche académique souhaitant réutiliser les données d'un réseau social.

juridique général relatif au partage de données par voie d'API. A cet égard, les participants sont invités à consulter les travaux passés de la CNIL, et notamment :

- en ce qui concerne **l'attribution de la responsabilité du traitement, la charge de la mise en œuvre des obligations légales** (la réalisation d'une AIPD, l'exercice des droits, l'information des personnes, etc.), ou encore **la contractualisation**, pourront être pertinents. En ce sens, la CNIL propose des fiches pratiques pour travailler avec un sous-traitant, ainsi que des exemples de clauses contractuelles.
- en ce qui concerne **la réponse à apporter aux demandes d'exercice des droits** : la page « Respecter les droits des personnes » sur le site de la CNIL.
- en ce qui concerne **l'authentification** :
  - la page « Sécurité : authentifier les utilisateurs » ;
  - la page « Les jetons individuels de connexion ou token access » ;
  - la recommandation relative aux mots de passe et autres secrets partagés ;
- en ce qui concerne **la journalisation**, la recommandation dédié ;
- en ce qui concerne **le rôle des délégués à la protection des données** : les fiches pratiques du site de la CNIL, accessibles depuis la page d'accueil « le délégué à la protection des données (DPO) ».

Par ailleurs, un groupe de travail sur l'ouverture et le partage de données à la CNIL viendra préciser prochainement le cadre juridique applicable à certains cas de partage de données par API. Ses missions seront de clarifier, par des critères et des exemples très concrets, la façon dont les textes doivent s'appliquer en matière de publication et de réutilisation des données, qu'elles soient publiques ou privées, publiées sur le web ou partagées entre des entreprises.

Sur le périmètre de la recommandation

Sur les traitements visés par la recommandation

Il ressort des réponses à la consultation que si dans l'ensemble, les traitements visés par la recommandation API sont clairs, certains cas précis nécessitent d'être détaillés. Les participants ont souvent demandé des exemples pour les cas visés par la recommandation. Pour répondre à ce besoin, **la publication finale de la recommandation est accompagnée de cas concrets de partage de données mettant en œuvre les mesures préconisées**.

En particulier, sur les types d'organismes visés par la recommandation, **tout type d'organisme public ou privé est concerné** dès lors qu'il participe à un partage de données personnelles par API.

De plus, certaines contributions ont questionné le rôle des **fournisseurs d'outils ou de service de développement ou de gestion des API** (tels que les fournisseurs de passerelle d'API, de portail développeur, d'outils de gestion ou de test d'API). **Ceux-ci sont bien visés par la recommandation API et leur rôle est généralement celui de gestionnaire d'API**, en ce que les outils ou services proposés impactent directement les conditions techniques du partage de données. Toutefois, l'organisme ayant recours à cet outil ou à ce service sera certainement lui aussi gestionnaire d'API, puisqu'il met en œuvre l'API. En outre, ce rôle technique ne donne aucune indication sur la responsabilité juridique au titre du RGPD des fournisseurs, qui devra être déterminée au cas par cas.

Par ailleurs, **les API visées par la recommandation sont autant les API publiques que les API à accès restreint**, bien que certaines recommandations puissent ne s'appliquer qu'à une de ces catégories d'API. Ces cas particuliers sont précisés au sein de la recommandation.

Enfin, certains participants ont questionné les occurrences pour lesquelles les recommandations devaient être prises en compte : s'appliquent-elles pour chaque partage de données, pour chaque API, ou encore pour chaque réutilisateur ? Par nature, certaines recommandations sont nécessairement liées à une API (comme la documentation), à un partage (comme l'exactitude des données), ou encore à un réutilisateur (comme son authentification). Celles-ci n'appellent pas de précision particulière.

En revanche, certaines **recommandations plus générales**, comme l'analyse des risques, **devraient être mises en œuvre pour chaque partage de données, ou pour chaque modification significative du partage et être prises en compte tant que le partage aura lieu**. Le partage désignant ici la situation où

au moins un détenteur de données et au moins un réutilisateur utilisent les outils fournis par le gestionnaire d'API pour échanger des données.

Sur la répartition des rôles techniques

Les réponses aux questions fermées à la contribution indiquent deux tendances :

- **11 participants sur 24 ont répondu non** à la question « Les trois rôles proposés (détenteur de données, gestionnaire d'API, réutilisateur) vous semblent-ils pertinents et suffisamment clairs ? » ;
- **19 participants sur 24 ont répondu oui** à la question « Si vous utilisez des API, votre rôle (ou vos rôles) est-il clair ? Les mesures préconisées vous semblent-elles pertinentes au regard de votre rôle ? ».

Les rôles proposés semblent ainsi suffisamment clairs et concrets pour que les organismes se les approprient. Toutefois, au regard des réponses à la première question et des réponses aux questions ouvertes, les participants semblent interroger la pertinence de ces trois rôles pour certains cas particuliers.

Les trois rôles choisis permettent de formuler des recommandations concrètes sur le parcours des données que n'auraient pas permis les rôles de fournisseur et de consommateur d'API usuels. Toutefois, sur 17 réponses à la question « Vous semble-t-il nécessaire de clarifier comment les rôles techniques usuels de fournisseur et consommateur d'API s'articulent avec les trois rôles retenus ? », 7 participants ayant demandé des clarifications, notamment au moyen d'exemples, **des précisions sur l'articulation de ces rôles sont apportées en complément de la recommandation, ainsi que dans les définitions en annexe I**.

