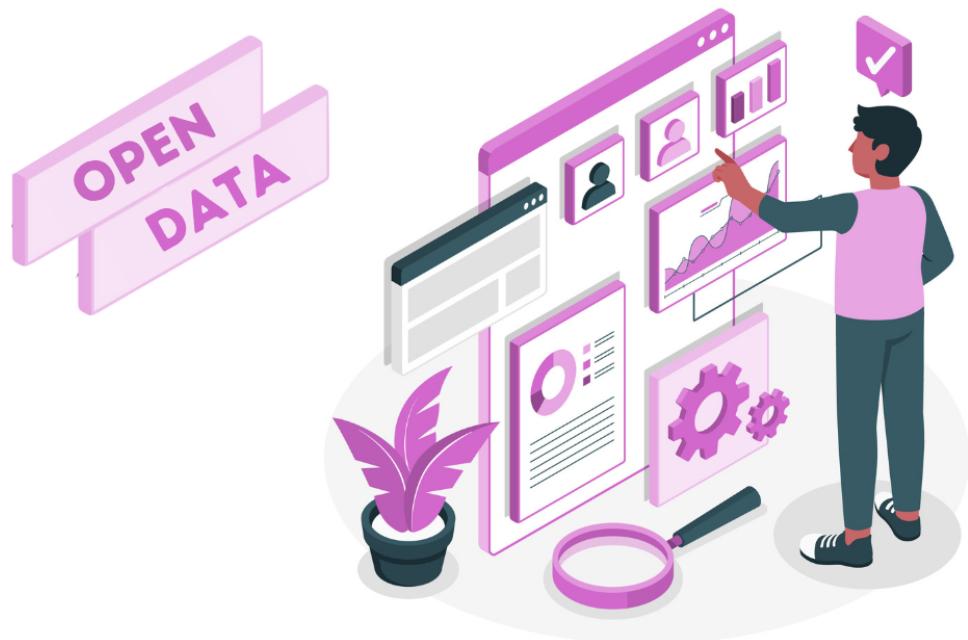




Open Data

Marie, Yasmina, Alice, Maxime, Clara

L'open Data c'est quoi ?



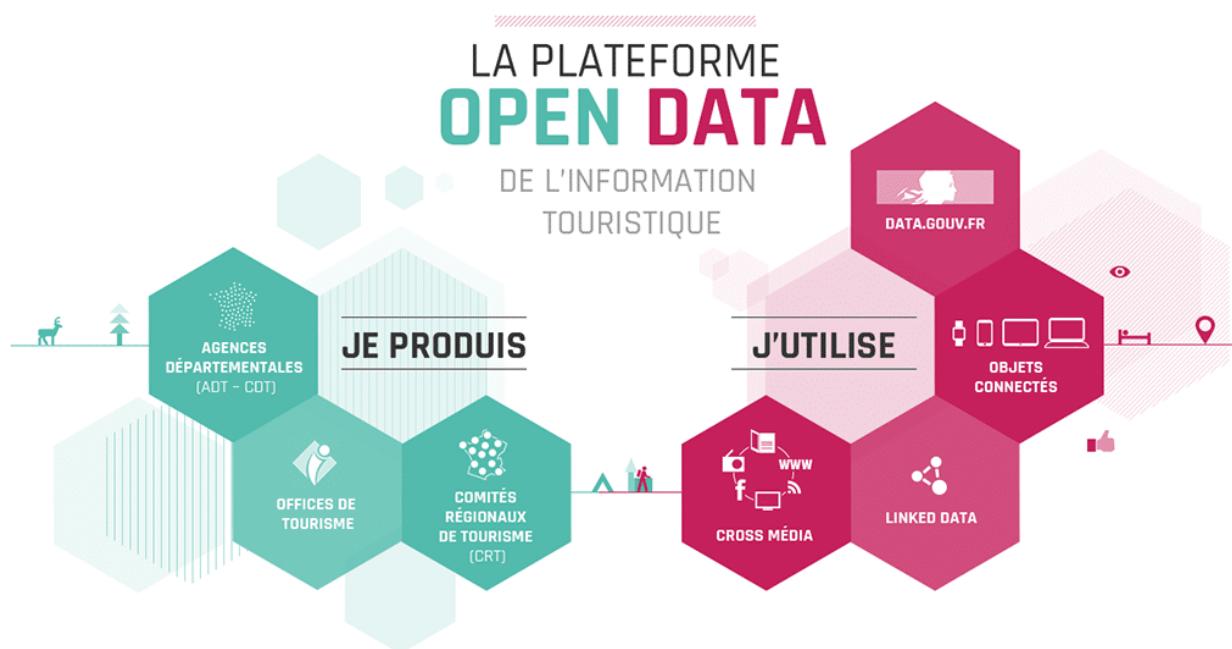
Il s'agit de données auxquelles tout le monde peut accéder et que tout le monde peut utiliser et partager.

