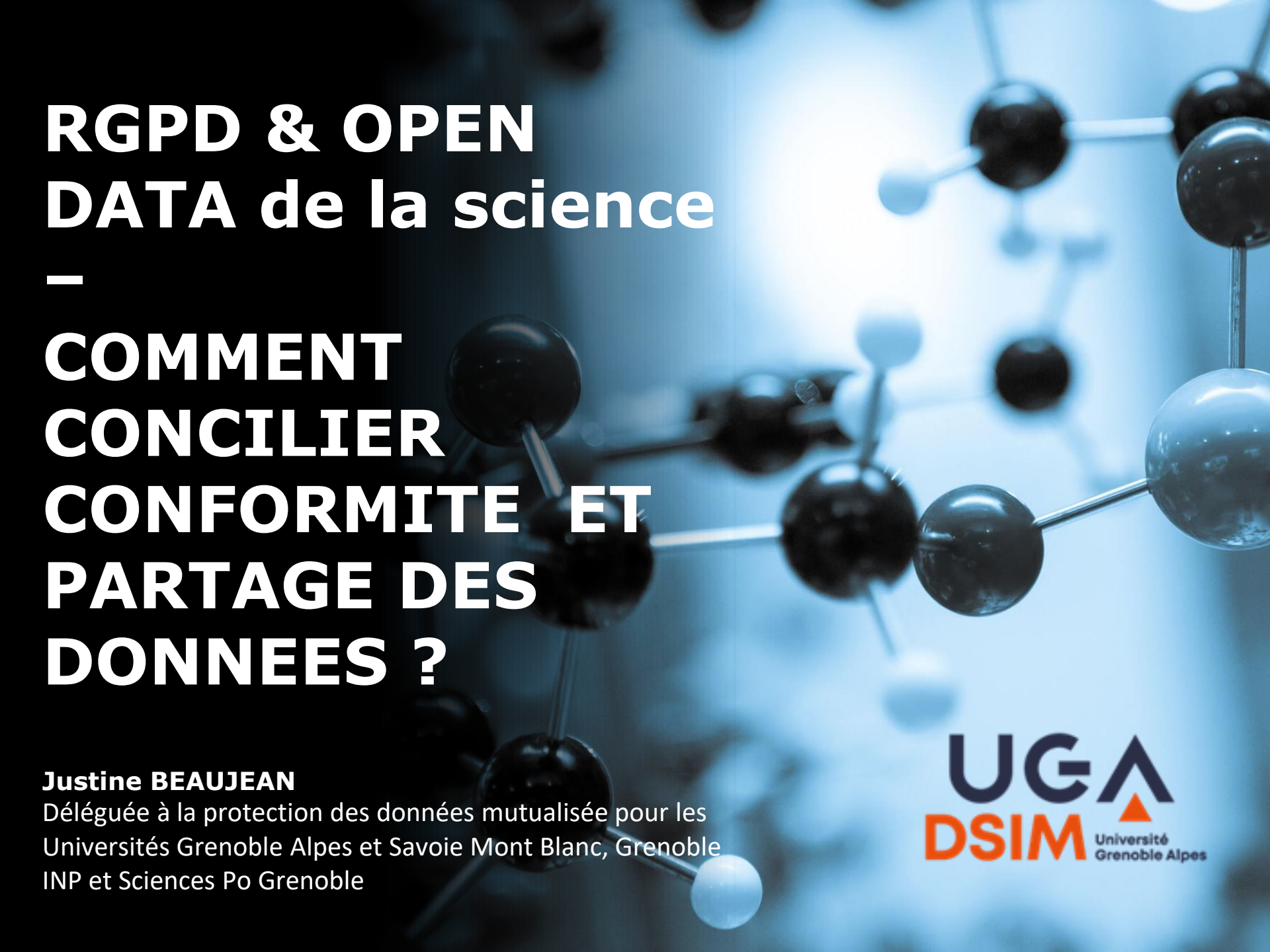
RGPD & OPEN DATA de la science – COMMENT CONCILIER CONFORMITE ET PARTAGE DES DONNEES ?



Justine BEAUJEAN
Déléguée à la protection des données mutualisée pour les Universités Grenoble Alpes et Savoie Mont Blanc, Grenoble INP et Sciences Po Grenoble

UGA DSIM Université Grenoble Alpes

Plan

- Contexte et enjeux
- Identifier une donnée à caractère personnel
- Principes fondamentaux du RGPD
- Dérogations et exceptions dans le cadre de la recherche
- L'anonymisation
- Les bonnes pratiques

Contexte et enjeux

L'Open data de la science peut-elle coexister avec une protection rigoureuse des données personnelles ?

