

- Onthaal
- Inleiding
- Sessie 1: Een nieuwe methodologie voor een nieuwe census
- **Sessie 2: De bescherming van de gegevens door swapping**

Koffiepauze

- Sessie 3: Demografie, huishoudens en familiekeren
- Sessie 4: Arbeidsmarkt – Van registers tot variabelen

Middagpauze

- Sessie 5: Onderwijsdata – Integratie van de gegevens van de gemeenschappen
- Sessie 6: Aanmaak van een woningendatabank

Koffiepauze

- Sessie 7: Hoe worden de gegevens van de census verspreid ?
- **Slotwoord**

Studiedag Census 2011

Sessie 2: De bescherming van de gegevens door swapping

20 januari 2015



Sessie 2: Agenda

1. Waarom record swapping ?
2. Werking van record swapping
3. Gevolgen van record swapping

Beveiliging van gegevens: oude methode

- Tot nu toe werd in België een post-tabulaire methode gebruikt.
- Primaire confidentialiteit: bescherming van cellen met lage frequenties door waarden te verbergen
- Secundaire confidentialiteit: Voorkomen dat verborgen waarden opnieuw berekend kunnen worden => extra waarden verbergen. Software: tau-argus

1. Waarom record swapping ?

Fictief voorbeeld van secundaire confidentialiteit

Gemeente	Land van geboorte				Totaal
	Frankrijk	Duitsland	Nieuw-Zeeland	...	
AAAAA	100	150	0	4750	5000
BBBBB	200	300	8	9492	10000
CCCCC	300	450	1	14249	15000
DDDDD	400	600	9	18991	20000
EEEEEE	500	750	7	23743	25000
Totaal	1500	2250	25	71225	75000

1. Waarom record swapping ?

- Wat is het probleem met de gekozen post-tabulaire methode ?
 - Secundaire confidentialiteit is zeer moeilijk te berekenen bij een hoog aantal dimensies. Zeer lage performantie. Mogelijkheden tau-argus beperkt.
 - Praktisch probleem: alle tabellen in één keer beschermen
 - Te veel cellen worden verborgen.
- Voordelen van een pre-tabulaire methode:
 - Eenmaal micro-gegevens beschermd => zeer eenvoudig kubussen aan te maken.
 - Consistentie van gegevens tussen de kubussen
- Extra voordeel van record swapping
 - Frequentieverdelingen van de variabelen blijven ongewijzigd

