Confidentialité des données statistiques et RGPD

Confidentialité des données : entre le prescrit et la pratique, l'enjeu de la maîtrise des risques de réidentification





INSEE - Unité des Affaires juridiques et contentieuses

INSEE
INSTITUT NATIONAL DE LA STATISTIQUE ET DES ÉTUDES ÉCONOMIQUES

Les enjeux de confidentialité relatifs aux personnes physiques

- Back to basics
- Qu'est-ce qu'une donnée à caractère personnel ?
- Les trois formes d'identification
- Quand peut-on parler de données anonymes ?
- Que peut/doit-on faire ?
- En guise de conclusion

Charte des droits fondamentaux de l'Union européenne

- Article 7, Respect de la vie privée et familiale :
  - Toute personne a droit au respect de sa vie privée et familiale, de son domicile et de ses communications
- Article 8, Protection des données à caractère personnel
  - Toute personne a droit à la protection des données à caractère personnel la concernant.

Qu’est-ce qu’une donnée à caractère personnel ?

- Information qu’on peut relier à une personne physique reconnaissable immédiatement ou après une recherche d’une ampleur « raisonnable » :
  - RGPD, article 4-1
  - RGPD, considérant 26
- Indépendant de la nature, de la sensibilité, de la « publicité » ou « notoriété » de la donnée

RGPD, article 4-1, « données à caractère personnel »

« toute information se rapportant à une personne physique identifiée ou identifiable (ci-après dénommée « personne concernée ») ; est réputée être une « personne physique identifiable » une personne physique qui peut être identifiée, directement ou indirectement, notamment par référence à un identifiant, tel qu'un nom, un numéro d'identification, des données de localisation, un identifiant en ligne, ou à un ou plusieurs éléments spécifiques propres à son identité physique, physiologique, génétique, psychique, économique, culturelle ou sociale »

RGPD, considérant 26, « les moyens raisonnables »

« *Pour déterminer si une personne physique est identifiable, il convient de prendre en considération l'ensemble des moyens raisonnablement susceptibles d'être utilisés par le responsable du traitement ou par toute autre personne pour identifier la personne physique directement ou indirectement, tels que le ciblage. Pour établir si des moyens sont raisonnablement susceptibles d'être utilisés pour identifier une personne physique, il convient de prendre en considération l'ensemble des facteurs objectifs, tels que le coût de l'identification et le temps nécessaire à celle-ci, en tenant compte des technologies disponibles au moment du traitement et de l'évolution de celles-ci.* »

Organigramme simplifié de la direction générale de l'Insee

Le comité de direction

Pascal Rivière
Inspection générale

Jean-Luc Tavernier
Directeur Général

Benoît Ourliac
Cabinet du directeur général

Secrétariat général **Alain Bayet**	Sphère Informatique **Olivier Lefebvre**	Département de la production et de l'infrastructure informatique	Département des applications et projets	
	Département des affaires financières et de la programmation des travaux et moyens	Département des ressources humaines	Département du cadre de vie et des conditions de travail	
	Centre statistique de Metz	Unité des affaires juridiques et contentieuses	Unité de la coordination des activités transversales	Pilotage et animation du réseau des Directeurs régionaux **Etienne Traynard**
Direction de la méthodologie et de la coordination statistique et Internationale **Sylvie Lagarde**	Département des méthodes statistiques	Unité qualité	Département de la coordination statistique et internationale	
Direction des statistiques d'entreprises **Cristel Colin**	Département des répertoires, infrastructures et statistiques structurelles	Département des statistiques de court terme	Département des synthèses sectorielles	
Direction des études et synthèses économiques **Didier Blanchet**	Département de la conjoncture	Département des études économiques	Département des comptes nationaux	
Direction des statistiques démographiques et sociales **Chantal Cases**	Département de la démographie	Département de l'emploi et des revenus d'activité	Département des ressources et des conditions de vie des ménages	Unité des prix à la consommation et des enquêtes ménages	Unité des études démographiques et sociales
Direction de la diffusion et de l'action régionale **Françoise Maurel**	Département Insee info service	Département de l'offre éditoriale	Département de l'action régionale	Unité des ressources documentaires et de l'archivage	Secrétariat du Conseil national de l'information statistique

Trois formes d'identification

- Données sont **directement** identifiantes :
  - Les données sont associées à un élément indiquant clairement l'identité de la personne (nom, prénom, email nominatif, photo, etc.)
- Données sont **indirectement** identifiantes :
  - Numéro client, NIR, numéro de téléphone, numéro d'immatriculation d'un véhicule
  - Les données ne permettent pas, **isolément**, de savoir immédiatement à qui correspondent les informations
  - Il est possible de retrouver l'identité de la personne par association avec une base de données détenue en interne ou **par un tiers**
- La personne physique peut être identifiée par **toute combinaison d'informations**

Identification par combinaison

Définition du traitement de données à caractère personnel

LES DONNÉES
À CARACTÈRE PERSONNEL

3 Les combinaisons d'informations

Une enquête ne sera pas rendue anonyme seulement parce qu'elle ne demande pas le nom et le prénom de la personne interrogée. C'est le cas des enquêtes qui concernent peu de personnes, surtout si ces personnes sont sélectionnées selon une caractéristique prédéterminée.
Une seule réponse peut permettre de retrouver l'identité d'une des personnes interrogées.