Les gouvernements , les entreprises et les individus peuvent utiliser l'open data afin de créer des avantages sociaux.

On peut utiliser les données car elles sont disponibles sous une forme commune et lisible par des machines.

Ces données ouvertes peuvent être utilisées pour développer de nouvelles applications, améliorer les services publics ou effectuer des recherches. Par exemple, les données météorologiques ouvertes peuvent être utilisées pour créer des applications de prévisions météo plus précises. De même, les données de transport en commun ouvertes peuvent aider à optimiser les itinéraires et les horaires pour les usagers.

L'open data doit être sous licence. Cette licence doit en permettre l'usage libre et autoriser les utilisateurs à transformer, combiner et partager ces données, même à des fins commerciales.



- **Disponibilité et accès.** Les données doivent être pleinement accessibles, moyennant un coût de reproduction raisonnable. De préférence, elles se téléchargent sur Internet. La forme doit être confortable et modifiable.
- **Réutilisation et redistribution.** Les données doivent être fournies sous des conditions permettant la réutilisation et la redistribution, incluant le mélange avec d'autres ensembles de données.
- **Participation universelle.** Tout le monde doit être en mesure d'utiliser, de réutiliser et de redistribuer les données. Il ne doit y avoir aucune discrimination concernant les fins d'utilisation, ou contre des personnes ou des groupes.

Ces trois critères sont l'essence de l' Open Data, car ils autorisent l'interopérabilité. L'interopérabilité désigne la capacité de différentes entreprises ou systèmes à travailler ensemble. En l'occurrence, **l'interopérabilité est la capacité de mélanger différents ensembles de données.**

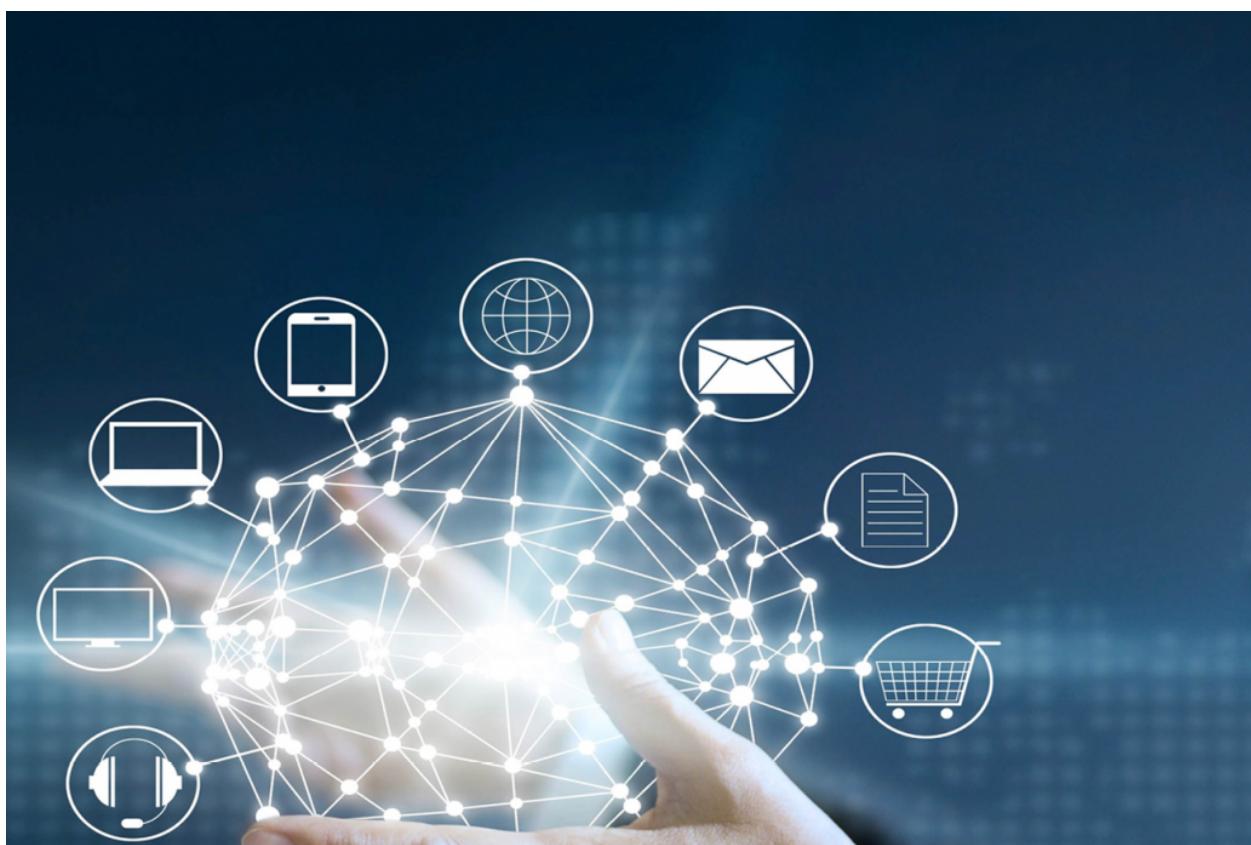
L'interopérabilité, essence de l'Open Data



