

Limites éthiques du partage des données de la recherche

Jean-Gabriel Ganascia

26 novembre 2024

Sorbonne Université, équipe ACASA du LIP6
ex-président du COMETS (comité d'éthique du CNRS)

Président comité d'éthique de France Travail

Président du comité d'orientation du CHEC

Jean-Gabriel.Ganascia@lip6.fr



Synoptique

1. Du partage au partage des données
2. Éthique vs. épistémologie des données
3. Limites du partage
4. PAPA revisité



1

DU PARTAGE AU PARTAGE DES DONNÉES

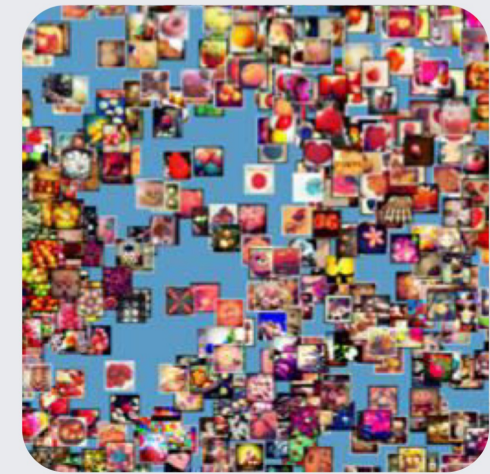
La logique et les bienfaits du partage

Dans le partage matériel,
où l'on perd l'usage de ce
que l'on met en partage...

La légende de Saint-martin



... Dans la partage
numérique, rien n'est perdu.



Photos Shared
Per Day

2×10^9

Le partage numérique est « gagnant – gagnant »!

Histoire du partage des données

Partage de données scientifique:

- Origine: biologie, *Bermuda principles*, 1996
- Déclaration de Berlin, 2003 et 2005
- CNRS, 2014

Vient des communautés scientifiques

- Astrophysique, biologie, ...
- Médecine ?? – industrie pharmaceutique?

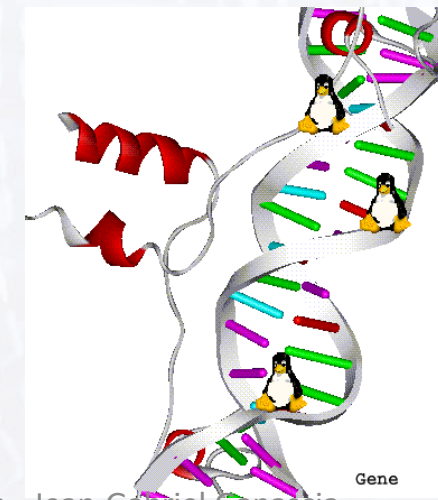
Objectifs

- Mettre la science à la disposition de tous
- Empêcher les fraudes en permettant à tous de reproduire les expériences

Partage des données: de quoi parle-t-on?

Données de recherche

- Nature des données
 - astronomie, génome, images, parole, ...
- Types de données:
 - Données primaires:
 - radiotélescope 7 Po/mn!
 - Données secondaires
 - dérivées des données primaires
 - Filtrées, annotées, enrichies interprétées
 - Métadonnées
 - Structurent, gèrent, facilite l'accès
 - Tracent l'origine des données



Accès ouvert – « Open Access »

Définition: fournir accès en ligne gratuit à l'« **information scientifique** » - déclaration de Budapest, 2002

« Information scientifique »:

- a) *Articles de recherche*, éventuellement évalués par les pairs
- b) *Données de recherche* ayant servi aux publications (données primaires et secondaires)

Gratuité du partage ...

Le partage des données ne coûte rien...

... mais il n'est pas nécessairement gratuit!

