PROGRAMME

- Accueil
- Introduction
- Session 1 : une nouvelle méthodologie pour un nouveau census
- Session 2 : la protection des données grâce au swapping

Pause-café

- Session 3 : démographie, ménages et noyaux familiaux
- Session 4 : marché du travail - Des registres aux variables

Pause de midi

- Session 5 : enseignement - Intégration des données des communautés
- Session 6 : création d'une base de données « logements »

Pause-café

- Session 7 : comment s'effectue la diffusion des données du Census ?
- Conclusion

economie
SPF Economie, P.M.E., Classes moyennes et Energie

Journée d'étude Census 2011

Session 2 : la protection des données grâce au *swapping*

20 janvier 2015

.be

Session 2 : plan

1. Pourquoi le *record swapping* ?
2. Fonctionnement du *record swapping*
3. Conséquences du *record swapping*

economie
SPF Economie, P.M.E., Classes moyennes et Energie

Protection des données : ancienne méthode

- Recours jusqu'à présent à une méthode post-tabulaire en Belgique
- Confidentialité primaire : protection des cellules avec de faibles fréquences en masquant les valeurs
- Confidentialité secondaire : éviter que des valeurs masquées puissent à nouveau être calculées => en masquant des valeurs supplémentaires. Logiciel : tau-argus

economie
SPF Economie, P.M.E., Classes moyennes et Energie

Exemple fictif de confidentialité secondaire

	Pays de naissance				
Commune	France	Allemagne	Nouvelle-Zélande	…	Total
AAAAA	100	150	0	4750	5000
BBBBB	200	300	8	~~9492~~	10000
CCCCC	300	450	1	14249	15000
DDDDD	400	600	9	~~18991~~	20000
EEEEE	500	750	7	23743	25000
Total	1500	2250	25	71225	75000

1. Pourquoi le *record swapping* ?

- Pourquoi la méthode post-tabulaire choisie pose-t-elle problème ?
  - Il est très difficile de calculer la confidentialité secondaire avec un nombre élevé de dimensions. Performance très limitée. Possibilités restreintes avec tau-argus.
  - Problème pratique : protéger tous les tableaux simultanément
  - Trop de cellules masquées
- Avantages d'une méthode pré-tabulaire :
  - Une fois les microdonnées protégées => très simple de créer des cubes
  - Cohérence des données entre les cubes
- Autre avantage du *record swapping*
  - La distribution des fréquences des variables reste identique.

