

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

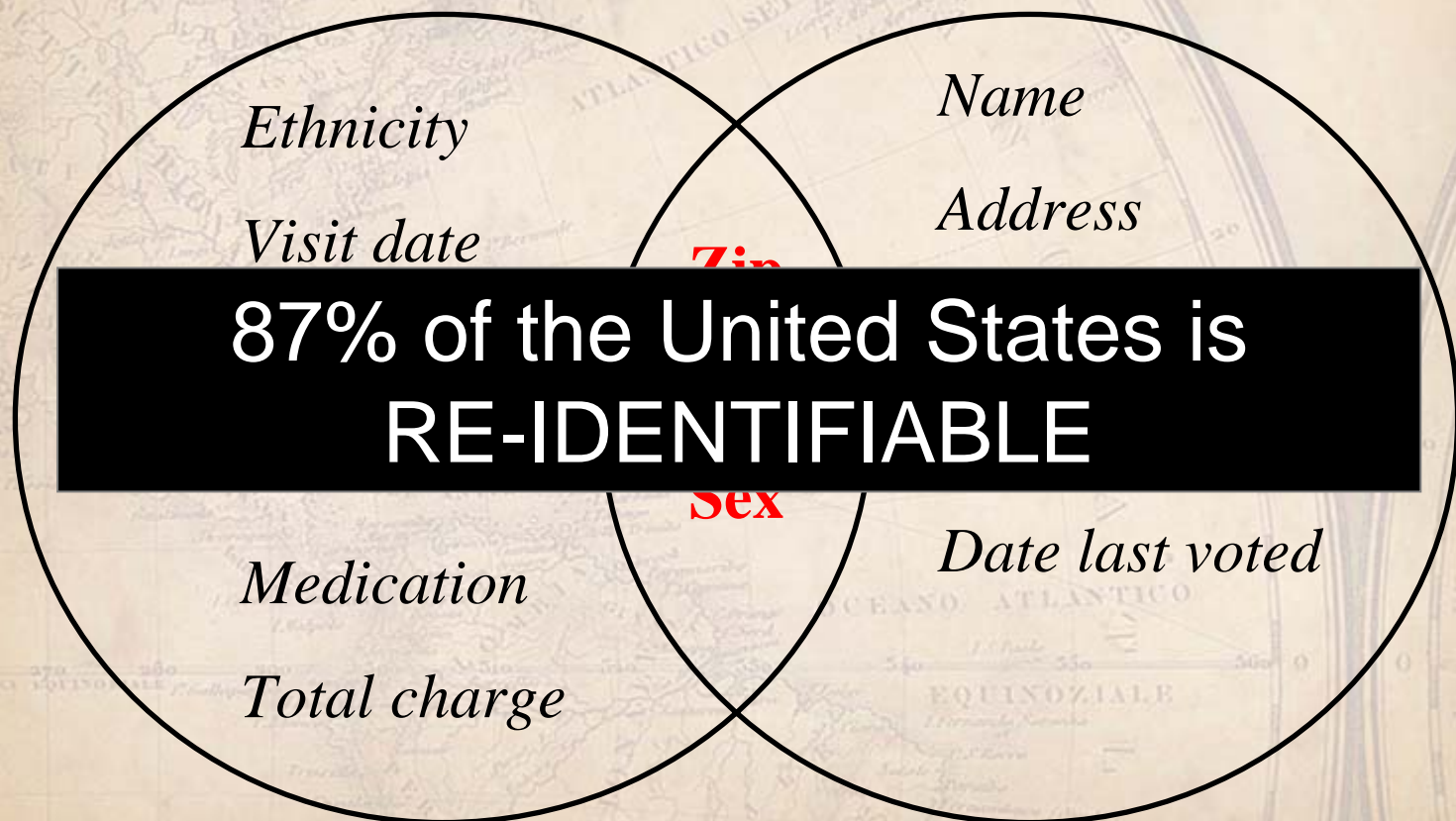
du 25 au 28 septembre 2007
Montréal, Canada

De-identification Risk and Resolution

Bradley Malin, Ph.D.
Assistant Professor
Vanderbilt University

De-identified is not Anonymous

(Sweeney 1998, 2000)