Même une enquête qui porte sur un grand nombre de personnes interrogées au hasard, dans la rue par exemple, peut contenir des réponses qui combinées les unes aux autres permettent de retrouver l'identité des sondés. C'est le cas lorsque les questions sont très nombreuses ou précises (par ex., "je suis née à Paris, dans le 14ème arrondissement, je travaille à la CNIL et je suis fan d'animaux, de jazz et de jeux vidéo").

LES DONNÉES
À CARACTÈRE PERSONNEL

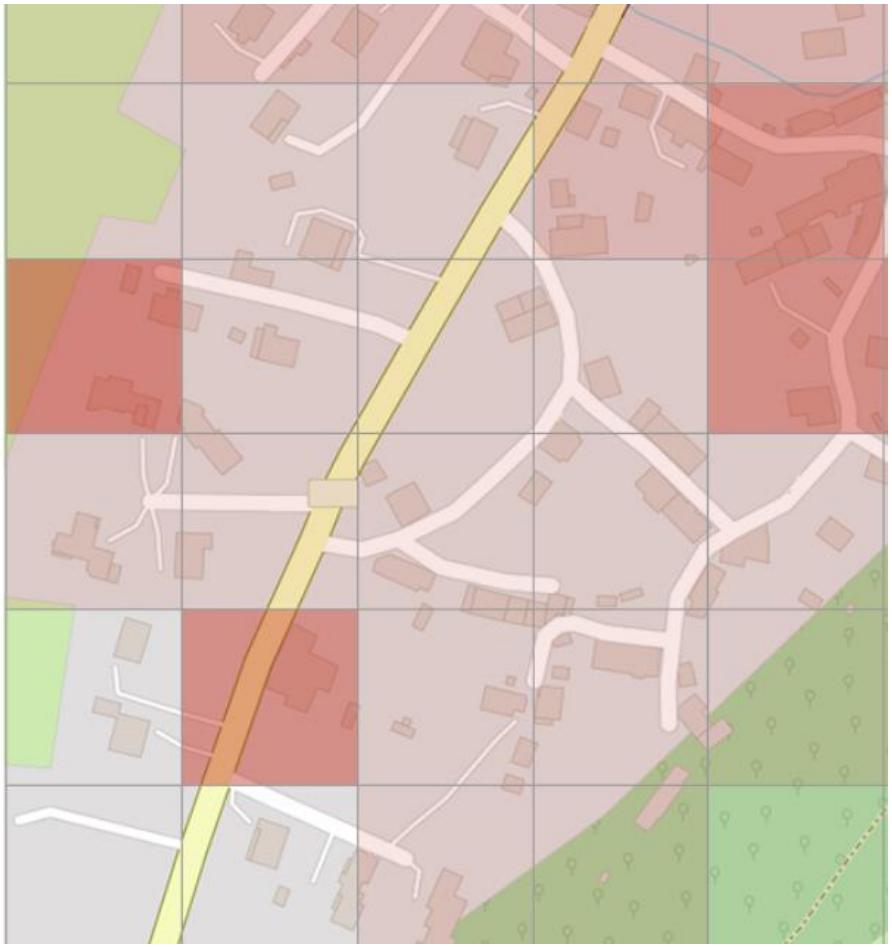
3 Les combinaisons d'informations

Par exemple en matière d'open data (ouverture et mise à disposition des données produites et collectées par les services publics) donnant la possibilité d'accéder à certaines informations personnelles.

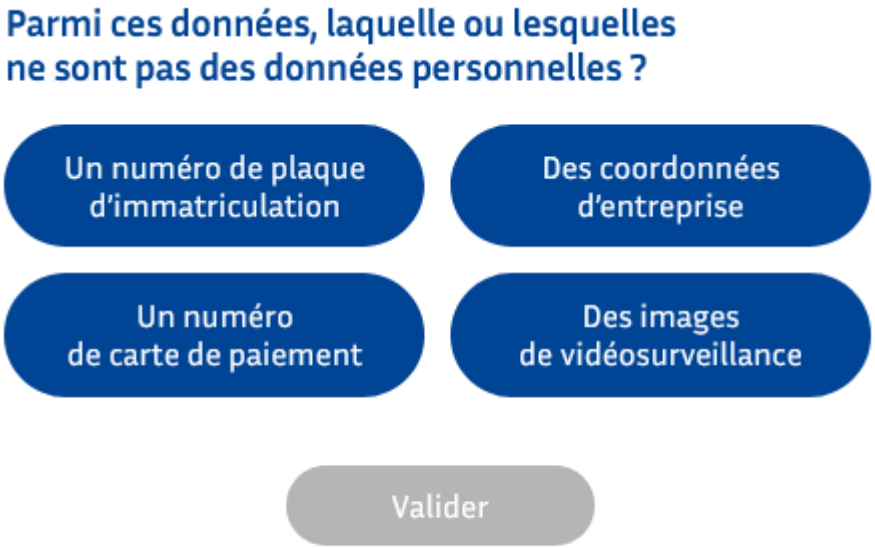
Supposons que la France soit quadrillée, **chaque carreau étant associé à des données socio-démographiques, dont l'imposition moyenne des habitants, la population active, le niveau de diplôme, etc.**