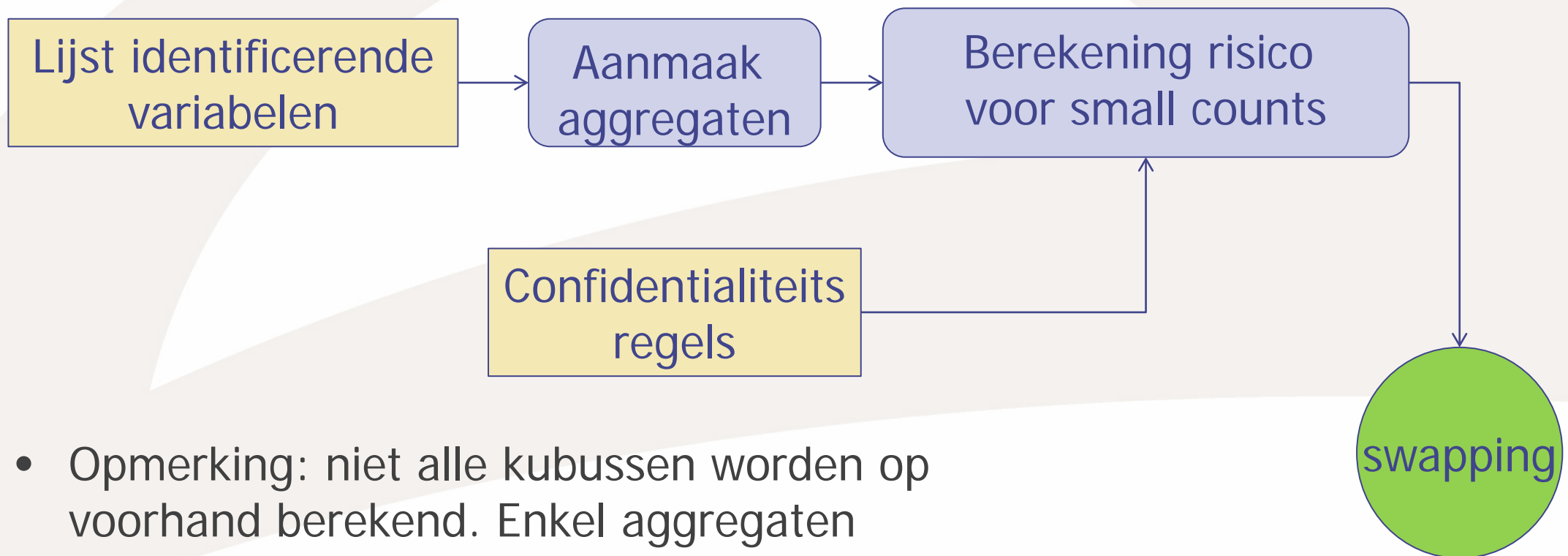


Sessie 2: Agenda

1. Waarom record swapping ?
2. Werking van record swapping
3. Gevolgen van record swapping

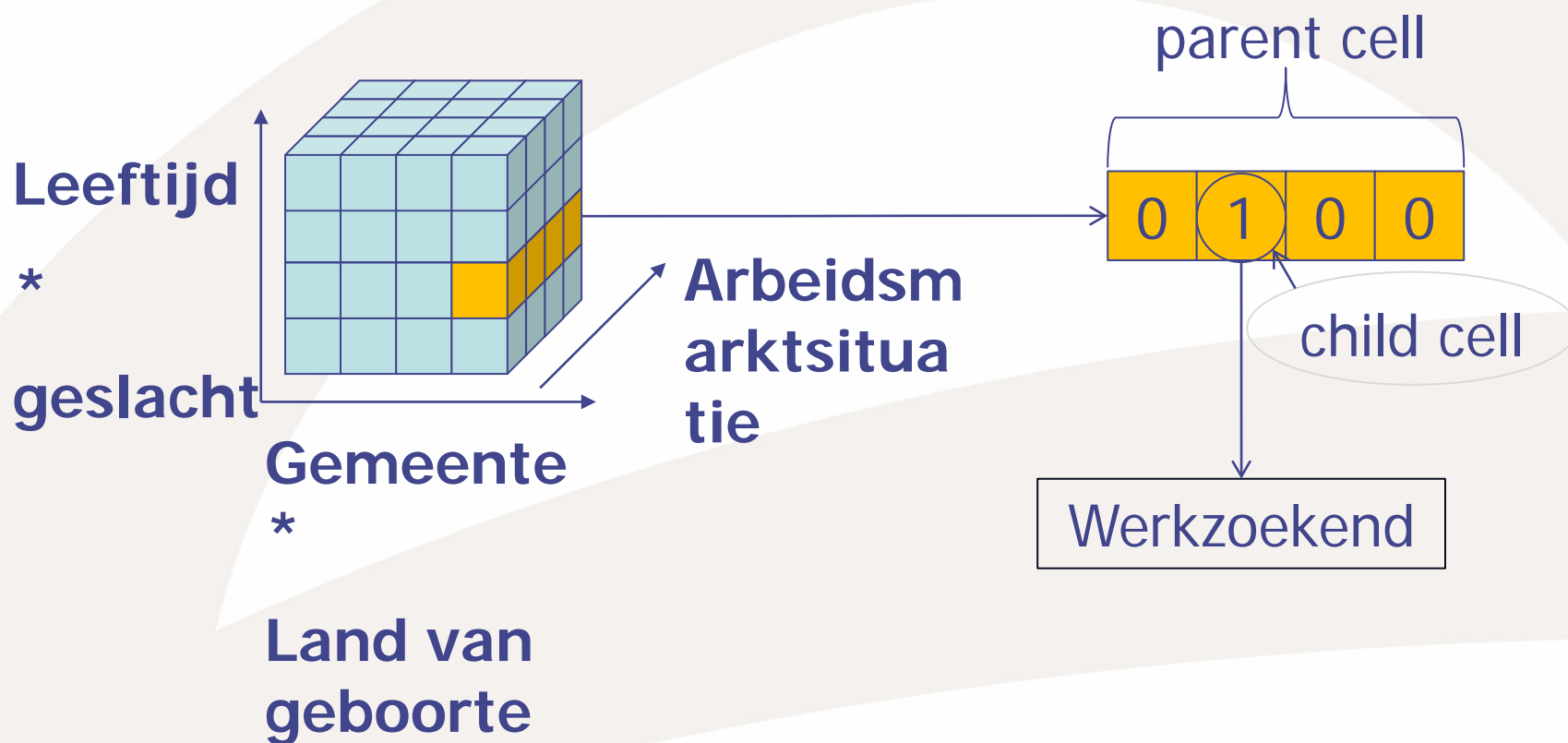
- Kenmerken en principe van record swapping
 - Pre-tabulaire methode
 - Voor een aantal records worden waarden van sommige variabelen omgewisseld tussen twee records.
 - Geen verborgen cellen meer in de kubussen
- Methode
 - Stap 1: Identificeren van “te swappen” records
 - Stap 2: Op zoek gaan naar een tegenhanger om te swappen en de omwisseling uitvoeren
- Extra bescherming: geen communicatie van lijst met confidentialiteitsregels

Stap 1: Identificeren van “te swappen” records



- Opmerking: niet alle kubussen worden op voorhand berekend. Enkel aggregaten bestaande uit combinaties van identificerende variabelen.