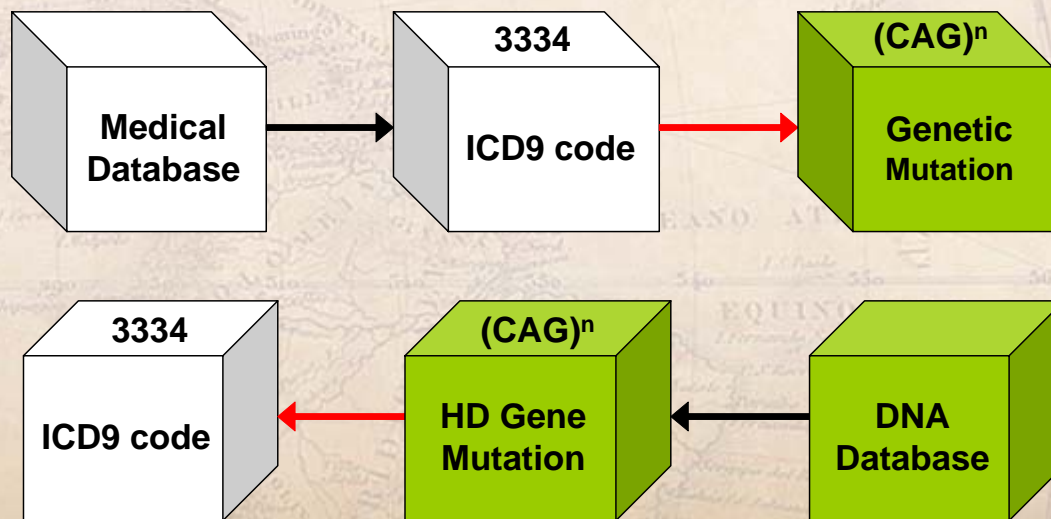


Hospital Discharge Data

Voter List

DNA Re-identification

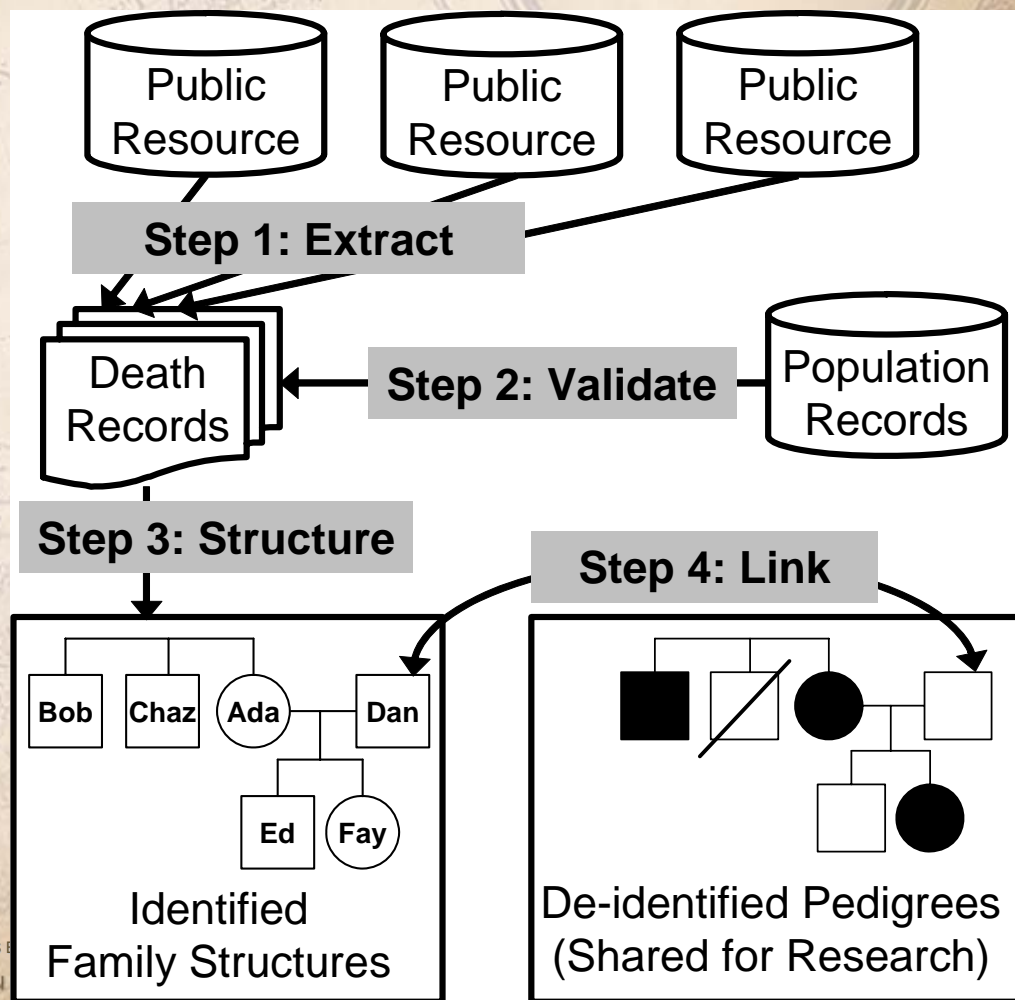
- Many deployed genomic privacy technologies leave DNA susceptible to re-identification (*Malin 2005*)
- DNA is re-identified by automated methods, such as:
 - Genotype – Phenotype Inference (*Malin & Sweeney, 2000, 2002*)



