

Gestion des données sensibles et personnelles en sciences sociales

Pablo Diaz, FORS, CER-UNIL

La gestion des données, une affaire légale ?

Les **données personnelles** ne peuvent être traitées sans une certaine connaissance des lois.

- En effet, le traitement de ce type de données est encadré par des règlements en matière de **protection des données** (autodétermination informationnelle ≠ sécurité).
- Les règlements en matière de protection des données fournissent les **définitions**, les **limites**, les **droits** et **devoirs** de chacun·e.

Bases légales

Un certain nombre de textes fondamentaux garantissent l'auto-détermination informationnelle:

- La Déclaration Universelle des Droits de l'Homme (art. 12).
- La Convention Européenne des Droits de l'Homme (art. 8).

THE UNIVERSAL DECLARATION OF HUMAN RIGHTS



Adopted by the General Assembly of the United Nations in 1948, the Universal Declaration states fundamental rights and freedoms to which all human beings are entitled.

You have the responsibility to respect the rights of others.

We are all born free and equal.

Everyone is entitled to these rights no matter your race, religion, sex, language, or nationality.

Everyone has the right to life, freedom, and safety.

No one can take away any of your rights.

No one has the right to hold you in slavery.



No one has the right to torture you.



You have a right to be recognized everywhere as a person before the law.



We are all equal before the law and are entitled to equal protection of the law.



You have the right to seek legal help if your rights are violated.



No one has the right to wrongly imprison you or force you to leave your country.



You have a right to a fair, public trial.



Everyone is innocent until proven guilty.



You have the right to privacy. No one can interfere with your reputation, family, home, or correspondence.



You have the right to travel.



You have the right to seek asylum in another country if you are persecuted in your own.



Everyone has the right to a nationality.



All consenting adults have the right to marry and to raise a family.



You have the right to own property.



Everyone has the right to belong to a religion.



You have the right to think and voice your opinions freely.



Everyone has the right to gather as a peaceful assembly.



You have the right to participate in the governance of your country, either directly or by helping to choose representatives in free and genuine elections.



You have the right to social security and are entitled to economic, social, and cultural help from your government.



Every adult has the right to a job, a fair wage, and membership in a trade union.



You have the right to leisure and rest from work.



Everyone has the right to an adequate standard of living for themselves and their family.



Everyone has the right to an education.



Everyone has the right to freely participate in the culture and scientific advancement of their community, and their intellectual property as artist or scientist should be protected.



We are all entitled to a social order in which we may enjoy these rights.



Everyone's rights and freedoms should be protected unless they obstruct the rights and freedoms of others.



No State, group, or person can use this Declaration to deny the rights and freedoms of others.



This is a simplified version of the UDHR. For the complete text, visit www.un.org

Bases légales

En Suisse il y a des lois à deux niveaux:

- Fédéral (**Constitution, FADP**)
- Cantonal (**LPrD** pour Vaud.)

En UE il y a le Règlement Général sur la Protection des Données (**RGPD**).

Au niveau international, 128 pays (sur 194) ont des lois en matière de protection des données.

https://unctad.org/en/Pages/DTL/STI_and_ICTs/ICT4D-Legislation/eCom-Data-Protection-Laws.aspx

Bases légales

En Suisse, c'est le « statut légal » de la personne qui traite les données qui détermine quelles lois s'appliquent (cantonales ou fédérales)

Personne privée	Institution fédérale	Institution cantonale
Lois fédérales	Lois fédérales EPFL, ETH, FORS, etc.	Lois cantonales UNIL, UNIGE, UNIBE, UNIZH, HES, etc.

La logique \neq territoire

Bases légales

LPD

– Art. 2 Champ d'application

¹ La présente loi régit le traitement de données concernant des personnes physiques et morales effectué par:

- a. des personnes privées;
- b. des organes fédéraux.

LPrD

✓ Art. 3 Champ d'application [4, ...]

¹ La présente loi s'applique à tout traitement de données des personnes physiques ou morales.

