

Gepseudonimiseerde persoonsgegevens zijn gegevens die slechts met een code in verband kunnen worden gebracht met geïdentificeerde of identificeerbare natuurlijke personen (of nog: gegevens die niet meer aan specifieke personen kunnen worden gekoppeld zonder gebruik te maken van aanvullende gegevens die afzonderlijk worden bewaard). De organisatie die ze verwerkt voor onderzoeksdoeleinden kan ze zelf niet herleiden tot de betrokken natuurlijke personen.

De organisatie die de persoonsgegevens verwerkt voor onderzoeksdoeleinden beschikt voor de identificatie van de betrokkenen enkel over een inhoudsloze code. Het volstaat overigens niet om het identificatienummer, de naam, de voornaam en het adres van de betrokkene te verwijderen om van gepseudonimiseerde persoonsgegevens te kunnen gewagen. Er moet worden vermeden dat diens kenmerken al te precies beschikbaar zijn.

De KSZ zal derhalve bijzondere aandacht besteden aan het minimaliseren van de mogelijkheid om de betrokken personen te heridentificeren. Zodoende moeten de gegevens die worden verstrekt voldoende veralgemeend zijn (bijvoorbeeld door het gebruik van voldoende ruime waardenklassen) en mag de hoeveelheid te combineren selectiecriteria niet zo groot zijn dat het aantal personen dat eraan voldoet zeer klein is (en die personen bijgevolg heridentificeerbaar zijn).

De KSZ hanteert daartoe volgende regels:

Geen exacte data

Er worden geen exacte data meegedeeld. Bij voorkeur worden data teruggebracht tot jaar of jaar en kwartaal. Maand en kwartaal kunnen worden meegedeeld indien het onderzoek dit vereist. Een andere mogelijkheid is om het aantal dagen te bepalen tussen 2 data en dat aantal in te delen in klassen.

Geen exacte bedragen

Er worden geen exacte bedragen meegedeeld. Bedragen moeten steeds in klassen worden ingedeeld.

Geen precieze woonplaatsen

Er worden geen adressen van woonplaatsen meegedeeld. Dit gegeven wordt bij voorkeur zo hoog mogelijk geaggregeerd. Hoe lager het niveau, hoe meer moet worden aangetoond waarom dit niveau noodzakelijk is.

Toepassen van de 1 tot 3-regel of 1 tot 9-regel (herkomst)

Deze regel wordt toegepast in het geval er kruistabellen worden gevraagd. Indien een bepaalde cel (een bepaalde combinatie van variabelen) een te klein aantal personen vertegenwoordigt, i.e. minder dan 3 of minder dan 9, dan wordt het effectief aantal vervangen door resp. 1 tot 3 of 1 tot 9.

Variabelen in ruimere klassen indelen

Deze regel kan zowel worden toegepast in het geval van gepseudonimiseerde persoonsgegevens als in het geval van anonieme (geaggregeerde) gegevens. Leeftijdsklassen, loonklassen enz kunnen verruimd worden.

Van één kruistabel meerdere maken, waarbij telkens een deel van de variabelen gekruist wordt

Wanneer blijkt dat een kruistabel te veel kleine cellen bevat met kleine waarden, dan kan het een oplossing zijn om van één kruistabel meerdere te maken, waarbij telkens een deel van de variabelen gekruist wordt.

Woonplaats	geslacht	leeftijdsklasse	nationaliteit	Socio-economische positie	aantal
Brugge	1	18-24	Belg	Werkend	1
Gent	2	25-34	EU	Werkloos	2
Kortrijk	2	25-34	EU	Werkend	1
Aalst	1	35-49	Afrika	Werkloos	1
...					



Woonplaats	geslacht	leeftijdsklasse	aantal	Woonplaats	nationaliteit	Soc-eco positie	aantal
Brugge	1	18-24	14	Brugge	Belg	Werkend	37
Gent	2	25-34	19	Gent	EU	Werkloos	26
Kortrijk	2	25-34	24	Kortrijk	EU	Werkend	18
Aalst	1	35-49	20	Aalst	Afrika	Werkloos	54
...							

2-fasen methode

Deze methode wordt toegepast in het geval van gepseudonimiseerde persoonsgegevens waarbij een grote databank moet geconsulteerd worden. Deze databank wordt op een server van de KSZ geplaatst in een beveiligde omgeving. De onderzoeker krijgt toegang tot deze databank binnen de lokalen van de KSZ. Ter voorbereiding wordt aan de onderzoeker een kleine steekproef uit die databank meegedeeld. De idee is dat de onderzoeker hierop zijn applicaties ontwikkelt en test en ze vervolgens uitvoert op de gehele databank in de lokalen van de KSZ. De onderzoeker neemt alleen de resultaten van die uitvoering mee, in de vorm van geaggregeerde gegevens.

Scrambelen van data

Scrambelen is een techniek waarbij de reële data worden vervangen door data die geen werkelijke situatie meer voorstellen. Waarden kunnen vervangen worden door fictieve, of er kan een omwisseling van waarden gebeuren (bv. de woonplaats Antwerpen wordt vervangen door de woonplaats Hasselt). Er kunnen ook bepaalde voorwaarden worden opgelegd (bv. waarden moeten binnen een bepaalde range blijven, bepaalde waarden mogen niet voorkomen, enz.). Er kunnen ook voorwaarden worden opgelegd tussen bepaalde variabelen zodat bepaalde verbanden bewaard blijven (bv. een waarde X van variabele A gaat altijd samen met een waarde Y van variabele B). Gescrembelde datasets dienen niet om er analyses op uit te voeren, maar om applicaties voor te bereiden die kunnen uitgevoerd worden op de reële data op een beveiligde server binnen de lokalen van de KSZ. De onderzoeker neemt alleen de resultaten van die uitvoering mee, in de vorm van geaggregeerde gegevens.