Genealogy Re-identification

(Malin 2006)

- *IdentiFamily*:
 - software that links de-identified pedigrees to named individuals
 - Uses publicly available information, such as obituaries, death records, and the Social Security Death Index database to build genealogies




Genealogy Re-identification

(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.

System Susceptibility

(Malin, JAMIA 2005)

Privacy Protection Systems

What	Trusted Third Party	Semi-Trusted Third Party	Denominalization	De-identification
Where	deCode Genetics Inc.	University of Gent, Custodix	University of Montreal	University of Utah, University of Sydney, Australian National University

Susceptibility to Attack

Family Structures	■	□	■	■
Trails	□	■	□	■
Genotype-Phenotype	■	■	■	■
Dictionary	■	■	□	□

Susceptible

Not Susceptible

Altering Data Does not Guarantee Protection

- Science Magazine (*Lin et al, 2004*)

- < 100 “SNPs” make DNA unique

DISCLAIMER:

Uniqueness Does not Guarantee Privacy will be Compromised

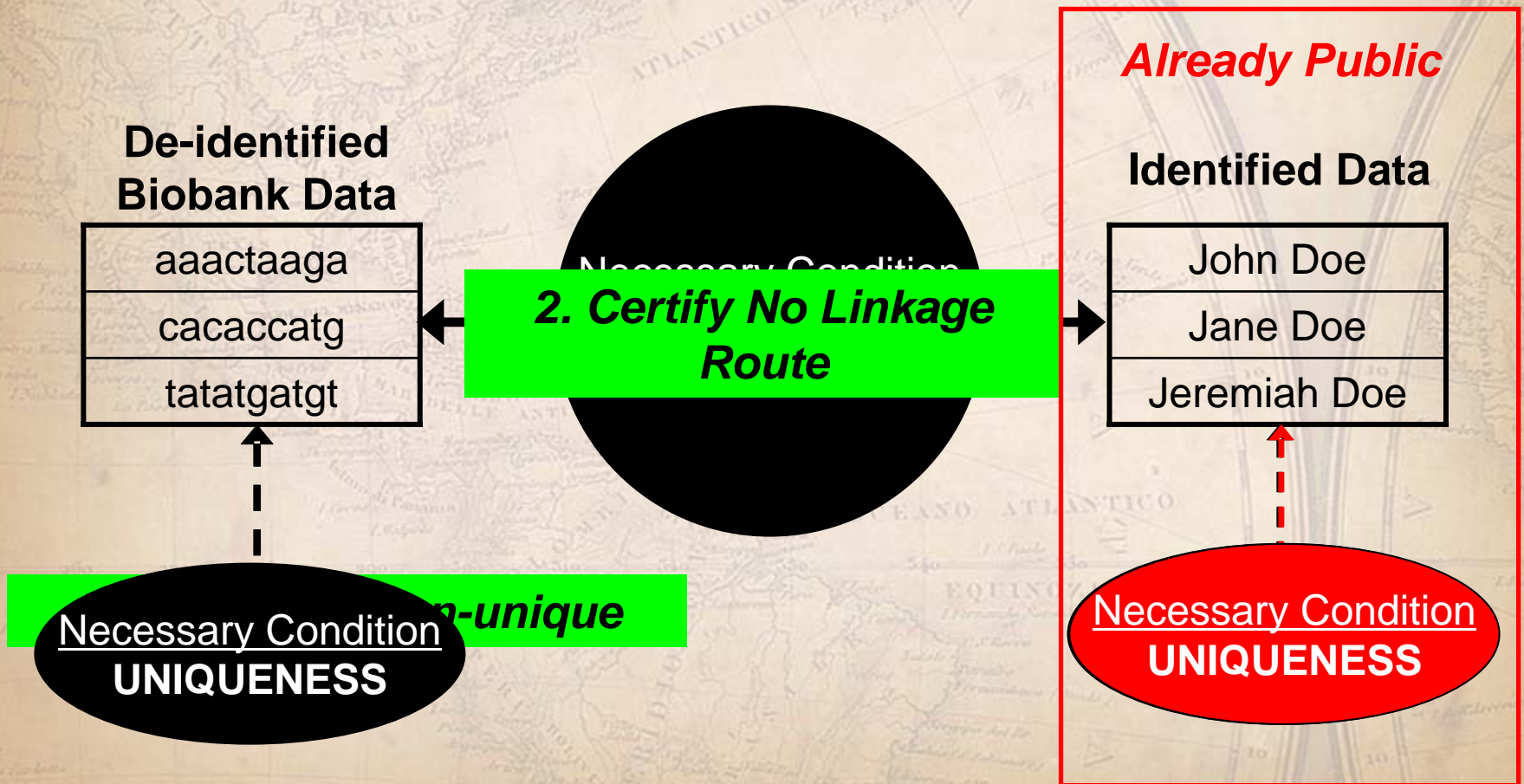
- internal correlations (see graph)

- Conclusions

- Too much perturbation needed to prevent linkage
- Keep records under lock and key