Session 2 : plan

1. Pourquoi le *record swapping* ?
2. Fonctionnement du *record swapping*
3. Conséquences du *record swapping*

2. Fonctionnement du *record swapping*

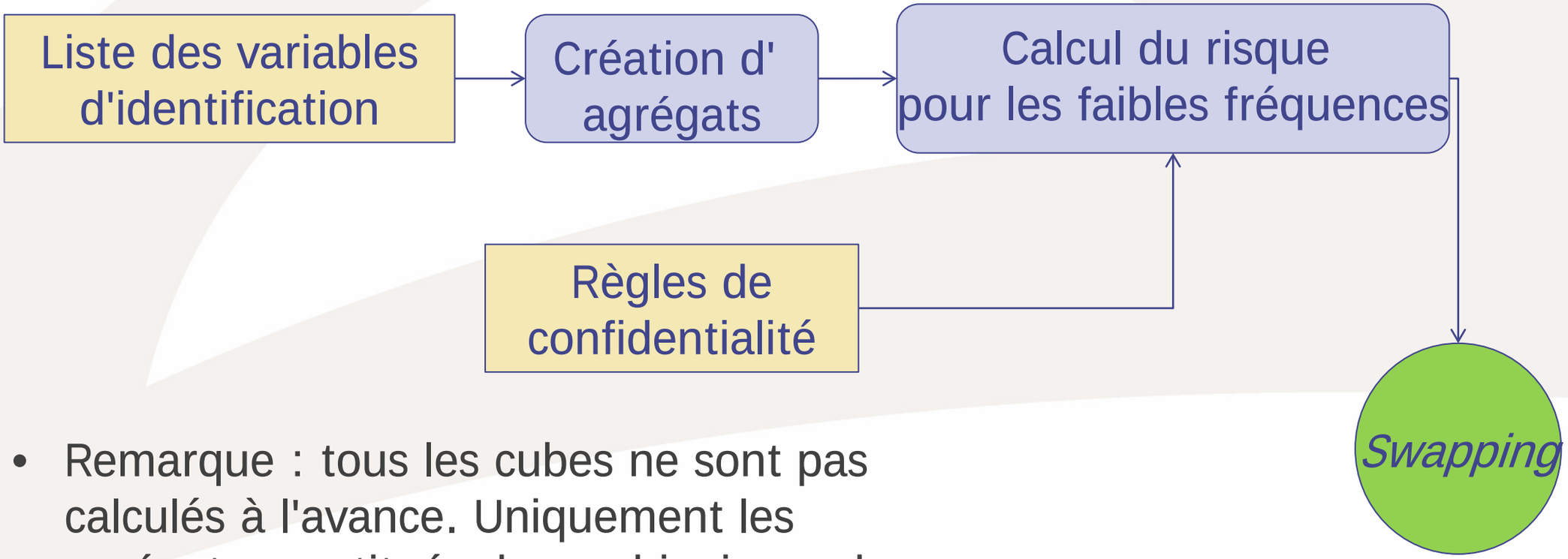
- Caractéristiques et principe du *record swapping*
  - Méthode pré-tabulaire
  - Pour certains enregistrements, les valeurs de certaines variables sont échangées entre deux enregistrements.
  - Plus aucune cellule masquée dans les cubes

- Méthode
  - Étape 1 : identification des enregistrements à échanger
  - Étape 2 : recherche d'un enregistrement voisin pour effectuer l'échange