Voorbeeld van privacycontrole



Voorbeeld van privacycontrole: uitwerking identificatie

- Identificerende variabelen:
 - Gemeente, land van geboorte, geslacht, leeftijd
- Deelkubus van 4 dimensies heeft een “parent cell” met frequentie 1 (small count)
- “Child cell” (kubus van 5 dimensies) bevat een cell met frequentie 1 voor “arbeidsmarktsituatie = werkzoekend”

Voorbeeld van privacycontrole: uitwerking identificatie

- Confidentialiteitsregel: indien small count voor een “parent cell” EN “parent cell” bevat info over land van geboorte EN “child cell” is een werkzoekende => gevoelige info, dus confidantieel

=> “Drill down” van identificeerbare cel (dimensie 4) naar een cel (dimensie 5) met confidantiele informatie => leert extra info over de persoon in kwestie !

Voorbeelden van privacycontrole: berekening risico

Gemeente	Land van geboorte	Geslacht	Leeftijd	Arbeidsmarktsituatie	N
Merelbeke	Nieuw-Zeeland	vrouw	24	werkzoekend	1
Merelbeke	Nieuw-Zeeland	vrouw	24	werkend	0
Merelbeke	Nieuw-Zeeland	vrouw	24	student	0
Merelbeke	Nieuw-Zeeland	vrouw	24	...	0

Small count = 1. Risico op een gevoelige cel is $1/1 = 100\%$

Voorbeelden van privacycontrole: berekening risico

Gemeente	Land van geboorte	Geslacht	Leeftijd	Arbeidsmarktsituatie	N
Merelbeke	Nieuw-Zeeland	vrouw	24	werkzoekend	3
Merelbeke	Nieuw-Zeeland	vrouw	24	werkend	0
Merelbeke	Nieuw-Zeeland	vrouw	24	student	0
Merelbeke	Nieuw-Zeeland	vrouw	24	...	0

Small count = 3. Risico op een gevoelige cel is $3/3 = 100\%$

Alle 3 de personen zijn werkzoekend: dus je bent "zeker" dat de persoon werkzoekend is.

Voorbeelden van privacycontrole: berekening risico

Gemeente	Land van geboorte	Geslacht	Leeftijd	Arbeidsmarktsituatie	N
Merelbeke	Nieuw-Zeeland	vrouw	24	werkzoekend	1
Merelbeke	Nieuw-Zeeland	vrouw	24	werkend	2
Merelbeke	Nieuw-Zeeland	vrouw	24	student	1
Merelbeke	Nieuw-Zeeland	vrouw	24	...	0

Small count = 4 (parent cell). Risico op een gevoelige cel is $1/4 = 25\%$. Enkel werkzoekend wordt hier als gevoelig beschouwd. 25% is al minder reden om te swappen.

Indien geen small count => geen swapping

Stap 2: Swapping uitvoeren

- “Afstand” tussen twee records: mate waarin twee records van elkaar verschillen. Aan een aantal variabelen worden gewichten toegekend (niet alle variabelen). Afstand wordt berekend op basis van de gewichten van de variabelen met verschillende waarden.
- Zoek een record (nog niet geswapt) binnen zelfde arrondissement / provincie => zo weinig mogelijk verschillend van oorspronkelijke record, maar met een verschillende waarde voor de gevoelige variabele. Bepaalde basiscriteria moeten voldoen. Swap de gemeente.

Stap 2: Swapping uitvoeren

- Indien geen niet-geswapt record volgens basiscriteria gevonden => swap binnen hetzelfde gewest.
- Indien niet mogelijk => binnen België
- Indien niet mogelijk => swap de leeftijd

Stap 2: Swapping uitvoeren – fictief voorbeeld

Micro gegevens – voor swapping

gevoelig gegeven

Gemeente	Land van geboorte	Geslacht	Leeftijd	Arbeidsmarkt situatie	Burgerlijke Staat
Merelbeke	Nieuw-Zeeland	vrouw	24	werkzoekend	ongetrouwd
Melle	Nieuw-Zeeland	vrouw	24	werkend	getrouwd
Melle	Nieuw-Zeeland	vrouw	24	werkend	ongetrouwd
Hasselt	Nieuw-Zeeland	vrouw	24	werkend	ongetrouwd

