

Comment apparier données sanitaires et *linguistiques* lorsqu'on a accès à aucunes d'entre elles

Arnaud Casteigts

Université d'Ottawa

4è Colloque international des Programmes Locaux
et Régionaux de Santé

—
27 juin 2011



 Institut de recherche
de l'Hôpital **Montfort**



uOttawa

L'Université canadienne
Canada's university

Principales données

- Données sur la population
 - Information démographique (assurance santé)
 - Statistiques vitales (registraire général)

Principales données

- Données sur la population
 - Information démographique (assurance santé)
 - Statistiques vitales (registraire général)
- Données sur l'utilisation des services
 - Hôpitaux, urgences, soins primaires, longue durée...

Principales données

- Données sur la population
 - Information démographique (assurance santé)
 - Statistiques vitales (registraire général)
- Données sur l'utilisation des services
 - Hôpitaux, urgences, soins primaires, longue durée...
- Données sur les professionnels
 - (via actes facturés)

Principales données

Institute for Clinical Evaluative Sciences

- Données sur la population
 - Information démographique (assurance santé)
 - Statistiques vitales (registraire général)
- Données sur l'utilisation des services
 - Hôpitaux, urgences, soins primaires, longue durée...
- Données sur les professionnels
 - (via actes facturés)

ICES



Principales données

Institute for Clinical Evaluative Sciences

- Données sur la population
 - Information démographique (assurance santé)
 - Statistiques vitales (registraire général)
- Données sur l'utilisation des services
 - Hôpitaux, urgences, soins primaires, longue durée...
- Données sur les professionnels
 - (via actes facturés)

ICES

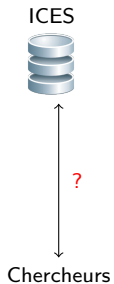


Le problème

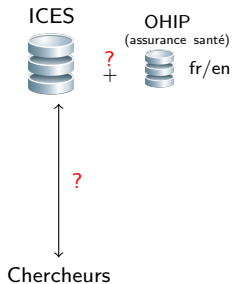
→ Pas de variable linguistique à ICES.

Étude des communautés francophones ?