- Protection supplémentaire : pas de communication des règles de confidentialité

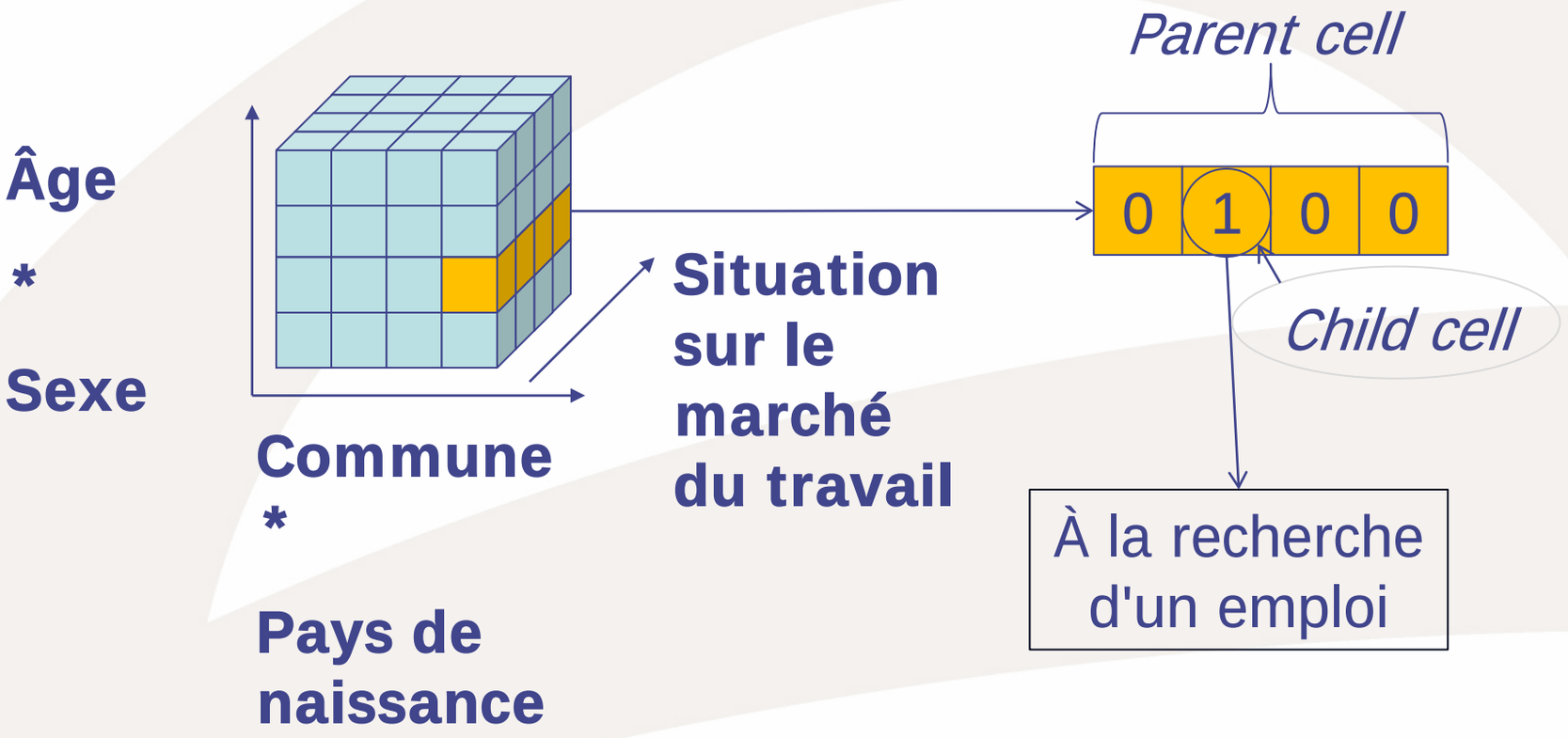
2. Fonctionnement du *record swapping*

Étape 1 : identification des enregistrements à échanger



- Remarque : tous les cubes ne sont pas calculés à l'avance. Uniquement les agrégats constitués de combinaisons de variables d'identification.

Exemple de contrôle de la vie privée



economie
SPF Economie, P.M.E., Classes moyennes et Energie

Exemple de contrôle de la vie privée : processus d'identification

- Variables d'identification :
  - commune, pays de naissance, sexe, âge

- Le cube partiel de 4 dimensions a une *parent cell* ayant une fréquence de 1.

- La *child cell* (cube de 5 dimensions) contient une cellule ayant une fréquence de 1 pour « situation sur le marché du travail » = « à la recherche d'un emploi ».

economie
SPF Economie, P.M.E., Classes moyennes et Energie

Exemple de contrôle de la vie privée : processus d'identification

- Règle de confidentialité : si une *parent cell* a une faible fréquence ET si la *parent cell* contient des données sur le pays de naissance ET si la *child cell* est « à la recherche d'un emploi » => données sensibles, donc confidentielles.

=> *Drill down* d'une cellule identifiable (dimension 4) à une cellule (dimension 5) contenant des données confidentielles => fournit des données supplémentaires sur la personne en question !

economie
SPF Economie, P.M.E., Classes moyennes et Energie

Exemples de contrôle de la vie privée : calcul du risque

Commune	Pays de naissance	Sexe	Âge	Situation sur le marché du travail	N
Merelbeke	Nouvelle-Zélande	Femme	24	À la recherche d'un emploi	1
Merelbeke	Nouvelle-Zélande	Femme	24	Occupée	0
Merelbeke	Nouvelle-Zélande	Femme	24	Étudiante	0
Merelbeke	Nouvelle-Zélande	Femme	24	…	0

Faible fréquence = 1. Le risque d'une cellule sensible est 1/1 = 100 %.

