

Maskeren kolommen met bijzondere persoonsgegevens

Het principe is eenvoudig:

- De kolommen die bijzondere persoonsgegevens bevatten onleesbaar maken door bv een zwarte balk of iets dergelijks.

Dit kan voor een groot deel geautomatiseerd gebeuren als je series met scans hebt die een grote uniformiteit bezitten. Dan is nabewerking altijd nog nodig voor afwijkingen op de kaarten, maar dat is redelijk overzichtelijk.

De kolom "religie" op de gezinskaarten is wel redelijk smal; dat maakt het lastig om goed af te dekken dan wel niet te veel af te dekken.

In Amsterdam zijn enkele velden (vierkante/rechthoekige vakjes, geen kolommen) op de persoons- en archiefkaarten (opvolgers van de gezinskaarten) afgedekt. Er is daarbij sprake van vier modellen kaarten die gebruikt zijn. Eenvoudig herkenbaar, mede doordat de verschillende modellen in verschillende series gescand zijn. En steeds op min of meer dezelfde wijze gecropt dat is: afgesneden. Dit kan verschillende 'kantlijnen' van de scans geven zodat steeds een ander afdekmodel nodig is voor de kolom die je wilt afdekken. Er is voor elk model kaart een vast afdekmodel gemaakt. En die zijn geautomatiseerd op de gescande kaarten 'gelegd'. Als een kaart ietsje anders gecropt is, dan is het zwarte vakje dus ietsje verkeerd geplaatst. De kosten voor het maken van de afdekmodellen en het vervolgens toepassen op de afzonderlijke kaarten waren verwaarloosbaar klein. Er is uiteraard wel sprake van dubbele opslag in je e-depot: de serie kaarten zonder afgedekte velden en de serie kaarten met afgedekte kaarten.

Het is minder effectief als de scans geen series bevatten met dezelfde layout/vorm waardoor het erg bewerkelijk wordt om de betreffende kolom(men) geautomatiseerd onleesbaar te maken.

Zo zijn in Amsterdam door de tijd heen in ieder geval drie verschillende modellen voor de gezinskaarten gebruikt. Daarnaast zijn de originele gezinskaarten in de oorlog verloren gegaan; de scans zijn gemaakt vanaf een microfilm waarop de gezinskaarten staan. Maar niet elke gezinskaart is met evenveel lege ruimte langs de rand van het filmbeeld gefotografeerd, waardoor ook bij gezinskaarten van eenzelfde model de zwarte balk niet automatisch op de goede plek zal komen te staan.

Bij het Haags Gemeentearchief is dit ook het geval: zeker drie formaten zijn in de verzameling gezinskaarten terug te vinden. De kaarten zijn ook verschillend gecropt waardoor de af te dekken gegevens zich niet op vaste plekken op de scan bevinden. Geautomatiseerd afdekken lijkt daarom erg lastig te zijn.

Bovendien moeten met deze methode de bijzondere persoonsgegevens van de hele periode afgedekt blijven totdat zeker is dat alle personen op de kaart niet meer in leven kunnen zijn. Dit houdt in dat de kaarten tot minstens 2049 gedeeltelijk afgedekt zullen blijven. Dit is wat het Haags Gemeentearchief betreft onwenselijk.

Kosten van het maskeren: Hangt af van grootte van het bestand en hoeveel modellen van de kaarten er in de loop van de tijd zijn gebruikt, dus hoeveel typen maskers er gemaakt moeten worden. Voor een groot bestand van circa een miljoen kaarten bij een grotere dienst waren de kosten van de software voor de afdekmodellen inclusief het toepassen op de scans ongeveer € 7.500. Bij een kleinere dienst is eigen software aangeschaft voor € 4.500,-, maar heeft men er uren van eigen medewerkers in moeten investeren. Afhankelijk van de beginsituatie en de formaten is de schatting dat dit ongeveer 300 uur gaat kosten.

De vraag die deze methode oproept: als deze kolommen (bv. godsdienst, gezindte) onleesbaar zijn, mag je dan toch de gegevens uit bevolkingsregister en/of gezinskaart online en doorzoekbaar aanbieden?

Wordt betrokken in de risico-analyse (PIA) die de werkgroep AVG in de periode februari-april 2019 uitvoert.