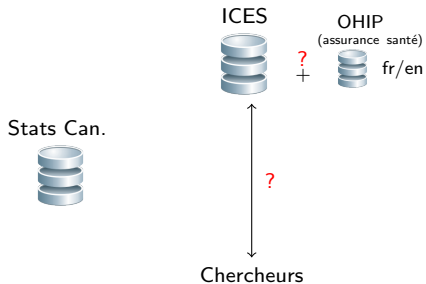
Solution 1 - interaction tripartite



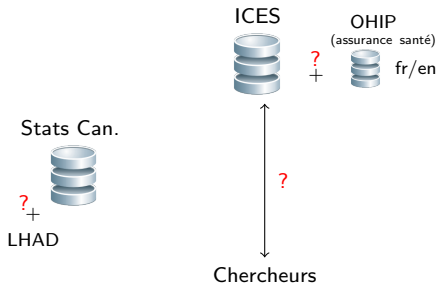
Solution 1 - interaction tripartite



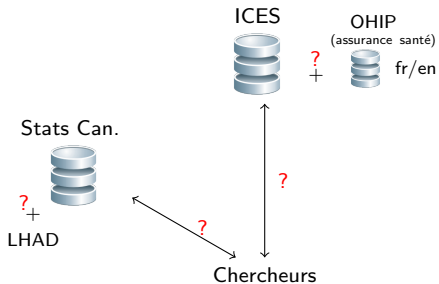
Solution 1 - interaction tripartite



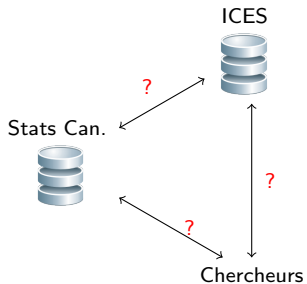
Solution 1 - interaction tripartite



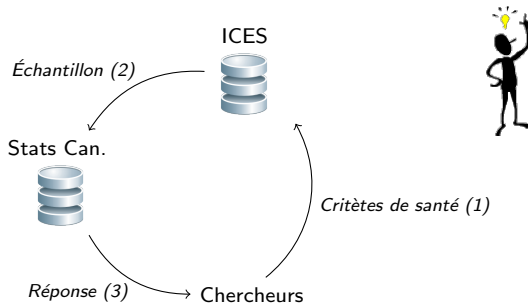
Solution 1 - interaction tripartite



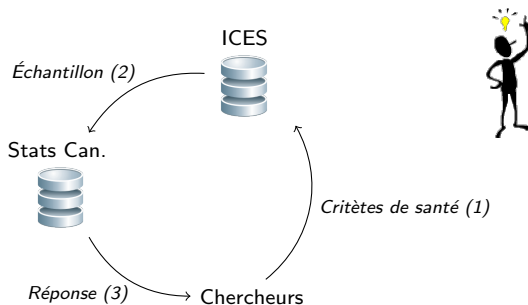
Solution 1 - interaction tripartite



Solution 1 - interaction tripartite

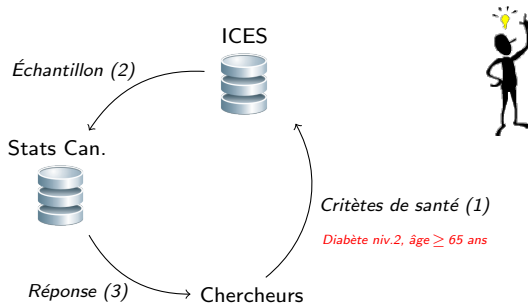


Solution 1 - interaction tripartite



*Exemple : taux de diabète chez les séniors de plus de 65 ans ?
(francophones vs. anglophones)*

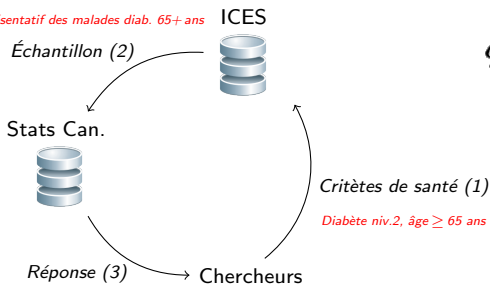
Solution 1 - interaction tripartite



*Exemple : taux de diabète chez les séniors de plus de 65 ans ?
(francophones vs. anglophones)*

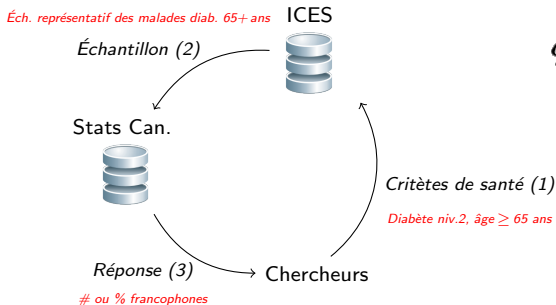
Solution 1 - interaction tripartite

Éch. représentatif des malades diab. 65+ ans



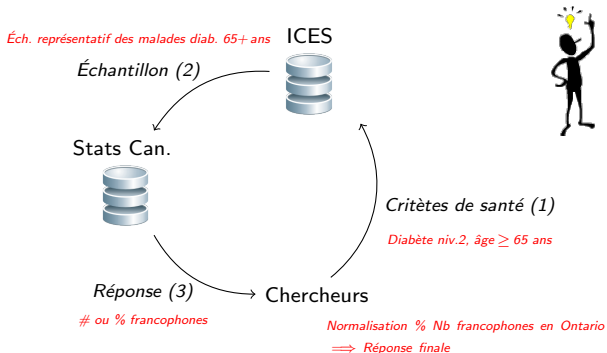
*Exemple : taux de diabète chez les séniors de plus de 65 ans ?
(francophones vs. anglophones)*