Dans les zones peu peuplées, **les carreaux, selon leur taille, peuvent ne comprendre qu'un nombre limité de foyers, voire un seul, dont il peut être aisé de retrouver l'adresse, l'identité, la physionomie générale**, notamment par un croisement de ces données avec des informations fournies par un moteur de recherche ou un service de cartographie en ligne.

Ainsi, selon la granularité du carreau, **la combinaison d'informations peut se révéler identifiante pour les personnes.**



Petit test


Parmi ces données, laquelle ou lesquelles ne sont pas des données personnelles ?
Un numéro de plaque d'immatriculation
Des coordonnées d'entreprise
Un numéro de carte de paiement
Des images de vidéosurveillance
Valider

Parmi ces données, laquelle ou lesquelles ne sont pas des données personnelles ?

- Un numéro de plaque d'immatriculation
- **Des coordonnées d'entreprise**
- Un numéro de carte de paiement
- Des images de vidéosurveillance

Parmi ces propositions, seules les coordonnées d'une entreprise ne sont pas considérées comme des données personnelles.
Il s'agit de données génériques (adresse postale, numéro de téléphone du standard, email de contact générique, etc.) qui ne correspondent pas à un individu.
Attention : certaines données non nominatives peuvent toutefois être identifiantes. Par exemple, le Directeur des ressources humaines de la société X.

Anonymisation et anonymat

- Anonymisation : lignes directrices du G29 (ex-groupe des CNIL européennes) :
  - Revue des méthodes d'anonymisation (randomisation, agrégation) : avantages/inconvénients, écueils à éviter
- Anonymat implique **irréversibilité** de l'anonymisation :
  - Données anonymisées, données traitées de façon à ne plus pouvoir être utilisées pour identifier une personne physique
  - Vaut aussi bien pour les tiers que pour le responsable de traitement, notamment s'il y a conservation des données initiales
- Des données qui demeurent identifiables moyennant des informations supplémentaires sont des données pseudonymisées et comme tels restent des données à caractère personnel
- Les risques résiduels de réidentification ne sont jamais exclus, compte tenu notamment de l'évolution des moyens disponibles :
  - Au mieux, on ne peut qu'atténuer les risques, en cherchant le meilleur compromis entre l'utilité des données et le droit de chacun à la protection de ses données et de sa vie privée
  - Une réévaluation permanente des risques est nécessaire
- **L'anonymisation, même en excluant tout risque de réidentification, n'exclut un impact sur la vie privée (profilage géographique, information faussement rattachée par randomisation, identification probabiliste), voire sur d'autres secrets (Strava)**

Que faire ?

- Viser la maîtrise du risque de réidentification en fonction de la sensibilité des données
- Le RT doit faire son maximum :
  - Ses obligations s'appliquent à des moyens proportionnés aux risques
- Le RT n'a pas d'obligation de résultat :
  - Le RGPD admet la possibilité de violation de données personnelles (article 33 du RGPD), avec des réponses graduées en fonction de la gravité (notification ou non de la CNIL, notification ou non des personnes concernées)
- Les conditions de mise en œuvre d'un traitement s'appuie sur le principe de **risques résiduels peu élevés** :
  - Analyse d'impact (articles 35 et 63 du RGPD)

RGPD, article 24-1, les obligations du responsable de traitement

« Compte tenu de la nature, de la portée, du contexte et des finalités du traitement ainsi que des risques, dont le degré de probabilité et de gravité varie, pour les droits et libertés des personnes physiques, le responsable du traitement ***met en œuvre des mesures techniques et organisationnelles appropriées*** *pour s'assurer et être en mesure de démontrer que le traitement est effectué conformément au présent règlement. Ces mesures sont réexaminées et actualisées si nécessaire. »*

Diffusion RP - Rochefourchat

- https://www.insee.fr/fr/statistiques/2011101?geo=COM-26274#chiffre-cle-6
- http://www.leparisien.fr/municipales-2014/comment-fait-on-pour-elire-un-maire-dans-une-commune-de-1-habitant-05-12-2013-3379003.php

Données locales

Article Le Parisien

Petits effectifs RP