

PRIVACY HORIZONS: TERRA INCOGNITA

29th 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^e Conférence internationale des commissaires
à la protection des données et de la vie privée

du 25 au 28 septembre 2007
Montréal, Canada

Désidentification des données, risques et résolution

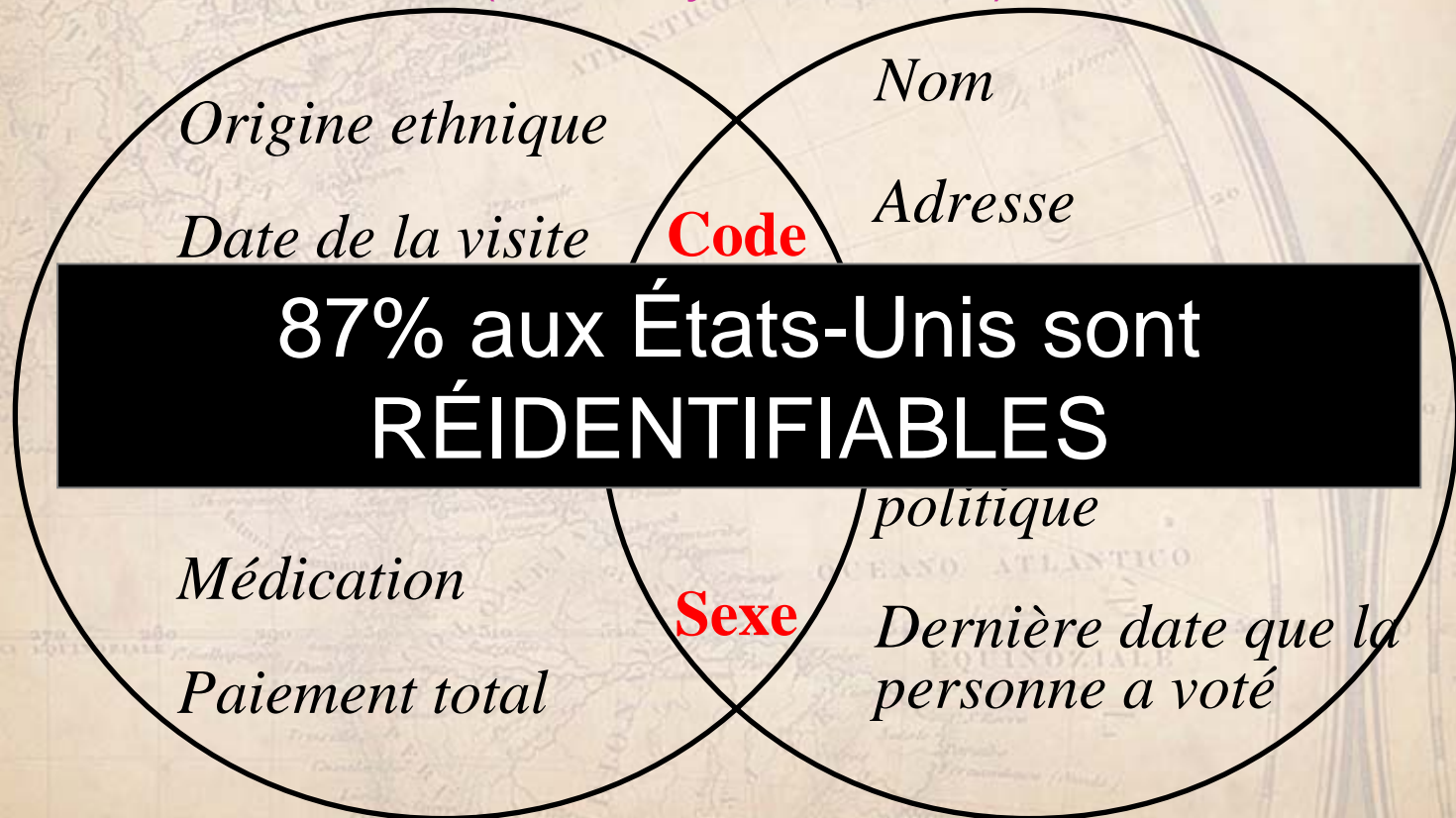
Bradley Malin, Ph.D.