Solution 1 - interaction tripartite



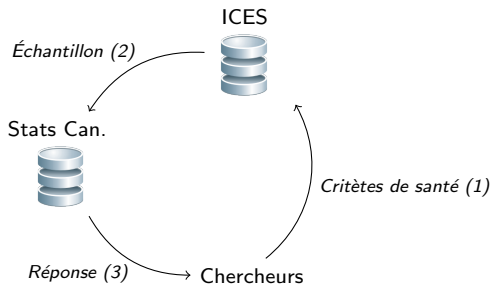
*Exemple : taux de diabète chez les séniors de plus de 65 ans ?
(francophones vs. anglophones)*

Solution 1 - interaction tripartite



*Exemple : taux de diabète chez les séniors de plus de 65 ans ?
(francophones vs. anglophones)*

Solution 1 - interaction tripartite

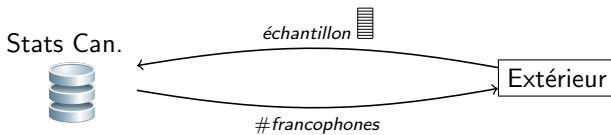


Oui mais..

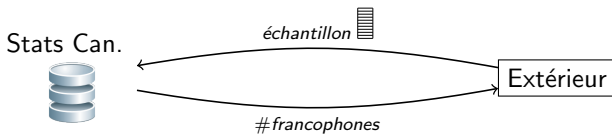
Statistique Canada ne distingue pas les entités externes !
(ICES et les chercheurs sont considérés comme une seule entité)



Le point de vue de Statistique Canada



Le point de vue de Statistique Canada

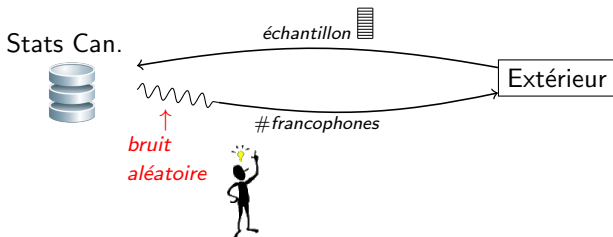


Conséquences..

Attaque ou collecte d'informations résiduelles possibles, par exemple :



Solution 2 - perturbation de la réponse



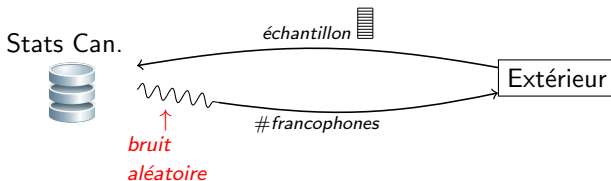
Conséquences..

extrêmement limitées

Attaque ou collecte d'informations résiduelles ~~possibles~~, par exemple :