Deux volontés :

- une accélération de la recherche grâce au partage des données
- Le respect des droits et libertés fondamentaux des personnes concernées, pour une recherche éthique.

L'article 9 du code civil (C. civ.) dispose « *chacun a droit au respect de sa vie privée* »

Article 1er al. 1 de la loi informatique et liberté du 1978 dispose « *L'informatique doit être au service de chaque citoyen. Son développement doit s'opérer dans le cadre de la coopération internationale. Elle ne doit porter atteinte ni à l'identité humaine, ni aux droits de l'homme, ni à la vie privée, ni aux libertés individuelles ou publiques*. »

Article 1er al.2 RGPD dispose « *Le présent règlement protège les libertés et droits fondamentaux des personnes physiques, et en particulier leur droit à la protection des données à caractère personnel. »*

Le RGPD c'est quoi ?

Le RGPD (Règlement général sur la protection des données), entré en vigueur depuis le 25 mai 2018, est un règlement européen visant à harmoniser les législations des Etats membres de l'UE en matière de protection des données à caractère personnel.

Ces dispositions ont été intégrées dans la **LIL** (Loi Informatique et Libertés), adoptée le 6 janvier 1978 et régulièrement mise à jour.

Ces textes constituent le cadre principal en France de la collecte, l'utilisation et la protection des données personnelles, sous le **contrôle de la CNIL (Commission Nationale Informatique et Liberté)** et s'articulent avec d'autres normes juridiques nationales et européennes (e-Privacy, Code de la santé publique, Code de l'éducation), qui apportent des règles spécifiques selon les secteurs, notamment en recherche et santé.

La Protection des données - Pourquoi ?

Pour prévenir des risques identifiés :

- **Atteinte à la vie privée** :

Divulgation non autorisée de données sensibles (ex. : données médicales, financières).

- **Usurpation d'identité** :

Exploitation frauduleuse de données pour accéder à des services ou effectuer des transactions au nom de la victime.

- **Discrimination et préjudice** :

Utilisation abusive de données (origine ethnique, orientation sexuelle, opinions politiques) pour discriminer ou porter atteinte à des droits fondamentaux.

- **Vol et exploitation des données** :

Par des cyberattaques ou des fuites, les données peuvent être revendues (dark web) ou utilisées à des fins malveillantes.

Respecter les principes fondamentaux du RGPD

- Ne collectez que les données vraiment nécessaires pour atteindre votre objectif (Minimisation) *Article 5(1)(c) du RGPD*
- Le traitement doit être licite, loyal et transparent et reposer sur une des 6 bases légales *article 5(1)(a) du RGPD*
- Soyez transparent (mention d'information*) Article 13 et 14 du RGPD*
- Organisez et facilitez l'exercice des droits des personnes *Article 12 du RGPD*
- Fixez des durées de conservation *Article 5(1)(e) du RGPD*
- Sécurisez les données et identifiez les risques *Article 32 RGPD*
- Inscrivez la mise en conformité dans une démarche continue (par exemple changement de finalité)

Établir préalablement **un plan de gestion des données** permettra de préparer la conformité en matière de protection des données à caractère personnel.

Respecter les principes fondamentaux du RGPD

En pratique :

- Chaque projet de recherche traitant des données personnelles doit faire l'objet d'une déclaration auprès du DPO afin d'inscrire et documenter la conformité de ce traitement sur un registre ;
- En cas de traitement impliquant des risques élevés pour les personnes, une analyse d'impact devra être réalisée ;
- En cas d'intervention d'un prestataire ou d'un partenaire, il convient de définir quels sont les droits et obligations de chaque partie dans une convention ;
- Rédiger une mention d'information pour les personnes concernées intégrant l'ensemble des mentions obligatoires.

Application du Privacy by design

Intégrer la conformité RGPD dès la conception de votre projet de recherche permet :

- D'anticiper les formalités et le schéma d'archivage des données ;
- Ne pas se retrouver coincé pour une publication si le projet n'a pas été enregistré dans le registre ;
- Rédiger une mention d'information conforme ;
- Définir et anticiper autant que possible l'ensemble des finalités du projet de recherche ;
- Utiliser des outils et services conformes au RGPD ;

Application du Privacy by design

Anticiper le schéma d'archivage des données

Réduction du risque de non-conformité

Promotion et réputation d'une recherche éthique

Facilitation

Utiliser des outils respectueux en matière de protection des données

UGA DSIM Université Grenoble Alpes

Dérogations pour s'adapter aux spécificités de la recherche

1. **Réutilisation des données pour de nouvelles finalités**
   Possible sans consentement explicite si compatible avec la finalité initiale.