Cette interopérabilité est importante, car elle permet à différents composants de fonctionner ensemble. C'est ce qui permet de créer des systèmes larges et complexes. Sans interopérabilité, c'est tout simplement impossible. On peut **prendre pour exemple le mythe de la Tour de Babel**. Dans cette légende, la communication impossible empêche complètement la construction de la tour.

Dans le cas des données, **la mise en commun repose sur la possibilité de mélanger librement ces données**. Cette interopérabilité est essentielle pour tirer des bénéfices de l'ouverture. Il est ensuite possible de développer des produits et des services en plus grande quantité, et d'une meilleure qualité.

L'interopérabilité des données ouvertes favorise également l'innovation et la créativité. En permettant à divers acteurs d'accéder et de combiner librement différentes sources de données, on ouvre la voie à de nouvelles idées et solutions. Cette approche collaborative peut conduire à des avancées significatives dans des domaines variés, de la recherche scientifique à l'amélioration des services publics.



Cette interopérabilité ne se repose pas seulement sur le partage de données. Il faut que les jeux de données utilisent un langage de programmation commun ou qu'un élément programmatique fasse l'intermédiaire entre ces informations. Pour cela, **le W3C (World Wide Web Consortium) prône l'application de standards de l' Open Data.**

Le créateur de la donnée doit en ce cas respecter des critères particuliers : certifier la provenance de la donnée, indiquer les métadonnées liées (typiquement la date et l'heure de création), garantir la qualité de l'information et si possible son auteur, etc.

On retrouve des informations ouvertes dans divers secteurs :

- de la science,
- des produits,
- de l'éducation,
- de l'environnement,
- de la cartographie,
- des bibliothèques,
- de l'économie,
- de la culture,
- du développement,
- des affaires,
- du design
- ou de la finance.



Sécurité et Conformité RGPD

Qu'est-ce que le RGPD ?



Le sigle RGPD signifie « Règlement Général sur la Protection des Données » (en anglais « General Data Protection Regulation » ou GDPR). Le RGPD encadre le traitement des données personnelles sur le territoire de l'Union européenne.

Le contexte juridique s'adapte pour suivre les évolutions des technologies et de nos sociétés (usages accrus du numérique, développement du commerce en ligne...).

Ce nouveau règlement européen s'inscrit dans la continuité de la Loi française Informatique et Libertés de 1978 et renforce le contrôle par les citoyens de l'utilisation qui peut être faite des données les concernant.



