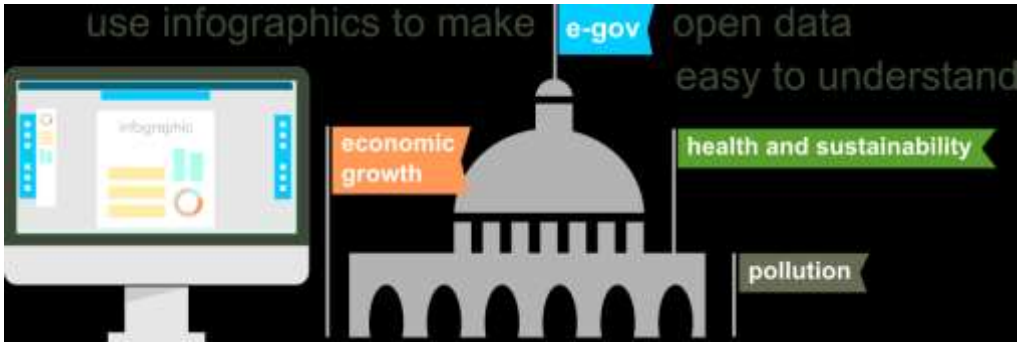
Anonymisation

Définition

L'**anonymisation de données** consiste à modifier le contenu ou la structure de ces données afin de rendre très difficile ou impossible la « *ré-identification* » des personnes (physiques ou morales) ou des entités concernées[1].
Le recours à l'anonymisation résulte d'un compromis entre la volonté de protection des données personnelles des individus et la nécessité de diffuser ou de partager ces données, souvent à des fins de recherche médicale ou de défense nationale. Ainsi l'anonymisation peut intervenir dans divers domaines : l'anonymisation des copies d'examen, des fichiers d'enquête pour des sondages ou dans le cadre de l'**open-data** sont des exemples. Dans le cas de l'open-data qui est une tendance en vogue et portant de nombreux bénéfices, l'anonymisation est indispensable.

L'open-data est une pratique qui consiste à publier et diffuser des informations produites par des collectivités, des services publics ou des entreprises à des fins d'intérêt général



Infographie open-data

Il faut garder à l'esprit que les données une fois rendues anonymes sortent du champ d'application des lois sur les données à caractère personnel, car théoriquement elles ne permettent plus l'identification des individus. Le problème qui se pose alors est qu'il est très ardu d'obtenir un ensemble de données anonymes en conservant une quantité d'informations suffisantes pour le traitement.

[1] Hull SC, Wilfond BS (2008), *What does it mean to be identifiable*

Techniques

1 Substitution ou pseudonymisation

Cette technique consiste à remplacer les identifiants d'une personne par des pseudonymes uniques qui peuvent être obtenus de différentes façons :

Une fonction de hachage permet de calculer l'empreinte d'une donnée fournie en entrée. Le résultat obtenu appelé *hash* est théoriquement unique et irréversible

- L'ensemble des pseudonymes (générés arbitrairement) est stocké dans une table de correspondance qui doit rester secrète et qui associe à chaque identifiant son pseudonyme. La sécurité de ce procédé est faible car elle repose exclusivement sur la confidentialité de la table de correspondance ; quiconque y accède est en mesure de retrouver les identifiants.
- Les pseudonymes sont calculés grâce des algorithmes de chiffrement ou des **fonctions de hachage**. Lorsqu'on a recours au chiffrement la réversibilité est toujours possible dès lors qu'on connait l'algorithme utilisé. Dans le cas du hachage le risque de réversibilité est faible mais il est possible de casser l'anonymisation en appliquant la fonction de hachage à l'ensemble des identifiants possibles ; il faut toutefois pour cela disposer d'une certaine puissance de calcul.
  La sécurité peut être renforcée par l'ajout de « sel », une chaîne de caractère secrète concaténée aux identifiants avant le hachage. On peut aller encore plus loin en appliquant un double hachage combiné à l'ajout de sel.

2 Ajout de bruit

Il s'agit ici de noyer l'information initiale pertinente parmi des données arbitraires ou de modifier les valeurs existantes. Cette méthode est très appropriée lorsque les informations contenues dans un jeu de données sont particulièrement sensibles pour les individus. Toutefois elle rend difficile la réutilisation des données en dehors de l'usage d'origine.

Dans le cas de l'ajout de bruit on peut par exemple stocker que des années de naissance au lieu des dates complètes ou modifier des valeurs telles que le poids (les diviser par 10)

3 Agrégation

Les données de même type sont rassemblées afin d'en extraire l'information nécessaire : par exemple pour estimer le taux de fréquentation, au lieu d'enregistrer les coordonnées de M. X, Y et Z ayant visité les locaux de l'entreprise E entre 8h et 9h, on pourra stocker que 3 personnes se sont présentées entre 8h et 9h.

Limites

Aucune technique d'anonymisation n'est infaillible. Il reste donc toujours possible de ré-identifier des personnes par recoupement et à partir de quelques données anonymisées. Citons à titre d'exemple le cas de l'entreprise de VOD Netflix qui, souhaitant améliorer son service, a publié les recommandations de 500 000 de ses clients en supprimant les identifiants des comptes. Cependant des informaticiens sont parvenus à identifier certains des clients en croisant les commentaires qu'ils avaient fait sur des films, la date de location et des appréciations qu'ils avaient faites sur d'autres forums.