2. **Conservation prolongée des données**
   Autorisée au-delà de la durée initiale, si nécessaire pour la recherche.

3. **Limitation des droits des personnes concernées**
   Les droits d'accès, rectification, effacement et opposition peuvent être restreints si leur exercice compromet l'objectif scientifique.

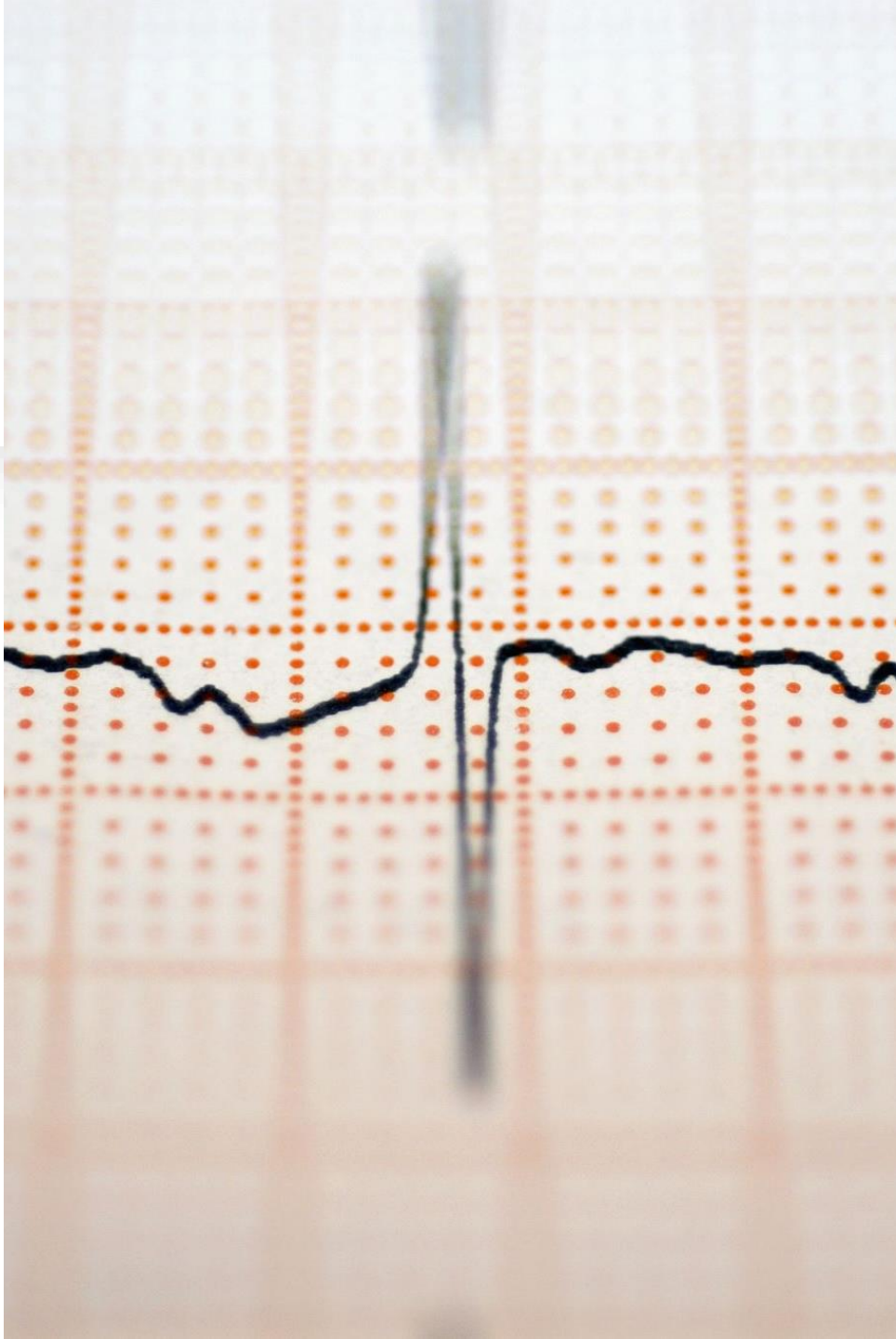
4. **Traitement des catégories particulières de données** (données sensibles) Autorisé si des garanties et mesures de sécurité appropriées sont mises en place (pseudonymisation, chiffrement, etc.).

5. **Information des personnes concernées**
   Si demande des **efforts disproportionnés** au regard des objectifs du traitement. Si cela compromettrait gravement la réalisation des objectifs de recherche.