Règlement Général sur la Protection des Données

Il harmonise les règles en Europe en offrant un cadre juridique unique aux professionnels. Il permet de développer leurs activités numériques au sein de l'UE en se fondant sur la confiance des utilisateurs.

(<https://www.cnil.fr/fr/rgpd-de-quoi-parle-t-on>)

Donnée personnelle



INFOGRAPHIE

Une donnée personnelle est toute information se rapportant à une personne physique identifiée ou identifiable. Mais, parce qu'elles concernent des personnes, celles-ci doivent en conserver la maîtrise.

Une personne physique peut être identifiée :

- directement (exemple : nom et prénom) ;
- indirectement (exemple : par un numéro de téléphone ou de plaque d'immatriculation, un identifiant tel que le numéro de sécurité sociale, une

adresse postale ou courriel, mais aussi la voix ou l'image).

L'identification d'une personne physique peut être réalisée :

- à partir d'une seule donnée (exemple : nom) ;
- à partir du croisement d'un ensemble de données (exemple : une femme vivant à telle adresse, née tel jour et membre dans telle association).

(<https://www.cnil.fr/fr/definition/donnee-personnelle>)

Les grands principes du RGPD

Le RGPD s'applique à toutes entreprises, organisations et autorités publiques qui traitent des données personnelles de résidents de l'UE, quelle que soit leur localisation géographique.



Les 6 grands principes du RGPD

Le RGPD est une réglementation de l'Union européenne visant à protéger la vie privée en régulant la collecte et le traitement des données personnelles.

1 - COLLECTEZ LES INFORMATIONS STRICTEMENT ESSENTIELLES



Les informations sont recueillies dans un but spécifique et légitime, sans être traitées ultérieurement de manière incompatible avec cet objectif initial.
Le principe de finalité restreint l'utilisation future des données et empêche la collecte de données "au cas où".
Le principe de minimisation restreint la collecte aux seules données essentielles à la réalisation de l'objectif.

2 - PRIVILÉGIEZ LA TRANSPARENCE

Il est crucial que les individus maintiennent le contrôle sur leurs données.
Cela implique une information claire sur l'utilisation prévue de leurs données dès leur collecte.
En aucun cas, les données ne doivent être collectées sans le consentement explicite des personnes concernées.
De plus, les individus doivent être informés de leurs droits et des procédures pour les exercer.



3 - FACILITEZ L'EXERCICE DES DROITS DES PERSONNES



Il est essentiel d'établir des procédures permettant aux individus d'exercer leurs droits de manière efficace. Répondez promptement aux demandes de consultation, d'accès, de rectification, de suppression des données, voire d'opposition, à moins que le traitement ne réponde à une obligation légale (par exemple, un citoyen ne peut s'opposer à figurer dans un registre d'état civil). Ces droits devraient être accessibles électroniquement via une adresse dédiée.



4 - DÉTERMINEZ DES PÉRIODES DE CONSERVATION

Il est impératif de ne pas conserver les données de manière indéfinie. Les informations sont maintenues en "base active", c'est-à-dire pour la gestion courante, pendant la période strictement nécessaire à la réalisation de l'objectif visé. Par la suite, elles doivent être détruites, anonymisées ou archivées conformément aux obligations légales relatives à la conservation des archives publiques.



5 - GARANTISSEZ LA SÉCURITÉ DES DONNÉES PAR DES MESURES ADAPTÉES

Assurez la sécurité des données par des moyens physiques et informatiques, tels que la sécurisation des locaux, des postes de travail, et une gestion stricte des autorisations. Veillez à limiter l'accès aux seuls tiers autorisés par la loi, adaptant ces mesures en fonction de la sensibilité des données et des risques éventuels en cas d'incident de sécurité.



6 - INTÉGREZ LA CONFORMITÉ DANS UNE APPROCHE PERMANENTE

La conformité n'est pas statique ni immuable. Elle repose sur le respect quotidien, à tous les niveaux, des principes et des mesures mises en œuvre par les agents. Il est essentiel de vérifier régulièrement que les traitements n'ont pas connu d'évolutions, que les procédures et les mesures de sécurité mises en place sont effectivement respectées, et de les ajuster si nécessaire.