***Privacy
(Perturbation)***

Formal Re-identification Model

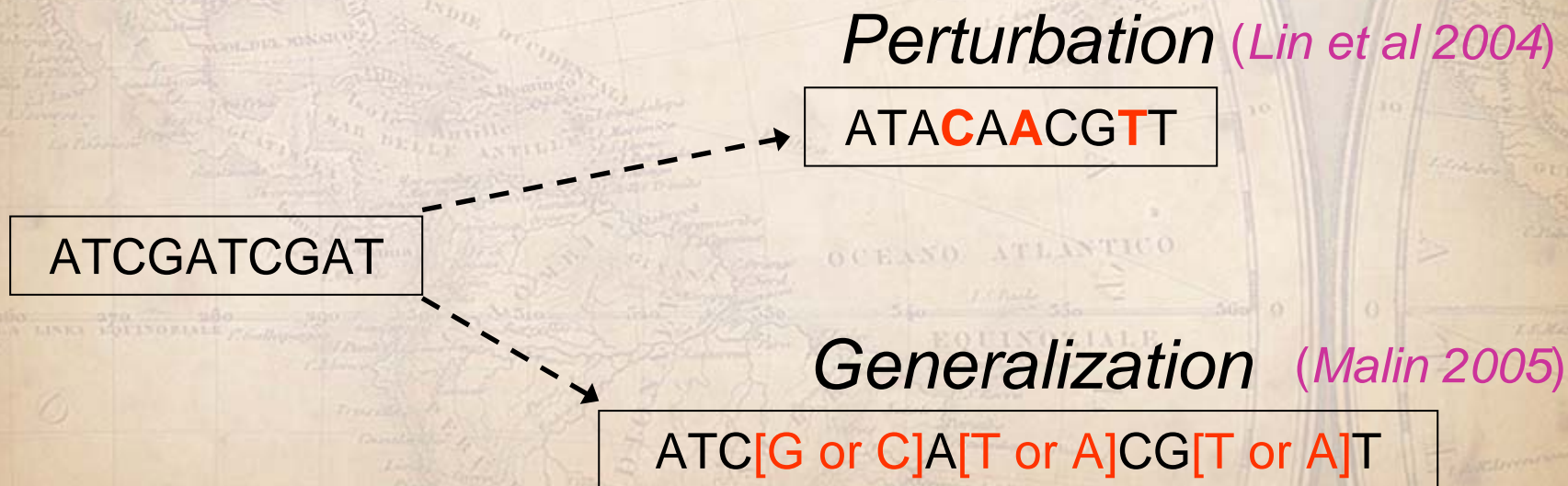


Formal Protection

- ***k*-Map** (Sweeney, 2002)
 - Each shared record refers to at least k entities in the population
- ***k*-Anonymity** (Sweeney, 2002)
 - Each shared record is equivalent to at least $k-1$ other records
- ***k*-Unlinkability** (Malin 2006)
 - Each shared record links to at least k identities via its trail
 - Satisfies *k*-Map protection model