economie
SPF Economie, P.M.E., Classes moyennes et Energie

Exemples de contrôle de la vie privée : calcul du risque

Commune	Pays de naissance	Sexe	Âge	Situation sur le marché du travail	N
Merelbeke	Nouvelle-Zélande	Femme	24	À la recherche d'un emploi	3
Merelbeke	Nouvelle-Zélande	Femme	24	Occupée	0
Merelbeke	Nouvelle-Zélande	Femme	24	Étudiante	0
Merelbeke	Nouvelle-Zélande	Femme	24	…	0

Faible fréquence = 3. Le risque d'une cellule sensible est 3/3 = 100 %. Les 3 personnes sont toutes à la recherche d'un emploi : il est donc certain que la personne est à la recherche d'un emploi.

economie
SPF Economie, P.M.E., Classes moyennes et Energie

Exemples de contrôle de la vie privée : calcul du risque

Commune	Pays de naissance	Sexe	Âge	Situation sur le marché du travail	N
Merelbeke	Nouvelle-Zélande	Femme	24	À la recherche d'un emploi	1
Merelbeke	Nouvelle-Zélande	Femme	24	Occupée	2
Merelbeke	Nouvelle-Zélande	Femme	24	Étudiante	1
Merelbeke	Nouvelle-Zélande	Femme	24	…	0

Faible fréquence = 4 (*parent cell*). Le risque d'une cellule sensible est 1/4 = 25 %. Seule la variable « à la recherche d'un emploi » est considérée ici comme sensible. 25 %, soit déjà moins de raisons de recourir au *swapping*. Si aucune faible fréquence => pas de *swapping*

economie
SPF Economie, P.M.E., Classes moyennes et Energie

Étape 2 : réalisation du *swapping*

- « Distance » entre les deux enregistrements : mesure dans laquelle les deux enregistrements diffèrent. Des poids sont attribués à certaines variables (pas à toutes). La distance est calculée en fonction des poids des variables ayant des valeurs différentes.
- Recherche d'un enregistrement (pas encore échangé) au sein du même arrondissement / de la même province => le plus proche possible de l'enregistrement initial, mais avec une valeur différente pour la variable sensible. Il faut respecter certains critères de base. Échange de la commune.

economie
SPF Economie, P.M.E., Classes moyennes et Energie

Étape 2 : réalisation du *swapping*

- Si aucun enregistrement non échangé n'est identifié selon les critères de base => échange au sein de la même région.

- Si impossible => au sein de la Belgique

- Si impossible => échange de l'âge

economie
SPF Economie, P.M.E., Classes moyennes et Energie

Étape 2 : réalisation du *swapping* – exemple fictif

Microdonnées – avant le *swapping* — donnée sensible

Commune	Pays de naissance	Sexe	Âge	Situation sur le marché du travail	État civil
Merelbeke	Nouvelle-Zélande	Femme	24	À la recherche d'un emploi	Célibataire
Melle	Nouvelle-Zélande	Femme	24	Occupée	Mariée
Melle	Nouvelle-Zélande	Femme	24	Occupée	Célibataire
Hasselt	Nouvelle-Zélande	Femme	24	Occupée	Célibataire