Données anonymes et pseudonymes

La maîtrise de la distinction entre ces deux types de données est essentielle, car une donnée anonyme n'est plus une donnée personnelle et n'est en conséquence plus soumise au RGPD.

Quels sont les critères pour garantir le caractère anonyme des données ?



Données pseudonymes

La pseudonymisation consiste à remplacer les données directement identifiantes (nom, prénoms, etc.) par des données indirectement identifiantes (alias, numéro séquentiel, etc.).

Contrairement à l'anonymisation, la pseudonymisation est une opération **réversible.**

Conséquence : ces données sont soumises au RGPD et à la Loi informatique et liberté.

Partage de données possible, mais sous conditions.

« Le Contributeur s'assure, en amont de la publication, que cette dernière respecte le cadre juridique en vigueur notamment le RGPD, la loi Informatique et libertés, le Code de la propriété intellectuelle et le CRPA. »
https://www.data.gouv.fr/fr/pages/legal/cgu/

Les critères d'anonymisation

l'individualisation : il ne doit pas être possible d'isoler un individu dans le jeu de données ;

la corrélation : il ne doit pas être possible de relier entre eux des ensembles de données distincts concernant un même individu ;

l'inférence :il ne doit pas être possible de déduire, de façon quasi certaine, de nouvelles informations sur un individu.

Les techniques d’anonymisation

- **Généralisation** : Il s’agit de généraliser les attributs du jeu de données en modifiant leur échelle ou leur ordre de grandeur afin de s’assurer qu’ils soient communs à un ensemble de personnes.
- **Randomisation :** il s’agit de modifier délibérément certaines données pour dégrader leur précision tout en préservant l'intégrité globale. Par exemple, il est possible de permuter certaines informations entre sujets d’étude pour que la distribution des données reste exacte, sans associer les détails précis à la personne correcte.

Contreparties : les données sont plus ou moins altérées en fonction des objectifs poursuivis et des méthodes appliquées. Cette méthode qui ne convient pas à l’ensemble des projets.

En pratique comment anonymiser

Étape 1 : Examiner les catégories de données
Identifier les types de données à anonymiser :

- Données structurées (tableaux, bases relationnelles).
- Informations de géolocalisation, etc.

Étape 2 : Supprimer les identifiants directs et valeurs rares
Éliminer les éléments permettant une identification directe (noms, adresses, numéros).
Supprimer ou généraliser les valeurs rares ou uniques (par exemple, l'âge exact des centenaires).

Étape 3 : Distinguer informations importantes et secondaires
Prioriser les données essentielles à la recherche.
Identifier les informations inutiles ou supprimables conformément au principe de minimisation.

Étape 4 : Définir la finesse des données
Déterminer le niveau de détail acceptable pour chaque donnée conservée.

- Exemples :
  - Tranches d'âge (25-34 ans) au lieu d'âges précis.
  - Regrouper des lieux géographiques en régions plus larges.

Étape 5 : Valider le caractère anonyme du jeu de données avec son DPO

Le principe de l'anonymisation et de la pseudonymisation

PROCESSUS	PSEUDONYMISATION	ANONYMISATION
STATUT DES DONNÉES	Personnelles (restent indirectement identifiantes et donc soumises au RGPD et à la loi Informatique et Libertés	Anonymes
RÉUTILISATION DES DONNÉES	Sous conditions	Sans restriction
UTILITÉ DES DONNÉES	Préservée car pas d'altération du niveau de détail des données	Plus ou moins altérée en fonction des objectifs poursuivis et des méthodes appliquées
MÉTHODES À METTRE EN OEUVRE	Compteur, générateur de nombres aléatoires, fonction de hachage, chiffrement à clé secrète, etc.	Dépend des objectifs poursuivis : confidentialité différentielle, randomisation, k-anonymat, l-diversité, t-proximité, etc.
COMPLEXITÉ DE LA MISE EN OEUVRE	Simple à moyenne	Dépend des objectifs poursuivis : simple dans certains cas comme l'agrégation ou le comptage et complexe dans d'autres

Recherche scientifique (hors santé) : enjeux et avantages de l'anonymisation et de la pseudonymisation | CNIL

Conclusion et recommandations

- Savoir identifier une donnée à caractère personnel est essential pour **éviter des sanctions** et **garantir l'éthique** de la recherche ;
- Intégrer la démarche de conformité dès le début du projet ;
- Anonymiser avant de diffuser ou pseudonymiser en appliquant les principes du RGPD ;
- Sélectionner les bons outils ;
- Collaborer avec le DPO de l'établissement.

L’EQUIPE DPO UGA PRETE A VOUS ACCOMPAGNER

dpo@grenet.fr