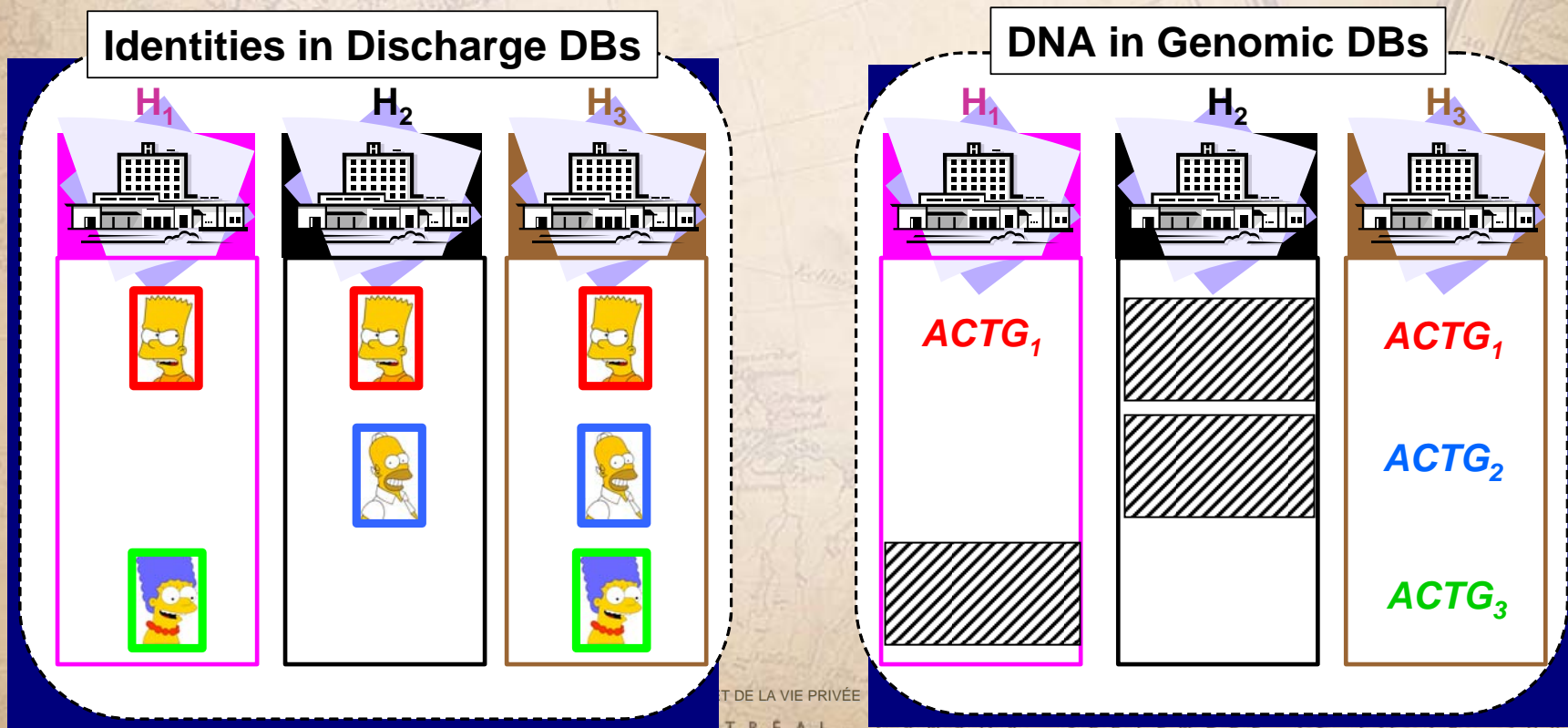
Beyond Ad hoc Protections

- Perturbation does not guarantee privacy