² Sont soumis à la présente loi les entités suivantes :

- a. le Grand Conseil ;
- b. le Conseil d'Etat et son administration ;
- c. l'Ordre judiciaire et son administration ;
- cbis. la Cour des comptes ;
- d. les communes, ainsi que les ententes, associations, fédérations, fractions et agglomérations de communes ;
- e. les personnes physiques et morales auxquelles le canton ou une commune confie des tâches publiques, dans l'exécution desdites tâches.

Bases légales

- Nous voyageons, avec les obligations légales de notre pays d'établissement (institutionnel).
- À l'étranger, en plus de nos lois nationales / locales, les lois du pays visité s'appliquent (que les participant·e·s y soient résidents ou non).
- Dans l'UE, lorsque l'on **cible** des populations/individus entant que ressortissants/résidents européens, les lois de leur pays de résidence s'applique, même si tout le traitement est organisé et mis en œuvre depuis la Suisse (questionnaire en ligne, téléphone, etc.)

Données personnelles

Traitement de données personnelles

Il est important de savoir si l'on traite ou non des données personnelles !

- À partir du moment où l'on effectue un traitement de données personnelles, les lois en matière de **protection des données** s'appliquent !
- C'est pourquoi l'anonymisation est souvent mise en avant comme LA solution...

Traitement de données personnelles

Traitement : **Toute opération ou ensemble d'opérations** effectuées ou non à l'aide de procédés automatisés et appliqués à des données personnelles, notamment la collecte, l'enregistrement, l'organisation, la conservation, l'adaptation ou la modification, l'extraction, la consultation, l'utilisation, la communication, la diffusion ou toute autre forme de mise à disposition, le rapprochement ou l'interconnexion, ainsi que le verrouillage, l'effacement ou la destruction

(art. 4, al. 5 LPrD).

Données personnelles

Toute information qui se rapporte à une personne identifiée ou identifiable est une donnée personnelle. Il s'agit d'une notion très large.

- Les données personnelles les plus « évidentes » : noms, adresses (postale, IP, mail, etc.), numéros de téléphone, photos, etc.
- Quelques données personnes plus subtiles : opinions et idées originales, distinctions / prix, profession atypique, etc.

Les données personnelles peuvent être objectives ou subjectives (pas de limite à leur nature).

Exemples de données personnelles

Contact

Prénom: Pablo
Nom de famille: Diaz
Téléphone: 123456
email: pablo@diaz.ch

Citations

“Il faut agir aussi vite
que possible, mais aussi
lentement que
nécessaire »

Photo



The way you dance is unique, and computers can tell it's you

Nearly everyone responds to music with movement, whether through subtle toe-tapping or an all-out boogie. A recent discovery shows that our dance style is almost always the same, regardless of the type of music, and a computer can identify the dancer with astounding accuracy.



Studying how people move to music is a powerful tool for researchers looking to understand how and why music affects us the way it does. Over the last few years, researchers at the Centre for Interdisciplinary Music Research at the University of Jyväskylä in Finland have used motion capture technology—the same kind used in Hollywood—to learn that your dance moves say a lot about you, such as how extroverted or neurotic you are, what mood you happen to be in, and even how much you empathize with other people.

Chinese ‘gait recognition’ tech IDs people by how they walk

By DAKE KANG November 6, 2018



 Click to copy

BEIJING (AP) — Chinese authorities have begun deploying a new surveillance tool: “gait recognition” software that uses people’s body shapes and how they walk to identify them, even when their faces are hidden from cameras.

Even Anonymous Coders Leave Fingerprints

Researchers have repeatedly shown that writing samples, even those in artificial languages, contain a unique fingerprint that's hard to hide.

RESEARCHERS WHO STUDY stylometry—the statistical analysis of linguistic style—have long known that writing is a unique, individualistic process. The vocabulary you select, your syntax, and your grammatical decisions leave behind a signature. Automated tools can now accurately identify the author of a forum post for example, as long as they have adequate training data to work with. But newer research shows that stylometry can also apply to *artificial* language samples, like code. Software developers, it turns out, leave behind a fingerprint as well.