Dans leurs contributions, les participants appellent des clarifications sur les cas particuliers suivants :

- **le cas où plusieurs organismes mettent à jour une unique base de données.** Dans ce cas de figure, chacun des organismes tient le rôle de détenteur de données, alors partagé entre tous. Si la réutilisation de la base de données est partagée par ces mêmes organismes, ils pourraient ainsi également tenir le rôle de réutilisateur. La coordination prévue par la recommandation devrait ainsi s'effectuer, dans la mesure du possible, entre tous les organismes à la fois détenteurs et réutilisateurs de données.
- **le cas où un intermédiaire réalise un calcul sur les données (tel qu'un traitement statistique) avant de les partager avec un ou plusieurs réutilisateurs.** Dans ce cas, et dès lors que l'intermédiaire réalise lui-même le calcul, le partage devrait être divisé en deux : dans un premier partage l'intermédiaire serait réutilisateur de données, alors que dans le second, il tiendrait le rôle de détenteur de données. Cette particularité ne devrait pas s'appliquer aux cas où le calcul est réalisé « à la volée » sans qu'on puisse distinguer deux bases de données distinctes (avant et après calcul) chez l'intermédiaire qui n'est alors éventuellement que gestionnaire d'API.
- **le cas du cumul de plusieurs rôles par un unique organisme.** Dans ce cas courant en pratique, seules les recommandations pertinentes devraient s'appliquer. Par exemple, les recommandations relatives à la coordination entre le détenteur de données et le gestionnaire d'API pour un organisme tenant ces deux rôles ne s'appliqueraient pas, en dehors de l'éventuelle coordination entre les personnes physiques mettant en œuvre le partage pour le compte de l'organisme. Lorsque différentes filiales d'un même organisme partagent des données entre elles, certaines mesures pourraient ne pas s'appliquer si la coordination entre ces filiales est assurée par ailleurs.

Ces problématiques seront détaillées en priorité dans les cas d'usage publiés sur le site de la CNIL.

Sur les recommandations générales

Sur les cas où l'utilisation d'une API est recommandée

À la question « Les cas où l'utilisation d'une API est recommandée sont-ils suffisamment clairs ? », **19 réponses positives sur 24 ont été données**, indiquant que la grille d'analyse proposée est bien comprise. Toutefois, certaines précisions semblent devoir être apportées :

- **les cas où l'utilisation d'une API est recommandée ne sont pas exhaustifs** : d'autres critères considérés comme secondaires en ce qui concerne la protection des données pourraient justifier le recours à une API (comme la nécessité d'unifier les données provenant de plusieurs sources) ;
- **dans un cas où l'utilisation d'une API est recommandée, cela n'est pas imposé** : des raisons propres à la situation particulière étudiée pourraient justifier l'utilisation d'un autre moyen de partage ;

- comme indiqué dans la recommandation**,** bien que l'utilisation d'une API puisse être parfois recommandée**, certains principes devraient être respectés** ;
- **l'API peut être complétée par un dispositif secondaire**, utilisé comme solution alternative, par exemple lorsque l'API n'est pas disponible. Cette solution devrait toutefois respecter un niveau de sécurité équivalent à celui de l'API.

Sur l'identification des facteurs de risques

Les réponses aux questions « La grille d'analyse des risques proposée vous permet-elle de cartographier les risques liés au partage de données par API liés à votre traitement ? » et « Vous permet-elle d'identifier le niveau de risque (négligeable, limité, important, maximal) correspondant à votre traitement ? » ont reçu respectivement **14 réponses positives sur 24** et **12 réponses positives sur 24**.

Si les risques mentionnés dans la grille proposée dans la recommandation API semblent suffisamment concrets et pertinents d'après les réponses obtenues, la procédure d'utilisation de cette grille pour évaluer les facteurs de risque relatif à un partage de données semble quant à elle manquer de clarté. En effet, les participants ont fréquemment indiqué que la **liste des facteurs de risques proposée devrait être transformée en une grille accordant un score à chaque niveau de risque et permettant de déduire le niveau de risque résiduel du partage**. Cette approche avait été écartée en raison de la difficulté de généraliser une telle grille à la diversité des traitements utilisant des API. Un autre outil pourrait toutefois être considéré puisqu'en l'état, les participants semblent indiquer ne pas être en mesure de déterminer les recommandations devant être suivies sans connaissance du risque qu'il est nécessaire d'atténuer. A cet égard, une méthodologie permettant d'appliquer les recommandations sera proposée sur le site de la CNIL. Il est à noter que la grille d'analyse proposée ne pourra en aucun cas remplacer la réalisation d'une AIPD, mais elle pourrait venir l'alimenter.

Par ailleurs, certaines précisions ont été demandées par les participants :

- D'autres risques mentionnés dans les réponses pourraient être pris en compte : une disparité dans le niveau de conformité au RGPD entre les organismes (pouvant entraîner une difficulté dans l'exercice des droits par exemple), ou encore un manque de visibilité sur le parcours des données en raison de la complexité du partage ou de l'implication d'un nombre trop important d'organismes ;
- La grille d'analyse devrait être utilisée pour chaque partage de données, ou à chaque modification significative du partage et les risques devraient être pris en compte tant que le partage aura lieu. Le partage désignant ici la situation où au moins un détenteur de données et au moins un réutilisateur utilisent les outils fournis par le gestionnaire d'API pour échanger des données. En effet, l'ajout d'un réutilisateur par exemple, pourrait venir modifier le niveau de risque du partage si celui-ci était d'une nature sensiblement différente des réutilisateurs existants.
- La grille d'analyse devrait être utilisée par chacun des organismes impliqués dans le partage. Toutefois, l'analyse des risques pourrait être plus ou moins poussée selon la capacité décisionnelle de l'organisme dans le partage. Par exemple, un réutilisateur n'ayant aucune capacité de décision sur la sélection des données ou sur les conditions du partage pourrait n'utiliser la grille d'analyse des risques que pour s'assurer que les mesures mises en œuvre par lui-même sont suffisantes.