- Alternative: Generalization of data

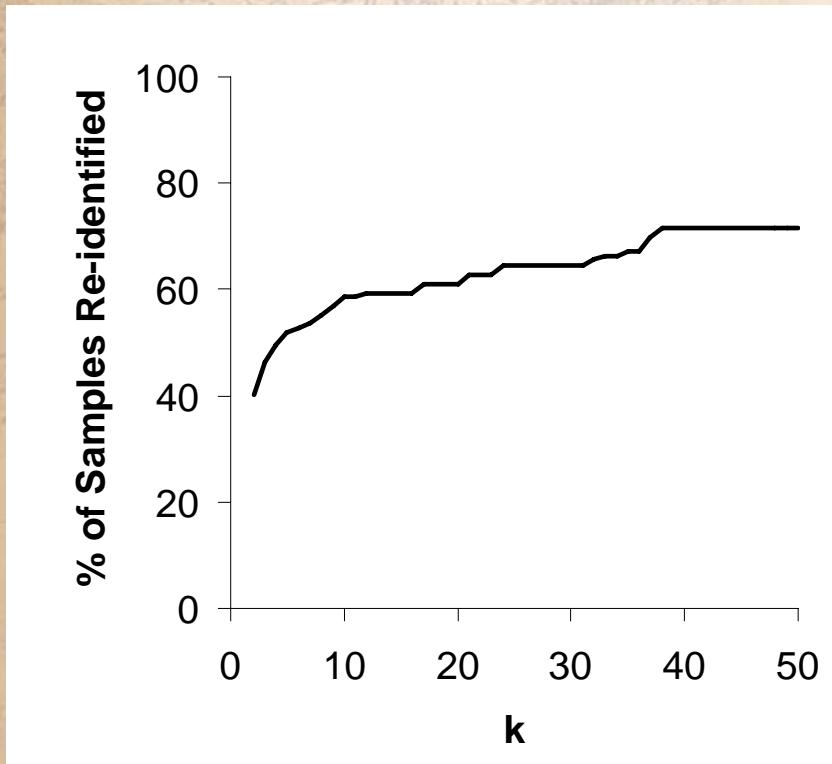


Learning Who You Are From Where You Have Been (“Trails”)