Professeur adjoint

Vanderbilt University

Désidentifié ne veut pas dire anonyme

(Sweeney 1998, 2000)

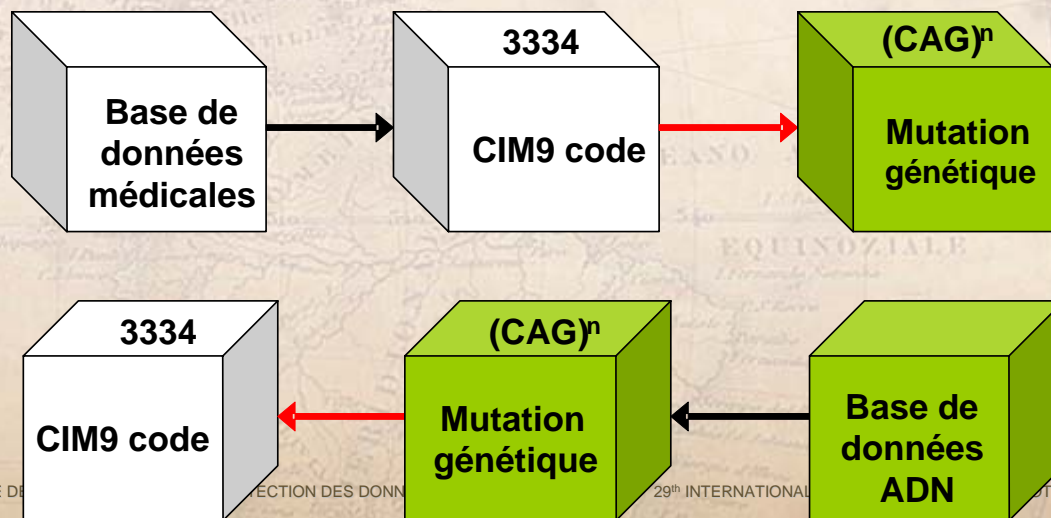


**Données sur les
congés des patients**

**Liste
d'électeurs**

Réidentification par empreintes génétiques

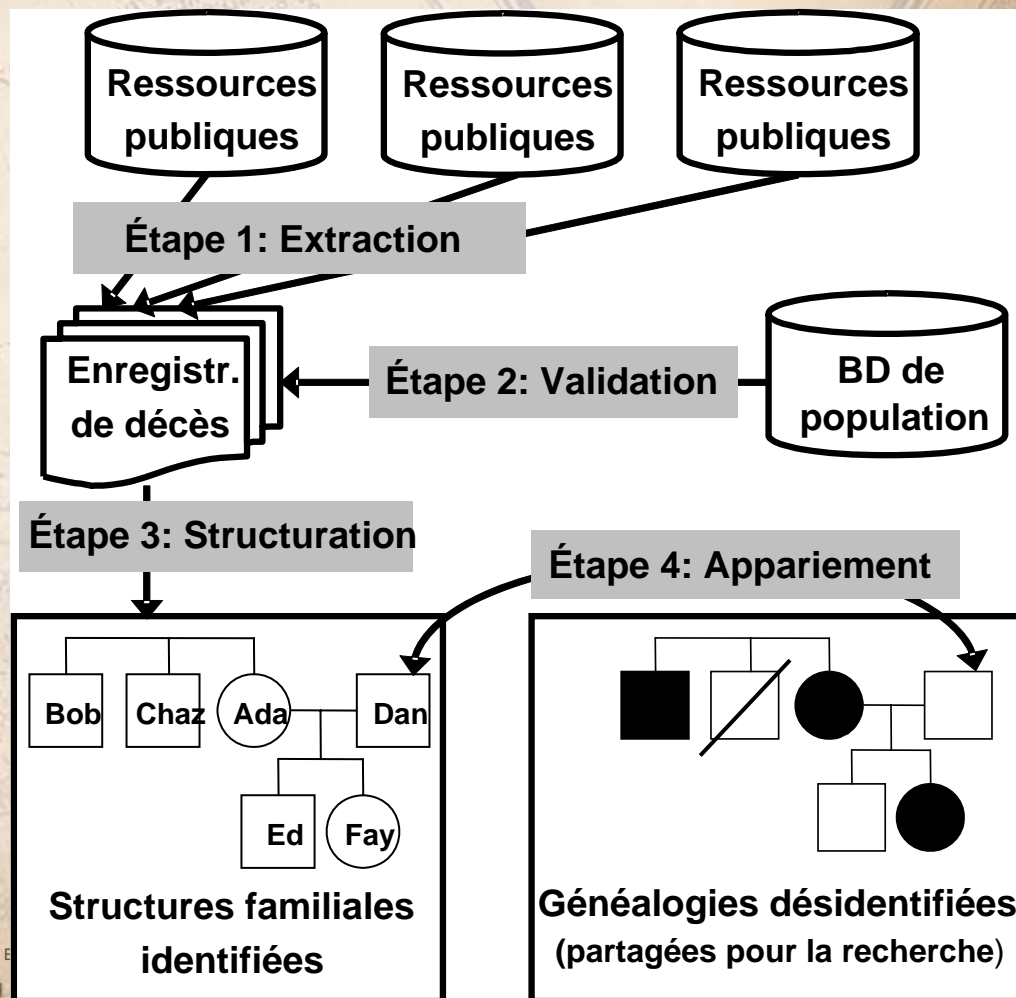
- Plusieurs des technologies de protection de la vie privée visant le génome permettent la réidentification par l'ADN (*Malin 2005*)
- L'ADN est réidentifié par des méthodes automatiques, par exemple :
 - Génotype – Inférence phénotype (*Malin & Sweeney, 2000, 2002*)