Rachel Greenstadt, an associate professor of computer science at Drexel University, and Aylin Caliskan, Greenstadt's former PhD student and now an assistant professor at George Washington University, have found that code, like other forms of stylistic expression, are not anonymous. At the DefCon hacking conference Friday, the pair will present a number of studies they've conducted using machine learning techniques to de-anonymize the authors of code samples. Their work, some of which was funded by and conducted in collaboration with the United States Army Research Laboratory, could be useful in a plagiarism dispute, for instance, but also has privacy implications, especially for the thousands of developers who contribute open source code to the world.

Artificial intelligence unmask anonymous chess players

Software that identifies unique styles poses privacy risks

By **Matthew Hutson**

Think your bishop's opening, queen's gambit, and pawn play are unique? A new artificial intelligence (AI) algorithm has got your chess style pegged.

AI software can already identify people by their voices or handwriting. Now, an AI has shown it can tag people based on their chess-playing behavior, an advance in the field of "stylometrics" that could help computers be better chess teachers or more humanlike in their game play. Alarming, the system could also be used to help identify and track people who think their online behavior is anonymous.

"Privacy threats are growing rapidly," says Alexandra Wood, a lawyer at the Berkman Klein Center for Internet & Society at Harvard University. She says studies like this one, when conducted responsibly, are useful because they "shed light on a significant mode of privacy loss."

Chess-playing software, such as Deep Blue and AlphaZero, has long been superhuman. But Ashton Anderson, a computer scientist at the University of Toronto and principal investigator of the new project, says the chess engines play almost an "alien style"

That required the system to recognize what was distinctive about each player's style.

The researchers tested the system by seeing how well it distinguished one player from another. They gave the system 100 games from each of about 3000 known players, and 100 fresh games from a mystery player. To make the task harder, they hid the first 15 moves of each game. The system looked for the best match and identified the mystery player 86% of the time, the researchers reported last month at the Conference on Neural Information Processing Systems (NeurIPS). "We didn't quite believe the results," says Reid McIlroy-Young, a student in Anderson's lab and the paper's primary author. A non-AI method was only 28% accurate.

"The work is really cool," says Noam Brown, a research scientist at Meta (the

parent company of Facebook) who has developed superhuman poker bots. He looks forward to chess bots that mimic Magnus Carlsen, the reigning world champion, and says style-aware AI could transform other computer interactions. "There's a lot of interest in chatbots, where you can have a chatbot that would speak in the style of Albert Einstein or something," he says.





**(DON'T) WORRY YOU ARE
UNIQUE!**

Données sensibles

Il s'agit de données personnelles se rapportant :

- aux opinions ou activités religieuses, philosophiques, politiques ou syndicales ainsi qu'à une origine ethnique;
- à la sphère intime de la personne, en particulier à son état psychique, mental ou physique;
- aux mesures d'aide individuelles découlant des législations sociales;
- aux poursuites ou sanctions pénales et administratives

⚠ La liste fournie par la LPrD est exhaustive

⚠ Selon le contexte, tout type d'information (nom, photo, etc.) peut être sensible (ce n'est pas déterminé à priori)

Exemples de données sensibles

Contact

First name: Sebastián
Last name: Calfuqueo
Genre: trans
Emploi: Syndicaliste

Citation

« L'humanité serait mieux armée si elle pensait plus souvent à ce qui est important pour elle. Nous nous énerverions moins à gagner toujours plus et à grimper l'échelle sociale ».

Photo



Donnée sensible ?



Donnée sensible ?



Kayan people (Myanmar)

From Wikipedia, the free encyclopedia

(Redirected from [Kayan \(Burma\)](#))

For the ethnic group from [Borneo](#), see [Kayan people \(Borneo\)](#).