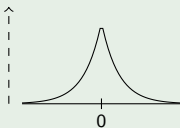
éch₁ =  éch₂ =  x

Le point de vue de Statistique Canada



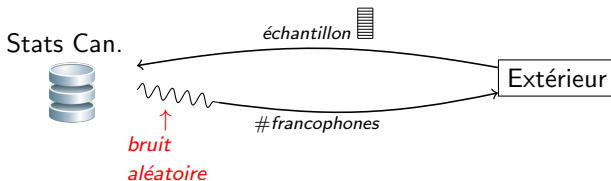
Ajout de bruit aléatoire

probabilité



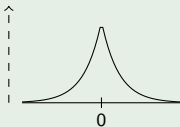
(a) Nombre aléatoire

Le point de vue de Statistique Canada



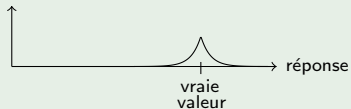
Ajout de bruit aléatoire

probabilité



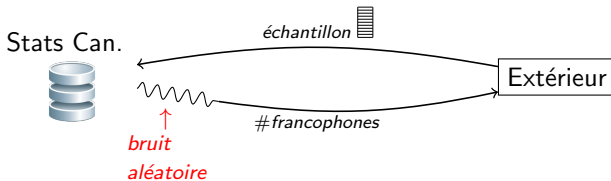
(a) Nombre aléatoire

probabilité

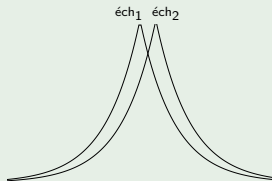


(b) Probabilité de réponse

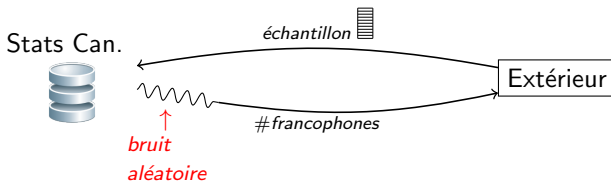
Le point de vue de Statistique Canada



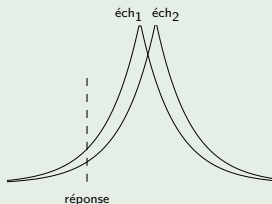
Ajout de bruit aléatoire



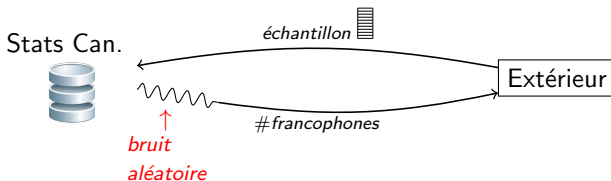
Le point de vue de Statistique Canada



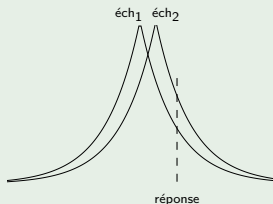
Ajout de bruit aléatoire



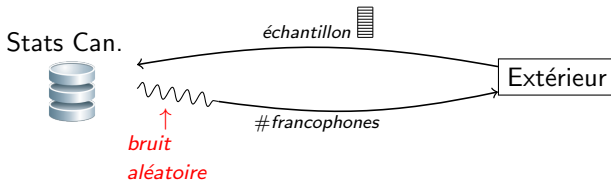
Le point de vue de Statistique Canada



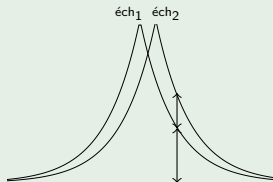
Ajout de bruit aléatoire



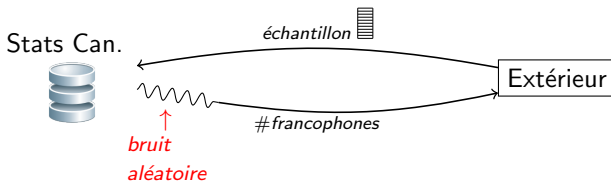
Le point de vue de Statistique Canada



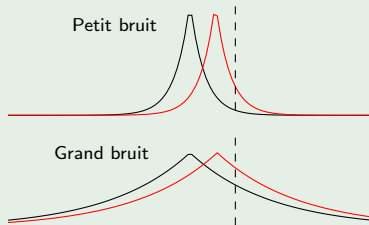
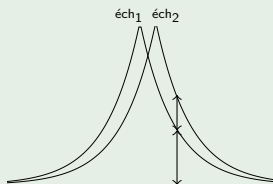
Ajout de bruit aléatoire