Réidentification généalogique

(Malin 2006)

- *IdentiFamily*:
 - logiciel qui apparie des généalogies désidentifiées à des personnes désignées
 - Se sert d'informations accessibles au public, p.ex., nécrologies, enregistrements de décès, la base de données de la *Social Security Death Index* pour établir des généalogies




Réidentification généalogique

(Malin 2006)

WyomingNews.com
Wyoming
Tribune-Eagle

Monday

[Home](#) | [News](#) | [Sports](#) | [Obituaries](#) | [Classifieds](#) | [Community](#) | [Real Estate](#) 

Subscribe
Today!



WTE
POLL

News

[Featured](#)
[Local News](#)
[National News](#)
[Outdoors](#)
[Entertainment](#)
[Special Projects](#)
[Story Archive](#)

Legislature 2007

[News](#)
[WTE Editorials](#)
[Guest Editorials](#)



Legislature 2007

[Click Here](#)

OBITUARIES

Richard R. Mann

1924-2007

Richard R. Mann, 82, of Cheyenne died Jan. 12 at Cheyenne Regional Medical Center.

He was born June 29, 1924, in Allentown, Pa., and had lived here since 1956.

Mr. Mann served in the Army Air Corp during World War II in South Africa and Italy.

He retired as a flight engineer for the Wyoming Air National Guard.

Mr. Mann was a member of St. Mary's Catholic Church, Elks, Moose and the Knights of Columbus, where he had been a past grand knight and state deputy.

He is survived by two sons, Gerald Mann and Thomas Mann, both of Cheyenne; seven daughters, Teresa Johnson, Kathryn Schroll, Judith Oldenburg, Cheryl Thibault, and Jon Cameron, all of Cheyenne, Lou Ann Golden of Sidney, Neb., and Kimberly Byron of Littleton, Colo.; his companion, Katie Heaton of Cheyenne; 25 grandchildren and two great-grandchildren.

He was preceded in death by his wife of more than 50 years, Patricia A. Mann; two daughters, Mary Constance Grant and Jeanane Rhodes; his parents, Russell and Viola Mann; two brothers, Roland Mann and Robert Mann; and a sister, Rochelle Behrandt.

Vulnérabilité du système

(Malin, JAMIA 2005)

Systemes de protection de la vie privée

| | | | | |
|-------------|----------------------|------------------------------|------------------------|--|
| Quoi | Tiers de confiance | Tiers de semi-confiance | Dénominalisation | Désidentification |
| Où | deCode Genetics Inc. | University of Gent, Custodix | Université de Montreal | University of Utah, University of Sydney, Australian National University |

Vulnérabilité à une attaque

| | | | | |
|------------------------------|--|--|--|--|
| Structures de famille | | | | |
| Sillage | | | | |
| Génotype-Phénotype | | | | |
| Dictionnaire | | | | |



Vulnérable



Pas vulnérable

La modification des données n'assure pas la protection

- Science Magazine (*Lin et al, 2004*)
 - < 100 SNPs rendent l'ADN unique

AVERTISSEMENT :

***L'unicité ne garantit pas que la vie
privée sera compromise***

- De nombreuses perturbations sont requises pour empêcher l'appariement
- Garder les enregistrements sous scellés