The **Kayan** are a sub-group of Red Karen ([Karenni people](#)), [Tibeto-Burman ethnic minority](#) of [Myanmar](#) (Burma). The Kayan consists of the following groups: Kayan Lahwi (also called **Padaung**, ပဒေါင်း[bə̀də̀u̯i̯]), Kayan Ka Khaung (Gekho), Kayan Lahta, Kayan Ka Ngan. Kayan Gebar, Kayan Kakhi and, sometimes, [Bwe people \(Kayaw\)](#). They are distinct from, and not to be confused with, the [Kayan people of Borneo](#).

Padaung (Yan Pa Doung) is a [Shan](#) term for the Kayan Lahwi (the group in which women wear the brass [neck rings](#)). The Kayan residents in Mae Hong Son Province in Northern Thailand refer to themselves as Kayan and object to being called Padaung. In *The Hardy Padaungs* (1967) [Khin Maung Nyunt](#), one of the first authors to use the term "Kayan", says that the Padaung prefer to be called Kayan.^[1] On the other hand, [Pascal Khoo Thwe](#) calls his people Padaung in his 2002 memoir, *From the Land of Green Ghosts: A Burmese Odyssey*.^[2]

In the late 1980s and early 1990s due to conflict with the military regime in Myanmar, many Kayan tribes fled to the [Thai](#) border area.^[3] Among the refugee camps set up there was a Long Neck section, which became a tourist site, self-sufficient on tourist revenue and not needing financial assistance.^[4]

Kayan

ကယန်း



Quid des données “publiques” ?

- Il arrive que des personnes rendent des données personnelles disponible via les réseaux sociaux
- Ces données restent des données personnelles !
- Il est important de vérifier les politiques de confidentialité des lieux (sites) où est faite la collecte.

Quid des données “publiques” ?

Attention donc aux termes et conditions contenus dans les **conditions générales d'utilisation (CGU)** des sites et des plateformes dont le contenu vous intéresse. En effet, chaque **site web** a la possibilité d'adapter ces dispositions à sa propre situation afin de permettre une meilleure protection de ses intérêts.

A titre d'exemple, LinkedIn interdit la réutilisation des données que le site contient à des fins commerciales. Les CGU de LinkedIn disposent que :

« Vous vous engagez à ne pas développer, prendre en charge ou utiliser des logiciels, des dispositifs, des scripts, des robots ou tout autre moyen ou processus (notamment des robots d'indexation, des modules d'extension de navigateur et compléments, ou toute autre technologie) visant à effectuer du web scraping des Services ou à copier par ailleurs des profils et d'autres données des Services ».

<https://www.captaincontrat.com/gestion/le-scraping-de-donnees-sur-linkedin-est-il-legal-me-kouhana>

Protection des données et recherche

En sciences sociales, il est quasi inévitable de travailler avec des données personnelles (et sensibles)

- Profils socio-démographiques
- Récits de vie
- Enquêtes longitudinales
- Enregistrements, etc.

Quelques questions

Puis-je collecter des données personnelles ?

- Les institutions publiques travaillent selon une logique de légalité (≠licéité).
- Il faut une base légale pour traiter des données personnelles.
- Le consentement est souvent la seule base légale disponible pour traiter des données sensibles.
- Lorsque les données sont disponibles à tout un chacun sans « opposition » -> base légale pour le traitement.
- Il faut également un **but** !

Puis-je collecter des données personnelles ?

Chapitre II Dispositions générales

Section I Principes

Art. 5 Légalité

¹ Les données personnelles ne peuvent être traitées que si :

- a. une base légale l'autorise ou
- b. leur traitement sert à l'accomplissement d'une tâche publique.

² Les données sensibles ne peuvent être traitées que si :

- a. une loi au sens formel le prévoit expressément,
- b. l'accomplissement d'une tâche clairement définie dans une loi au sens formel l'exige absolument, ou
- c. la personne concernée y a consenti ou a rendu ses données accessibles à tout un chacun.

Dois-je obtenir un consentement ?

- Même s'il y a des cas où l'on pourrait travailler sans consentement (données personnelles non sensibles)...
- Les lois sur la protection des données rendent obligatoire **l'information** des personnes pour toute collecte de données personnelles les concernant, même lorsque celles-ci sont collectées auprès de tiers.
- De plus, la participation à un projet de recherche est toujours facultatif (\neq d'obligation)
- Travailler sans consentement n'est pas plus facile !