Le point de vue de Statistique Canada

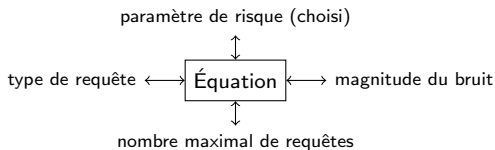


Ajout de bruit aléatoire



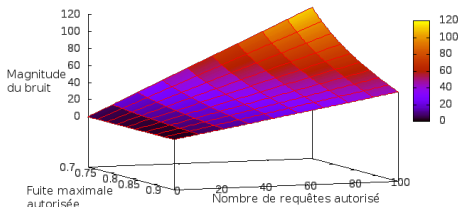
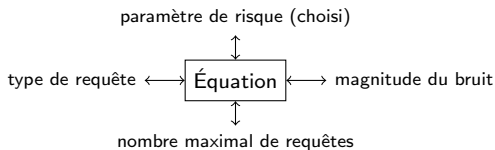
Un résultat théorique récent (2007)

Équation liant le niveau de bruit au risque statistique (Nissim *et. al*, STOC 2007)



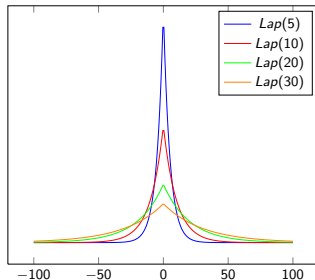
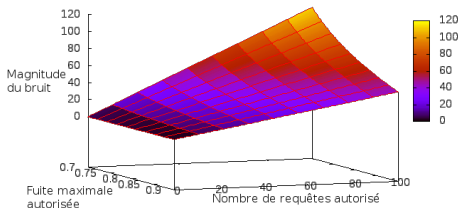
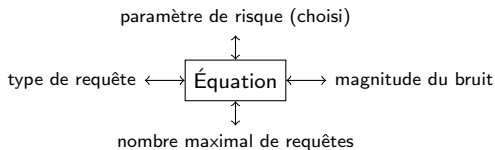
Un résultat théorique récent (2007)

Équation liant le niveau de bruit au risque statistique (Nissim *et. al*, STOC 2007)



Un résultat théorique récent (2007)

Équation liant le niveau de bruit au risque statistique (Nissim *et. al*, STOC 2007)



Questions ouvertes

- Quel niveau de bruit est acceptable pour les chercheurs ?

Questions ouvertes

- Quel niveau de bruit est acceptable pour les chercheurs ?
- Quel serait le risque induit par un usage *normal* (vs. pire scénario) ?

Questions ouvertes

- Quel niveau de bruit est acceptable pour les chercheurs ?
- Quel serait le risque induit par un usage *normal* (vs. pire scénario) ?
- Autres types de requêtes ?

Questions ouvertes

- Quel niveau de bruit est acceptable pour les chercheurs ?
- Quel serait le risque induit par un usage *normal* (vs. pire scénario) ?
- Autres types de requêtes ?

Perspectives

- Pour Statistique Canada.

Questions ouvertes

- Quel niveau de bruit est acceptable pour les chercheurs ?
- Quel serait le risque induit par un usage *normal* (vs. pire scénario) ?
- Autres types de requêtes ?

Perspectives

- Pour Statistique Canada.
- Pour ICES.

Questions ouvertes

- Quel niveau de bruit est acceptable pour les chercheurs ?
- Quel serait le risque induit par un usage *normal* (vs. pire scénario) ?
- Autres types de requêtes ?

Perspectives

- Pour Statistique Canada.
- Pour ICES.
- Pour les chercheurs en santé des populations

Le travail présenté :



Enabling Dynamic Linkage of Linguistic Census Data at Statistics Canada.

A. Casteigts, M.-H. Chomienne, L. Bouchard, G.-V. Jourdan

Rapport de recherche RRASFO, 2011

Le résultat fondamental utilisé :



Smooth sensitivity and sampling in private data analysis

K. Nissim, S. Raskhodnikova, A. Smith

39th ACM Symposium on Theory of Computing, 2007