Droits des individus

- **Droit à l'information** : droit de savoir quelles données personnelles sont collectées, par qui, dans quel but, et comment elles seront utilisées.
- **Droit d'accès** : droit d'accéder à leurs données personnelles détenues par une entreprise ou organisation, ainsi qu'à des informations sur la manière dont ces données sont traitées.
- **Droit de rectification** : droit de demander la correction de leurs données personnelles inexactes ou incomplètes.
- **Droit à l'effacement (Droit à l'oubli)** : droit de demander l'effacement de leurs données personnelles dans certaines circonstances (exemple: lorsque les données ne sont plus nécessaires aux finalités pour lesquelles elles ont été collectées)
- **Droit à la limitation du traitement** : droit de demander la limitation du traitement de leurs données personnelles dans certaines conditions.
- **Droit à la portabilité des données** : droit de recevoir leurs données personnelles dans un format structuré, couramment utilisé et lisible par machine, et de transmettre ces données à un autre responsable du traitement sans entrave.
- **Droit d'opposition** : droit de s'opposer au traitement de leurs données personnelles à des fins de marketing direct ou lorsque le traitement est fondé sur des intérêts légitimes

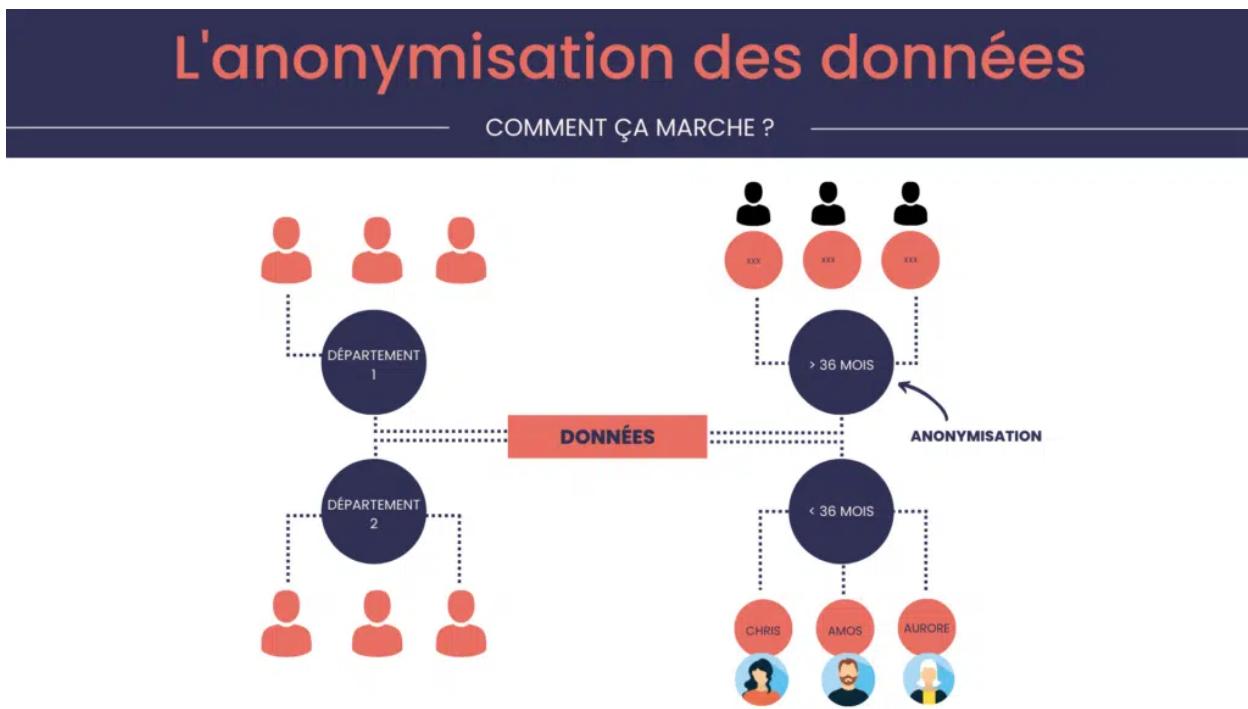
Quelles sanctions en cas de non-conformité au RGPD ?

- des sanctions financières importantes,
- des restrictions opérationnelles,
- des dommages à la réputation,
- des obligations de réparation et d'indemnisation.