Quelle forme doit prendre le consentement ?

Le consentement peut être oral ou écrit.

- Garder une trace est utile
- Parfois l'oral est la meilleure solution
- Ce qui compte = être compris

Attention : LRH -> Écrit

Quelle forme doit prendre le consentement ?

Lorsqu'il y a collecte de données personnelles, le consentement peut être implicite.

- Répondre à un questionnaire, un entretien, etc.

Lorsqu'il y a collecte de données sensibles, le consentement doit être explicite

- Seul un oui est un oui.

Comment informer ?

Le consentement doit être libre et **éclairé**

- Pour donner un avis il faut **savoir**
- L'information doit être **claire** (accessible/adaptée)
- Il faut suffisamment de **temps** pour réfléchir

Comment informer ?

Les informations obligatoires sont :

- Identité du responsable de traitement (PI + institution)
- La finalité du traitement (but de la collecte)
- Catégorie des destinataires si communication
- Droit d'accès aux données
- Droit de refuser de participer

Comment informer ?

Les informations suivantes sont conseillées:

- L'anonymisation/codage éventuel des données
- Politique de sécurité (codes, stockage, etc.)
- La préservation et réutilisation des données.
- La possibilité d'être informé des résultats

Comment informer ?

Il est conseillé d'opter pour une approche descriptive et de ne pas faire de promesses qui ne peuvent pas être tenues.

- Ne pas promettre l'anonymisation, le secret, etc.
- Ne pas parler d'impossibilité de réidentification
- Expliquer de manière transparente ce qui sera fait avec les données (suppression, etc.)

Puis-je pratiquer l'observation à couvert ?

Collecter des données personnelles à couvert est très difficile à justifier !

- Un·e chercheur·e qui voudrait collecter des données personnelles sans informer les personnes pourrait le faire à condition qu'il ou elle soit en mesure de prouver l'existence d'un **intérêt public prépondérant**.
- Prouver ce type d'intérêts est très difficile.

L'observation à couvert pourrait être possible si **pas de collecte de données personnelles**

Puis-je conserver des données personnelles indéfiniment ?

- On ne peut collecter / traiter des données que pour un but précis (principe de **finalité**).
- La **finalité** du traitement est une notion clé dans le sens où c'est elle qui détermine la limite temporelle de conservation des données personnelles.
- Si une conservation à moyen-long terme est envisagée, celle-ci doit être **cohérente avec le but de la collecte**. C'est la raison pour laquelle il convient d'informer clairement les participants à un projet de recherche de ses **objectifs** et de sa **temporalité**.

Puis-je communiquer des données personnelles à des tiers ?

Oui mais les personnes dont les données personnelles sont collectées doivent être informées de toute communication à des tiers (≠ membres de l'équipe de recherche).

- Sous-traitance pour les transcriptions, analyses
- Stockage des données sur le **cloud** (téléphone, Switch, etc.)
- Partage à d'autres équipes de recherche
- Archivage et réutilisation !



There is no cloud
it's just someone else's computer

Puis-je communiquer des données personnelles à des tiers ?

La communication transfrontière soulève des défis

- Communiquer vers entreprise US = problème
- Éviter les services Microsoft pour le traitement de données personnelles et sensibles !

Solution « patch » = chiffrement

[file:///Users/pdiaz/Downloads/20211115 Staatenliste f.pdf](file:///Users/pdiaz/Downloads/20211115_Staatenliste_f.pdf)

Puis-je partager des données personnelles via une archive ?

Oui, à certaines conditions:

- Consentement adéquat (ne pas « fermer » la porte)
- Le projet de recherche doit être conçu de manière à ce qu'il intègre le partage des données parmi ses objectifs (p.ex. histoire).
- La communication ne peut se faire qu'à des chercheur·e·s et à des fins de recherche (contrôle de l'accès)
- Attention aux sujets qui s'écarteraient trop de la problématique initiale

Merci!