Échange au sein de la même province

economie
SPF Economie, P.M.E., Classes moyennes et Energie

Étape 2 : réalisation du *swapping* – exemple fictif

Microdonnées – après le *swapping*

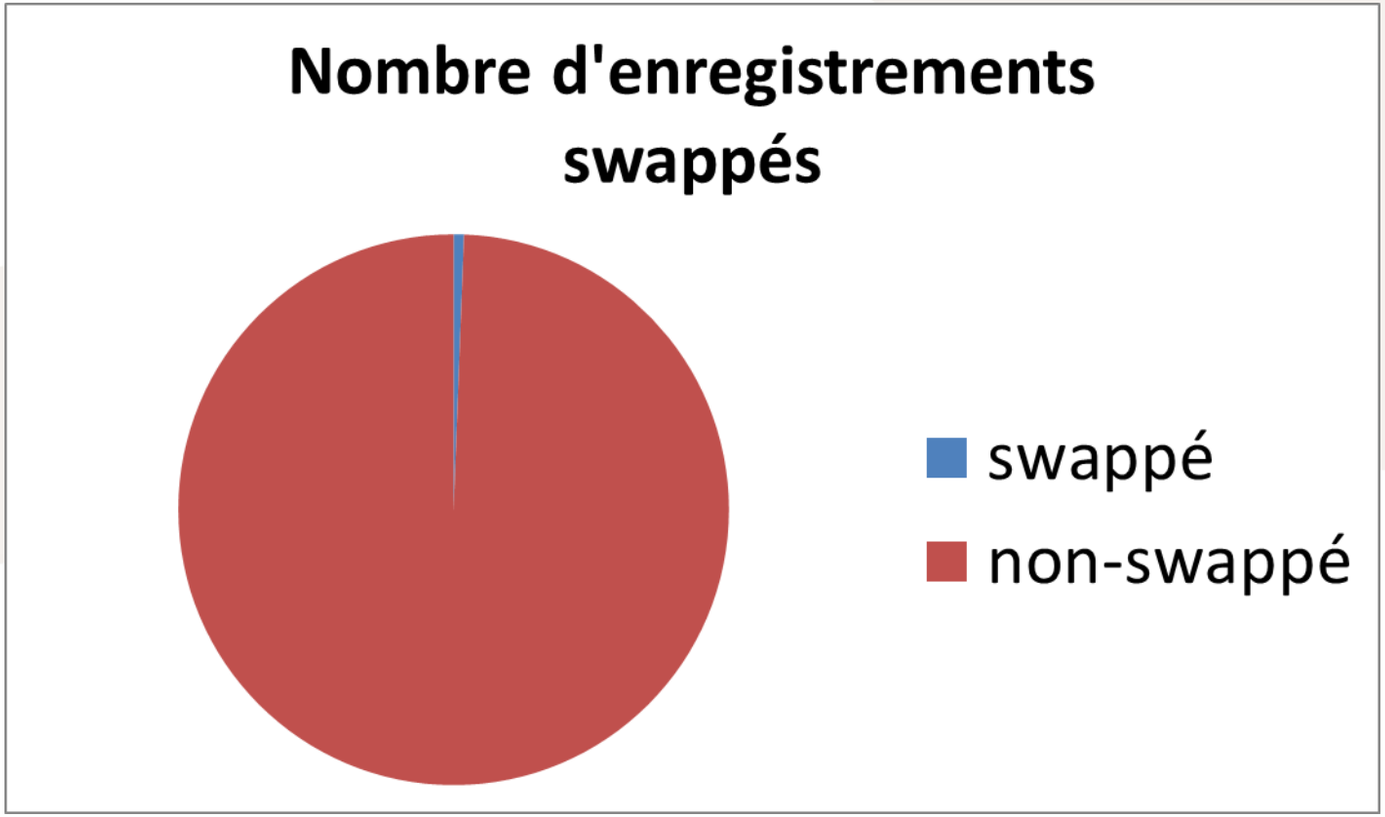
Commune	Pays de naissance	Sexe	Âge	Situation sur le marché du travail	État civil
Melle	Nouvelle-Zélande	Femme	24	À la recherche d'un emploi	Célibataire
Melle	Nouvelle-Zélande	Femme	24	Occupée	Mariée
Merelbeke	Nouvelle-Zélande	Femme	24	Occupée	Célibataire
Hasselt	Nouvelle-Zélande	Femme	24	Occupée	Célibataire

Session 2 : plan

1. Pourquoi le *record swapping* ?
2. Fonctionnement du *record swapping*
3. Conséquences du *record swapping*

3. Conséquences du *swapping*

- Faibles fréquences => incertitudes : possibilité d'un *swapping*
- Protection des 60 cubes (250 millions de cellules) et des autres tableaux grâce à l'échange d'un nombre restreint d'enregistrements.



economie
SPF Economie, P.M.E., Classes moyennes et Energie

Questions

http://economie.fgov.be