

Er zit ontzettend veel meer variabiliteit in (eerste letters van) voor- en achternamen dan in data zoals geboortedatum.

Voor voornamen zijn deze gegevens niet publiek beschikbaar. Als ik naar achternamen kijk in een [export](#) van de familienamenbank, gaat het (met accenten verwijderd) bijvoorbeeld om twee miljoen mensen met de eerste letter B, en slechts 15000 mensen met een eerste letter Q (initials_count.txt). Dat zijn overigens overschattingen vanwege afronding naar boven (minimaal 5 per unieke achternaam).

Neem je de lengte van de namen daarin mee (de asterisk-optie), dan deel je een achternaam beginnend met een B met lengte 4, 7, 6 elk met rond de 350k mensen (initial_length_count.txt). Met een long tail met veel unieke combinaties van letter en lengte.

Qua uitvoering aan de deur lijkt me de versie met asterisken lastig, vanwege de moeilijke vergelijking met het identiteitsbewijs. Ik zou dan eerder voor de geboortemaand kiezen om het bewijs mee uit te breiden in de gevallen dat dat nodig is.

Als ik dezelfde distributie van voornaam-initiaal neem als bij de achternaam, deel je de initialen BB (worst case) plus een dag van de maand met $2 \text{ mln} * 0.12 * 1/31$ met ~8 duizend andere mensen. Op het eerste gezicht lijkt dat veilig genoeg. Bij het toevoegen van een geboortemaand wordt dit verder teruggebracht naar ~700 (iets minder buiten de [zomermaanden](#)).

Als je wilt voorkomen dat mensen uniek identificeerbaar worden, en een initiaal is deel van de dataset, dan zal de set aan gegevens moeten verschillen per geval, bijvoorbeeld door minder gegevens in het testbewijs te ondertekenen als de combinatie al uniek genoeg is.

Met vriendelijke groet,

5.1.2e

On 04/03/2021 22:23, [5.1.2e](#) wrote:

AIL,

We hebben weinig tijd meer om een keuze te maken hoe ons niet-smart phone alternatief er uit ziet. Het wetsvoorstel met Memorie van Toelichting moet zsm afgerond worden en voor advies doorgestuurd worden naar de AP. Gistermiddag hebben we nog met translink gesproken over hun smartcard alternatief. We merken echter dat dat het nog niet helemaal is (er wordt niet aan alle eisen voldaan [zo moet de controle app online verificatie doen], vergt het aanpassingen aan de infrastructuur van de teststraten en vergt het simpelweg meer tijd om alle details uit te werken.

Daarom heb ik vanmiddag met [5.1.2e](#) en later telefonisch met [5.1.2e](#) en Spoor 2 nogmaals alle alternatieven verkend en zijn we uitgekomen op de variant dat de geteste persoon zelf zijn bewijs uitprint op gewoon papier. Deze voldoet aan bijna alle eisen, als we accepteren dat iemand wel toegang moet hebben tot een computer + printer of dat hij zich daarbij laat ondersteunen.

Zie de bijlage voor de beschrijving.

Graag jullie spoedige reactie morgen of jullie hier mee kunnen instemmen. Zo niet, dan uiteraard graag met een verbetervoorstel of aanvullende maatregelen.

Er zit nog wel een vraag aan vast. Een vervolgrisico is namelijk dat er een online handel ontstaat van testbewijscodes met corresponderende initialen en geboortedag. Om dit tegen te gaan zijn er verschillende maatregelen denkbaar (die gecombineerd ingezet kunnen / moeten worden):

1. Handhaven door dergelijke vormen van misbruik strafbaar te stellen en actief sites down te brengen.

2. De digitale handtekeningen van op dergelijke site verschenen QR codes ongeldig verklaren. De controle app download eens per x tijd een lijst van ongeldige handtekeningen en controleert daar op.

3. De kans verkleinen dat er een match is met de naam op het identiteitsbewijs door bv uitbreiden van het aantal letters of geboortedatum. Of als alternatief het opnemen van de initiaal + uitloopasteriksen voor elk karakter in de naam.

5.1.2e : Kunnen jullie uitrekenen voor punt 3 welk alternatief van persoonsgegevens de kans voldoende verkleind én voldoende privacy preserving is en die geen ingewikkelde uitvoeringsconsequenties heeft?

5.1.2e

5.1.2e

Ministerie van VWS

06- 5.1.2e

--

Beamed from the universe with a mobile device