- « Data Sharing »: *partage données scientifique*
Gratuit!
- « Open data »: *données publiques*
ouvertes, “Open means **anyone** can **freely access, use, modify, and share** for **any purpose** (subject, at most, to requirements that preserve provenance and openness).” Mais, les utilisateurs peuvent vendre le produit de l'utilisation des données.
- « Open access »: *accès ouvert (publications)*
Gratuit pour le lecteur, mais pas pour l'auteur...
Modèle gratuit – green
Modèle payant – gold (auteur payeur)



2

ÉTHIQUE VS. ÉPISTÉMOLOGIE DES DONNÉES

Ethique des données



CNIL – années 70:

- Finalité
- Proportionnalité

Le groupe PAPA – années 80:

- *Privacy*
- *Accuracy*
- *Property*
- *Accessibility*

Éthique de l'information:

Mal = entropie

Big Data — Masses de données

RGPD *Règlement Général sur la Protection des Données*

Data Act

Principes Data Act

- **Circulation:** éliminer les obstacles au partage
- **Protection des données**
- **Transparence des traitements**
- **Rétribution équitable**
- **Interopérabilité**



Principes RGPD

- **Finalité:** une organisation doit donner un objectif légitime avant de collecter des données personnelles
- **Transparence:** une organisation doit notifier les utilisateurs de la collecte et du partage de leur données personnelles.
- **Respects de droits de la personne:** les utilisateurs ont le droit d'accepter ou de rejeter la collecte de leurs données personnelles. Elles peuvent aussi en demander la correction ou l'effacement



Epistémologie vs. éthique des données

- Non conforme au RGPD



- Non conforme au HLEGIA et au AI act



- Non conforme au Data Act



L'âge du Peta Byte!

- Pas d'échantillonnage
- Enregistrement de tout, sans modèle a priori
- Pas d'analyse sémantique ou causale
- Promotion du partage
- "Correlation is enough"

3

LIMITES DU PARTAGE

Risques

- **Données sensibles et données à caractère personnel**
 - Médecine, interfaces, images, sciences sociales, etc.
 - Anonymisation: difficultés → croisement
- **Atteintes à l'intimité**
 - Interfaces cerveau-ordinateur



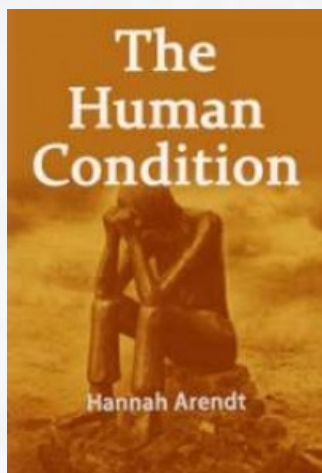
Privé, vie privée et intime

Distinguer

1. Intimité

1. Saint-Augustin
2. Jean-Jacques Rousseau

2. Extimité



3. Privé

4. Vie privée

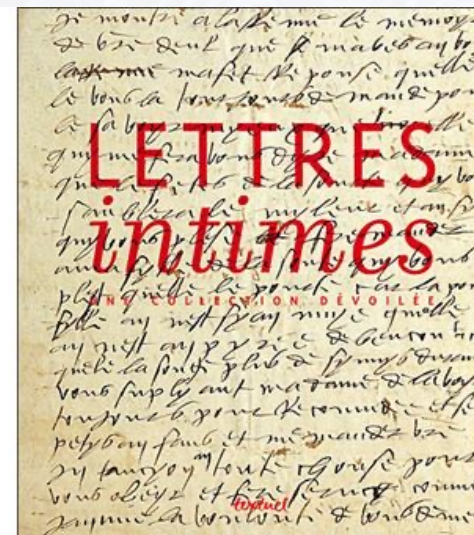
1. Warren & Brandeis – 1890
2. Totalitarisme – « 1984 » Orwell
3. Réseaux sociaux...

FRANÇOIS JULLIEN

De l'intime

Loin du bruyant
Amour

GRASSET



HARVARD LAW REVIEW.

VOL. IV. DECEMBER 15, 1890. NO. 5.

THE RIGHT TO PRIVACY.

"It could be done only on principles of private justice, moral fitness, and public convenience, which, when applied to a new subject, make common law without a precedent; much more when received and approved by usage."