***Protection de la vie privée
(Perturbation)***

Modèle formel de réidentification

Banque de données
biologiques déidentifiées

| |
|------------|
| aaactaaga |
| cacaccatg |
| tatatgatgt |

Condition nécessaire

**1. Rendre les données non
uniques**

Condition nécessaire

**2. Certifier l'absence de
chemin d'appariement**

**Déjà dans le
domaine public**

Données nominatives

| |
|--------------|
| John Doe |
| Jane Doe |
| Jeremiah Doe |

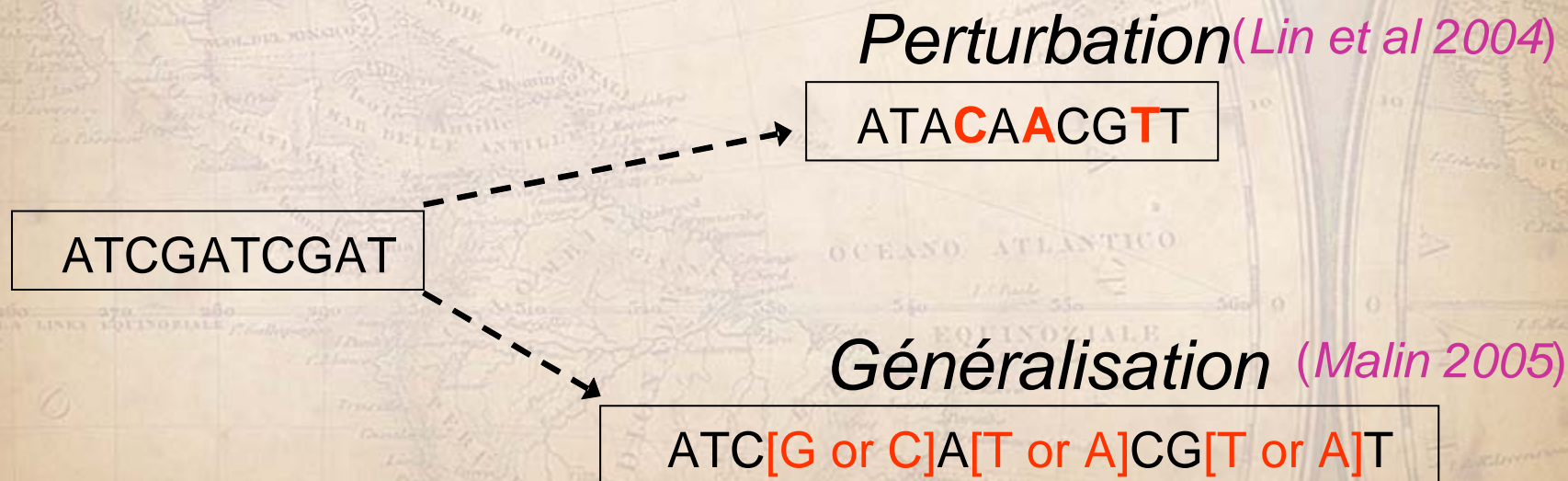
Condition nécessaire
UNICITÉ

Protection formelle

- **k -mappage** (Sweeney, 2002)
 - Chaque enregistrement partagé désigne au moins k unités dans la population
- **k -anonymat** (Sweeney, 2002)
 - Chaque enregistrement partagé est semblable à au moins $k-1$ autres enregistrements
- **k -non appariement** (Malin 2006)
 - Chaque enregistrement partagé s'apparie à au moins k identités à travers de son sillage
 - Satisfait le modèle de protection par k -mappage