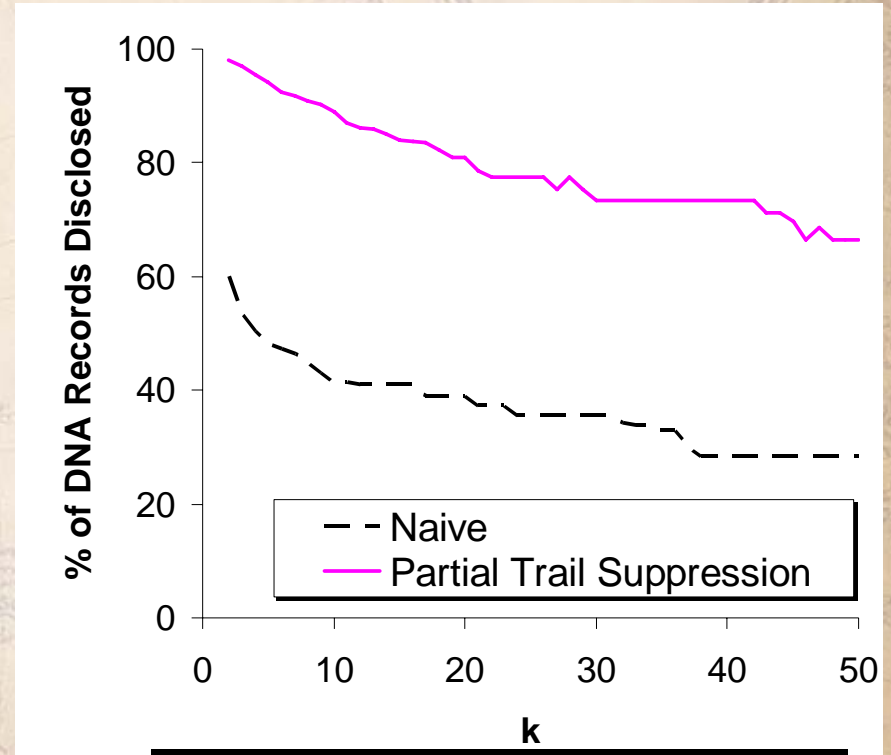
(Malin & Sweeney, 2001; 2004, Malin & Airoidi 2006)



Preventing Trails: Cystic Fibrosis Population (1149 samples)

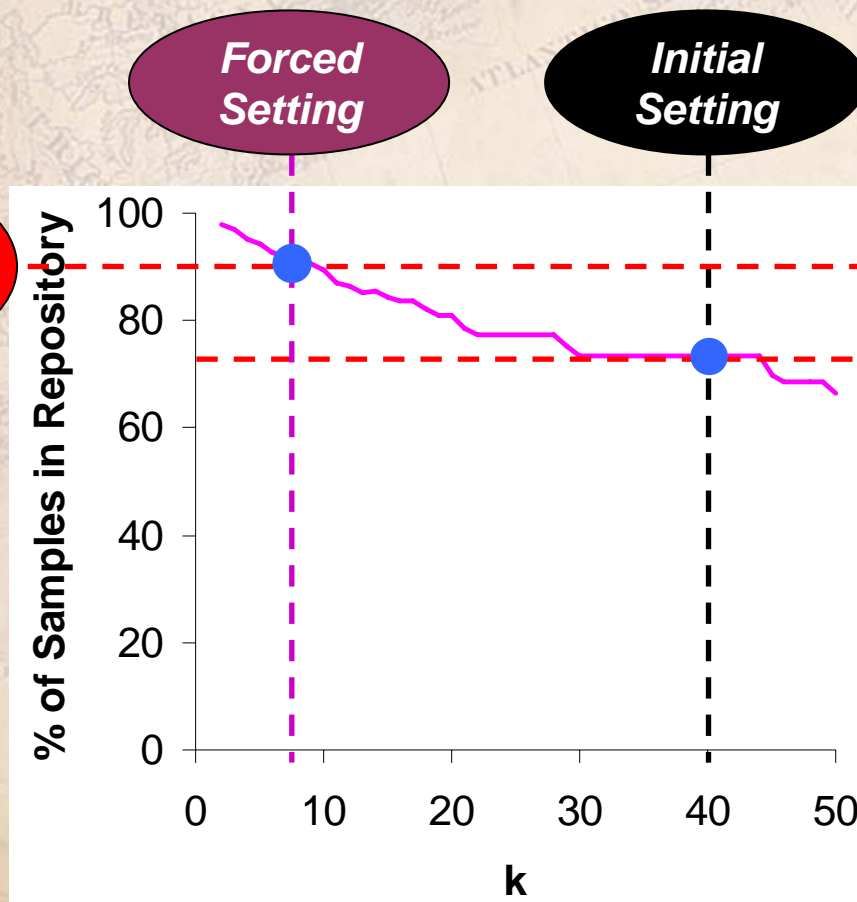


BEFORE STRANON
100% Samples In Repository



AFTER STRANON
0% Samples k-Re-identified

Benefit: Quantified Risk



- Change in re-identification risk
- Shift burden of increased risk to requesting analyst
- Ties together legal and computational models