<https://www.blog-qhse.com/protection-des-donnees-reglement-rgpd>

Les différentes techniques de protection des données personnelles (anonymisation, pseudonymisation...)

Comment anonymiser des données Big Data ?



Anonymiser des données Big Data peut se faire en deux étapes : la pseudonymisation et la dé-identification. L'idée est de prendre des données sensibles comme par exemple, les données de téléphone portable et les données médicales et de supprimer toute information qui peut les relier à un individu.

encryptage de colonnes: <https://dev.to/supratippb/enhancing-data-security-with-column-level-encryption-best-practices-2hol>

- **Anonymisation :** Il s'agit de transformer les données pour qu'elles ne permettent plus d'identifier une personne, même indirectement. Les données anonymisées ne sont plus considérées comme des données personnelles et sont donc hors du champ d'application du RGPD. Les techniques courantes incluent :

- La **suppression** des identifiants directs (nom, adresse, etc.),
- La **généralisation** (ex. : remplacer des âges précis par des tranches d'âge),
- La **permutation** (réassigner des valeurs de données aléatoirement au sein d'un même ensemble).
-



- **Pseudonymisation** : Cette technique consiste à remplacer les identifiants directs par des codes ou alias, permettant de limiter l'identification directe tout

en conservant un certain lien pour un usage interne. La pseudonymisation est une mesure de sécurité forte, mais contrairement à l'anonymisation, les données restent dans le champ du RGPD.

- **Agrégation** : Il s'agit de regrouper les données individuelles pour produire des statistiques ou indicateurs globaux. Cela permet de réduire le risque de réidentification tout en conservant la valeur analytique des données.

Sécurité des Données pour Prévenir les Fuites d'Informations Sensibles



- **Formation et sensibilisation** : Expliquez les risques liés aux fuites de données et les mesures préventives à prendre.
- **Gestion des accès** : Limitez l'accès aux données sensibles uniquement aux employés qui en ont besoin pour leurs fonctions. Utilisez des systèmes d'authentification robustes pour protéger les informations sensibles
- **Chiffrement et sécurité des données** : Chiffrez les données stockées et en transit. Effectuez des audits de sécurité réguliers pour identifier les vulnérabilités. Vous pourriez recommander des outils de chiffrement ou des

logiciels de sécurité spécifiques pour aider les entreprises à protéger leurs données.

- **Réaction en cas d'incident** : Élaborez un plan d'action en cas de fuite de données. Informez rapidement les parties concernées et collaborez avec les autorités compétentes. Les entreprises qui réagissent rapidement et de manière transparente en cas de fuite de données peuvent minimiser les dommages et regagner la confiance de leurs clients et employés.

[https://www.cci.fr/actualites/rh-et-protection-des-donnees-personnelles-comment-eviter-les-fuites#:~:text=Limitez l'accès aux données,l'authentification à deux facteurs.](https://www.cci.fr/actualites/rh-et-protection-des-donnees-personnelles-comment-eviter-les-fuites#:~:text=Limitez%20l'accès%20aux%20données,l'authentification%20à%20deux%20facteurs)

Le secret statistique pour protéger les données agrégées

Le secret statistique s'applique autant aux données individuelles qu'aux résultats agrégés obtenus à partir de celles-ci, dès lors que ces résultats agrégés pourraient permettre leur réidentification. De l'obligation de vérifier l'application du secret statistique aux résultats publiés découle la mise en œuvre de techniques diverses d'anonymisation ou de secrétisation, en particulier l'application des « règles du secret statistique ».

- **Suppression des petites cellules** : Lors de la publication de statistiques, il est courant de supprimer ou de masquer les données concernant des groupes de taille très réduite (par exemple, des sous-groupes de moins de 5 personnes) pour empêcher la réidentification.
- **Ajout de bruit statistique** : Des techniques telles que l'ajout de bruit aléatoire dans les données permettent de dissimuler les valeurs exactes tout en maintenant des tendances globales valides.
- **Regroupement de données** : Plutôt que de publier des données très détaillées, les données peuvent être regroupées (par exemple, par région plutôt que par ville) pour protéger les informations individuelles tout en rendant les données utiles pour les analyses.