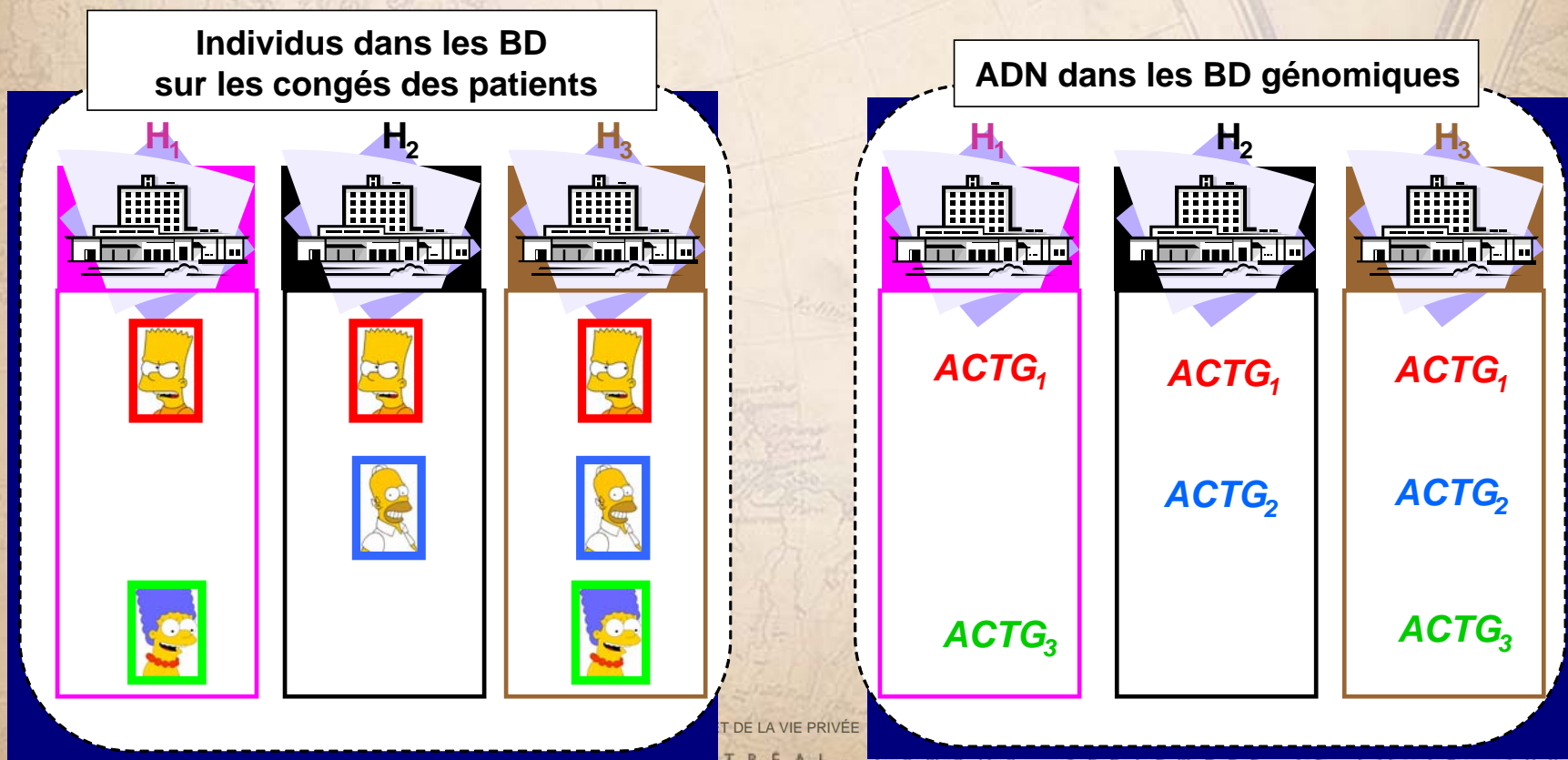
Au delà des protections *ad hoc*

- La perturbation ne garantit pas la protection de la vie privée
- Alternative : Généralisation des données



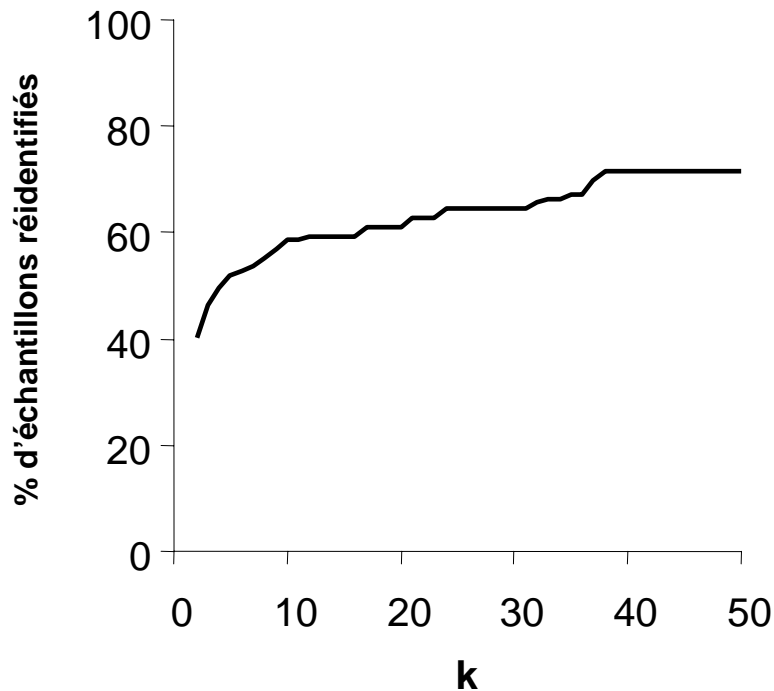
Savoir qui vous êtes à partir d'où vous avez été (« sillage »)

