

# PRIVACY HORIZONS: TERRA INCOGNITA

29<sup>th</sup> International Conference of  
Data Protection and Privacy Commissioners

September 25 to 28, 2007  
Montreal, Canada



## LES HORIZONS DE LA PROTECTION DE LA VIE PRIVÉE : TERRA INCOGNITA

29<sup>e</sup> Conférence internationale des commissaires  
à la protection des données et de la vie privée

du 25 au 28 septembre 2007  
Montréal, Canada

# Les problèmes de dépersonnalisation posés par les données génétiques et génomiques

William W. Lowrance, Ph.D.  
([lowrance@iprolink.ch](mailto:lowrance@iprolink.ch))

26 septembre 2007

# Fondement physique des problèmes

Le génome humain :

- est élaboré et extrêmement détaillé
- influence de nombreuses caractéristiques personnelles
- est intrinsèque au corps humain
- reste le même durant toute la durée d'une vie
- est unique à une personne

Le génome complet est porté par l'ADN dans chaque cellule du corps, à l'exception des globules rouges.

Les données génomiques ressemblent à

...tttccgtatgcgtagccagacttaccctcctagtag...

– à raison de 3 000 000 000 « cellules de données » qui portent chacune les bases a, t, g, c.

La modification ou l'insertion de quelques a, t, g, c peut faire une grande différence, quelle que soit la manière dont on envisage le génome :

– comme un ruban de programmation dynamique

– comme un « code à barres » intrinsèque.

# Les données génétiques ressemblent à ce qui suit :

- à l'échelle d'une séquence :  
|ctag...ctcca|
- à l'échelle d'un gène : « Gène porteur du diabète SLC308A »
- à l'échelle du corps : « cheveux roux », « dysplasie rénale héréditaire »
- à l'échelle de la famille : ascendance familiale, antécédents familiaux en matière de santé, autres indicateurs.

À mon avis, l'interprétation la plus utile de la notion d'identifiabilité pour les données génomiques est la suivante :

*« L'identifiabilité » est l'associativité potentielle des données précises à une personne.*

# Moyens par lesquels les données génomiques peuvent être personnalisées

- a) par association : en associant un génotype à des données génotypiques identifiables (p. ex. données policières, militaires ou par lien familial)
- b) par couplage : en couplant des données génomiques et d'autres données connexes (sur la santé, sociales, etc.) avec d'autres données
- c) par profilage : en décrivant de manière « probabilistique » l'apparence, les facteurs liés à la santé ou d'autres caractéristiques possibles.

# Tactiques permettant de dépersonnaliser les données génomiques

- a) en limitant la proportion de renseignements relatifs à un génome qui sera communiquée
- b) en faussant de manière statistique les données avant de les communiquer
- c) en dépersonnalisant les données de manière irréversible
- d) en séparant les identifiants et en effectuant un codage par clé.

# Tactique a) en limitant la proportion de renseignements relatifs à un génome qui sera communiquée

- s'effectue déjà et peut protéger les données;
- toutefois, elle limite souvent l'utilité des données, puisqu'on ignore le plus souvent quelles seront les portions du génome qui seront pertinentes;
- il est difficile de déterminer la « bonne quantité » de renseignements à communiquer.

## Tactique b) en faussant de manière statistique les données avant de les communiquer

- peut se faire, par exemple en substituant de manière aléatoire des a/t/g/c;
- fausse presque toujours l'utilité des données, puisque la plupart des analyses s'effectuent sur le plan de détails précis.

## Tactique c) en dépersonnalisant les données de manière irréversible

- s'effectue parfois, par exemple lorsque l'objectif consiste à sonder l'occurrence d'un phénomène particulier ou à fournir des données à des fins éducatives.

## Tactique d) en séparant les identifiants et en effectuant un codage par clé

- fonctionne bien lorsqu'elle est effectuée de manière adéquate, que la clé est bien protégée et que l'utilisation de la clé pour reconstituer les données est strictement supervisée;
- est de plus en plus utilisée dans le cadre d'activités comme la recherche en santé.

# Dépersonnaliser ou pas?

Les motifs justifiant la dépersonnalisation des données et les manières de le faire sont fonction :

- de la nature des données;
- du consentement;
- des usages prévus;
- du potentiel de couplage avec des données génotypiques de référence ou d'autres données;
- des protections.

# Dépersonnalisation : autres solutions et compléments

- Fournir l'accès aux données au moyen de communications supervisées (régies par contrat, supervisées par un comité de gérance, etc.).
- Sanctionner l'utilisation malveillante des données (comme la repersonnalisation inappropriée) ou l'utilisation abusive des données (comme la discrimination).

## Mot de la fin

*La dépersonnalisation est une forme de protection pratique et essentielle — pour les données génomiques comme pour les autres formes de données — et il faut fortement encourager son utilisation!*

Référence générale : LOWRANCE et COLLINS,  
« Identifiability in genomic research »,  
*Science* 317, pages 600 à 602, 3 août 2007.