WILSON, J., in *Miller v. Taylor*, 4 Burr. 521, 522.

THAT the individual shall have full protection in person and in property is a principle as old as the common law; but it has been found necessary from time to time to define anew the exact nature and extent of such protection. Political, social, and economic changes entail the recognition of new rights, and the common law, in its eternal youth, grows to meet the demands of society. Thus, in very early times, the law gave a remedy only



Atteintes à l'intimité

Elon Musk Neuralink l'esprit de la ruche

- “Over time I think we will probably see a **closer merger of biological intelligence and digital intelligence.**”
- “it's mostly about the bandwidth, the speed of the connection between your brain and **the digital version of yourself,** particularly output. ”



L'esprit de la ruche: Collaboration des esprits qui partagent les connaissances et les opinions les uns avec les autres
La "dentelle neurale" (I. Bank)

Risques

- **Données sensibles et données à caractère personnel**
 - Médecine, interfaces, images, sciences sociales, etc.
 - Anonymisation: difficultés → croisement
- **Atteintes à l'intimité**
 - Interfaces cerveau-ordinateur
- **Siphonage des données**



Données d'origine spatiale

- Agriculture
- Géographie
- Urbanisme
- Étude du climat
- Écologie
- Défense
- ...



Risques

- **Données sensibles et données à caractère personnel**
 - Médecine, interfaces, images, sciences sociales, etc.
 - Anonymisation: difficultés → croisement
- **Atteintes à l'intimité**
 - Interfaces cerveau-ordinateur
- **Siphonage des données**
- **Piratage, dépossession**
- **Biais dans les données**
- **Équité du partage!**
 - Utilisation par les grands acteurs du numérique ou par d'autres pays...



4

PAPA REVISITÉE

Éthique des données scientifiques

Groupe PAPA (1986)

- **Privacy**
 - *Lorsqu'on travaille sur des données non personnelles, ce n'est pas pertinent*
- **Accuracy**
 - *Est-ce requis avec les masses de données?*
- **Property**
 - *Qu'est-ce que ça signifie à l'âge du partage de données et de l'accès ouvert?*
- **Accessibility**



Éthique des données – PAPA revisitée

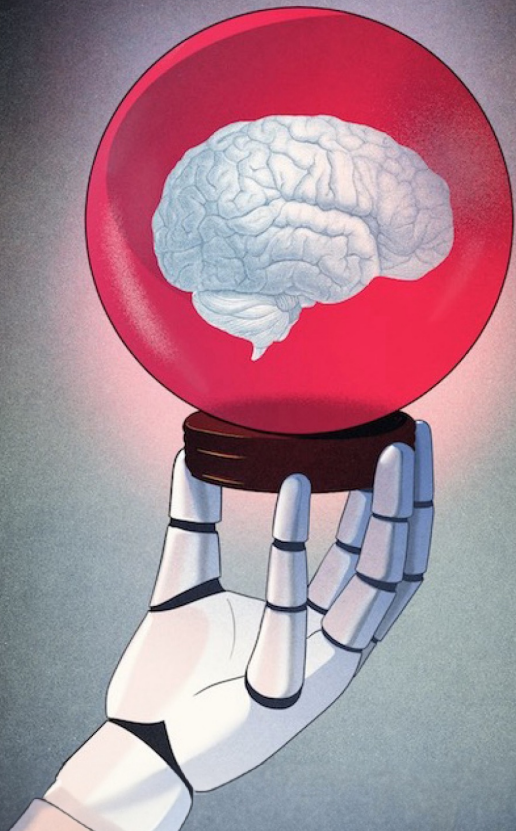
PAPA revisitée

- **Pertinency**
 - *La pertinence scientifique des données demande toujours à être questionnée*
- **Accuracy**
 - *Doit être qualifiée et certifiée...*
- **Place of Origin**
 - *Métadonnées indiquant l'origine des données*
- **Accessibility**
 - *En précisant les modalités...*



JEAN-GABRIEL GANASCIA

L'IA. expliquée aux humains



Seuil

Merci!