(Malin & Sweeney, 2001; 2004, Malin & Airoidi 2006)

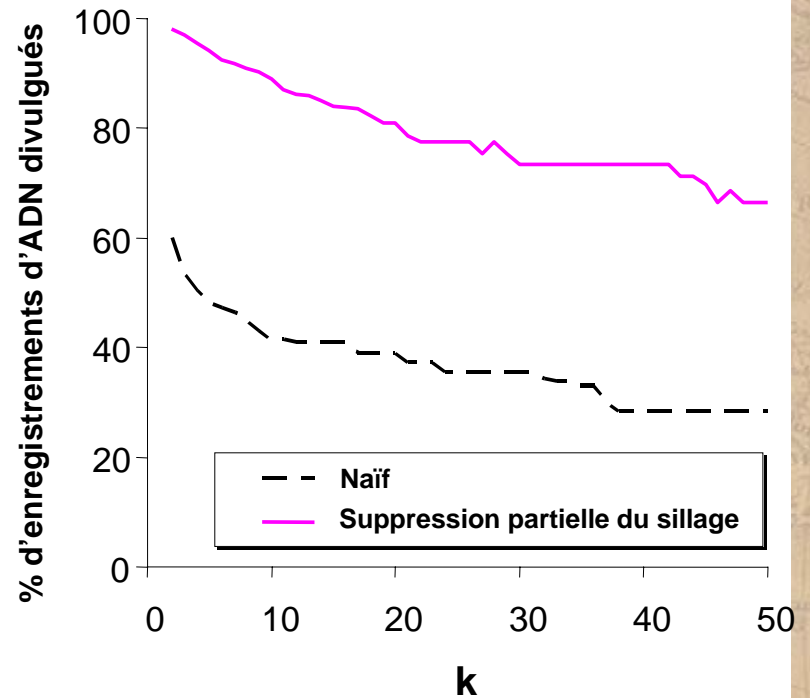


Empêcher le sillage : population avec la fibrose kystique

(1149 échantillons)



AVANT STRANON
100% échantillons en entrepôt



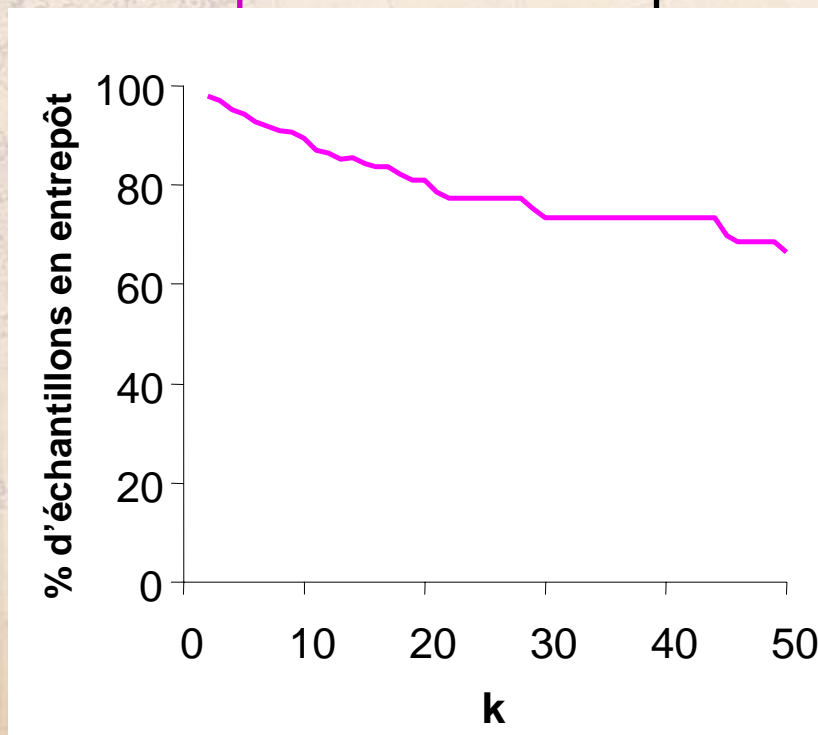
APRÈS STRANON
0% échantillons k-réidentifiés

Utilité : Risque quantifié

Réglage
forcé

Réglage
initial

Quantité
demandée



- Modification au risque de réidentification
- Déplace le fardeau de l'accroissement du risque vers l'analyste requérant
- Lie les modèles légal et informatique