Swapping binnen zelfde provincie

Stap 2: Swapping uitvoeren – fictief voorbeeld

Micro gegevens – na swapping

Gemeente	Land van geboorte	Geslacht	Leeftijd	Arbeidsmarkt situatie	Burgerlijke Staat
Melle	Nieuw-Zeeland	vrouw	24	werkzoekend	ongetrouwd
Melle	Nieuw-Zeeland	vrouw	24	werkend	getrouwd
Merelbeke	Nieuw-Zeeland	vrouw	24	werkend	ongetrouwd
Hasselt	Nieuw-Zeeland	vrouw	24	werkend	ongetrouwd

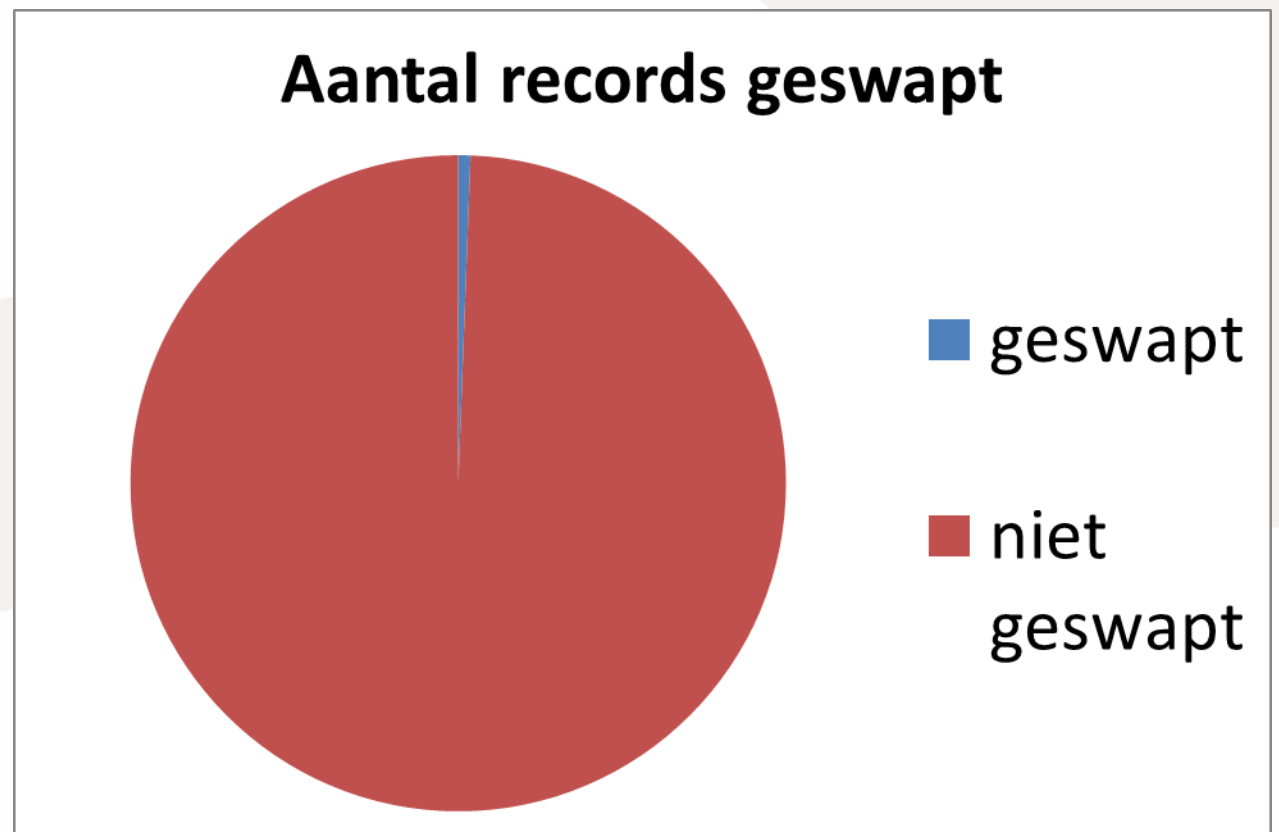


Sessie 2: Agenda

1. Waarom record swapping ?
2. Werking van record swapping
3. Gevolgen van record swapping

3. Gevolgen van de swapping

- Kleine frequentiewaarden => onzekerheid : misschien een swapping
- De 60 kubussen (250 miljoen cellen) en andere tabellen beschermd door een klein aantal records die geswapt zijn.



Vragen