Recommandations

Au vu des risques d'identification qui subsistent après l'anonymisation il est primordial de garder à l'esprit que les techniques d'anonymisation à elles seules ne peuvent garantir une protection maximale de la vie privée des personnes. Il est donc possible de combiner ces techniques ou de les complexifier pour renforcer la sécurité.
Il faut aussi effectuer des analyses de risques régulièrement afin de s'assurer que l'anonymat ne peut pas être compromis par une corrélation entre les données anonymisées et d'autres sources de données qui peuvent pas parfois échapper au contrôle du responsable du traitement.

Références

- Data Protection Working Party, Opinion 05/2014 on Anonymization Techniques
- http://www.senat.fr/rap/r13-469/r13-4697.html
- Référentiel AFAPDP des dispositifs d'anonymisation

Actualité IT et vie privée

Droit à l'oubli : Google dit non à la CNIL

Le bras de fer entre la firme de Moutain View et la CNIL se poursuit et s'intensifie après le refus de Google de se plier aux injonctions de la CNIL concernant le droit à l'oubli. En juin dernier, la CNIL mettait en demeure Google d'appliquer le droit à l'oubli à toutes les extensions du moteur de recherche, conformément à la décision de la Cour européenne de Justice de l'Union Européenne. Autrement dit, de ne pas se borner aux extensions de l'UE, quand bien même le droit à l'oubli ne concerne que les citoyens européens, mais à celles utilisées à travers le monde. Une page pourra être déréférencée pour des recherches effectuées depuis Google.com ou .ru et pas uniquement depuis Google.fr, ou .uk par exemple.
Selon Google une agence nationale de protection des données personnelles ne peut revendiquer une autorité à l'échelle mondiale.
Nous attendons la réponse de la CNIL
http://www.journaldugeek.com/2015/08/01/google-non-cnil-droit-a-loubli/

Espionnage de la vie privée : les batteries de smartphones s'y mettent

Le W3C travaille une nouvelle API qui permettrait aux navigateurs web de fournir une version allégée des pages visitées et ce en analysant l'état de la batterie des smartphones. Il sera donc possible grâce à du code javascript de suivre l'internaute pendant sa session de navigation, ce qui a amené de nombreux observateurs à souligner le côté intrusif d'une telle technologie et les risques qu'elle représente pour la vie privée des utilisateurs. Les API utilisant la batterie sont déjà inclues dans les navigateurs Opéra, Firefox et Chrome.
http://www.presse-citron.net/espionnage-de-la-vie-privee-meme-les-batteries-de-smartphones-vont-sy-mettre/

Facebook instaure le checkup de sécurité

Il est désormais possible pour les utilisateurs du numéro un des réseaux sociaux de vérifier le niveau de sécurité de leur compte grâce à une option insérée au-dessus du fil d'actualité. Elle permettra d'évaluer les paramètres de confidentialité, la force du mot de passe ou les alertes de connexion.
http://www.journaldugeek.com/2015/08/05/facebook-checkup-rapide-securite/

Do Not Track : vers une protection renforcée de la vie privée en ligne

Adblock, DuckDuckGo, et d'autres défenseurs de la vie privée se sont associés pour une refonte du standard Do Not Track, créé il y a plusieurs années afin de protéger les internautes et intégré dans la plupart des navigateurs. Ce standard qui a pour but d'empêcher la traçabilité lors de la navigation est malheureusement en perte de vitesse car non supporté par beaucoup d'acteurs du Web vivant de la publicité. Un compromis devra donc être trouvé afin de concilier les besoins des internautes, des éditeurs et des webmasters.
http://www.phonandroid.com/do-not-track-vers-une-plus-forte-protection-de-la-vie-privee-en-ligne.html

Empreintes digitales, des failles dans les smartphones

L'entreprise de sécurité FireEye Labs a démontré à l'occasion de la conférence BlackHat que l'utilisation de plus en plus répandue des empreintes digitales sur les smartphones (principalement Android) induit des failles de sécurité ; ces failles sont dues au fait que les empreintes sont stockées en clair sur le terminal et peuvent être accessibles depuis n'importe quelle application. Sont incriminés le HTC One Max et le Samsung Galaxy S5.

http://www.tomsguide.fr/actualite/empreintes-digitales-securite-smartphone,48116.html

ET si un crédit vous était refusé à cause de vos amis Facebook ?

Facebook a récemment déposé un brevet pour une technologie intitulée « Autorisation et authentification basées sur le réseau social d'un individu ». Au premier il s'agit juste d'authentifier un utilisateur grâce à ses amis, mais ce service pourrait être proposé à des banques. Il serait alors possible d'accorder ou de refuser un crédit ou un service à une personne donnée en se basant sur ses relations et en partant du principe que si vos amis sont de mauvais payeurs, alors vous l'êtes probablement aussi. Des entreprises telles que Lenddo et LendUp sont déjà positionnées sur ce secteur aux Etats-Unis. Espérons que comme beaucoup de brevets celui-ci ne sera pas utilisé.

http://lci.tf1.fr/high-tech/et-si-un-credit-vous-etait-refuse-a-cause-de-vos-amis-facebook-8643125.html

Directeur de publication

M. BILE Diéméléou

DG ARTCI

Rédacteur en chef

Mme Gbato Leontine Dorcas

Directeur de la protection des données personnel

Equipe de rédaction

M. Coulibaly Namongo

Chef de Département Technique

M. COULIBALY Nahoua

Chef de Service Expertise Technologique

M. KOUAME Cheik

Chef de Service Innovation et Prospective

Contacts:

Marcory Anoumabo, 18 BP 2203 Abidjan 18

autoritedeprotection@artci.ci

Tél: +225 20 34 43 54

Fax: +225 20 34 43 75