

BEVEILIGING EN UITLEVERING RESULTAATBESTANDEN NSE 2020

1. INLEIDING

Deze notitie beschrijft de wijze waarop de databestanden van de Nationale Studenten Enquête (NSE) 2020 worden beveiligd en uitgeleverd. Het betreffen het Openbare Benchmark bestand (BM) en de instelling specifieke bestanden waaronder het instellingsbestand (IB) en de bestanden met open vragen (OV) en die met betrekking tot het vragenblok 'Studeren met een Functiebeperking, Aandoening en Ziekte' (FAZ) .

Paragraaf 2 beschrijft het BM en de n5 beveiligingsregel (n=respons) die net als in voorgaande jaren hierop wordt toegepast

Paragraaf 3 beschrijft de instelling specifieke bestanden en de N5 regel (N=populatie) die in 2020 wordt toegepast op alle respons in het IB. Hierbij wordt ook ingegaan op de verwerking van de labels die optioneel door onderwijsinstellingen kunnen worden aangeleverd.

In 2020 doen alle studenten "anoniem" met de NSE mee. In tegenstelling tot 2019 wordt er in 2020 aan de student geen toestemming gevraagd om het studentnummer en/of de volledige (en eventueel indirect herleidbare) respons te delen met de instelling waar de student zijn/haar opleiding volgt. De respons van de studenten wordt opgenomen in het BM en het IB, zónder dat deze tot de persoon herleidbaar zijn. In 2020 zullen geen instellingsvragen worden opgenomen en verwerkt.

2. DE N5 REGEL EN HET BM

Identificatie

Het BM is in principe openbaar toegankelijk. Daarom moet het conform de AVG worden beveiligd tegen directe en indirecte identificatie zodat het geen persoonsgegevens in de zin van de AVG meer bevat (zie artikel 4 lid 1, <http://www.privacy-regulation.eu/nl/artikel-4-definitie-EU-AVG.htm>).

Directe identificatie geldt als er een persoonsleutel als BSN of studentnummer is opgenomen in het bestand waardoor gegevens direct tot personen herleidbaar zijn. Er is sprake van indirecte identificatie (of indirecte herleiding) als een combinatie van antwoorden (bv. " een bachelor student van 99 jaar") de respondent mogelijk kan identificeren. Deze gegevens noemen we (indirect) identificerende variabelen.

Beveiliging van het BM met n5

Voor de beveiliging van het BM gebruikt Studiekeuze123 de n5 regel. Deze zegt dat het aantal respondenten (studenten) in een groep binnen een "sleutel" minimaal 5 moet zijn om indirecte herleidbaarheid te voorkomen. Hierbij wordt de "sleutel" gevormd door de combinatie van *alle* identificerende variabelen in het record van de respondent. Records van kleinere groepen worden hierdoor verwijderd, evenals alle overige identificerende variabelen (zoals leeftijd) die geen deel uitmaken van de gekozen sleutel.

Voor de onderbouwing van de n5 regel verwijst Studiekeuze123 naar de al jaren bestaande n5 regel van het Deskundigenpanel voor het tonen van de gemiddelden op de website, die zij sinds 2018 ook toepast op het benchmarkbestand (zie <https://www.studiekeuze123.nl/nse-instellingen/beslisregels-voor-tonen-van-nse-resultaten-in-de-studiekeuzedatabase>). De n5 regel is relatief soepel, het CBS hanteert vaak n10 of nog strengere regels voor publieke bestanden. Daarbij wordt tevens rekening gehouden met

response kennis, waarbij de onthuller incidenteel kennis kan hebben van deelname en antwoorden van respondenten.

Gehanteerde beveiligingssleutel voor BM

Studiekeuze123 hanteert in 2020, net als in 2019, als "sleutel" voor het BM de 5-dimensionale combinatie 'Brin (Instelling), Brinvolgnummer (vestiging), Croho-code (opleiding), Studievorm (vt/dt), Type Student (regulier of pre/post master)'. In woorden, er worden met de respons in het BM geen andere variabelen opgenomen die naast deze 5 identificerend kunnen werken. Daarnaast worden responsgroepjes binnen deze sleutel, die kleiner zijn dan 5, verwijderd. In het vervolg van dit stuk wordt deze sleutel aangeduid als de O5-sleutel. Deze keuze leidt tot eliminatie van maximaal ca. 0,8% van alle records. Opgemerkt wordt dat bijzondere variabelen (zoals de vragen m.b.t. handicap en studie en door instellingen aangeleverde labels) niet voorkomen in het openbaar BM, net als vorige jaren.

Samengevat: een met de n5 regel beveiligd openbaar BM bevat geen persoonsgegevens in de zin van de AVG en is beschermd tegen zowel directe als indirecte identificatie van de individuele student. Studiekeuze 123 levert slechts één openbaar BM voor ministeries, journalisten, instellingen, etc. Dezelfde n5 regel vormt tevens het uitgangspunt voor alle door Studiekeuze123 op het web te publiceren resultaten (bv. de gemiddelde tevredenheidsscores op de opleidingspagina's van Studiekeuze123) en de gegevens in het NSE-dashboard.

3. HET INSTELLINGSBESTAND (IB)

Geen toestemmingsvraag

Iedere instelling krijgt een instellingsbestand (IB) van de eigen respondenten/studenten. In tegenstelling tot 2019 wordt er in 2020 aan de student geen toestemming gevraagd om het studentnummer en/of de volledige (en eventueel indirect herleidbare) respons te delen met de instelling waar de student zijn/haar opleiding volgt. Dat betekent dat in 2020 alle studenten 'anoniem' voor de eigen instelling deelnemen aan de NSE.

Beveiliging van het IB met N5

Omdat het IB geen openbaar bestand is en alleen gebruikt wordt door de instelling zelf wordt op de respons een iets minder strenge beveiligingsregel toegepast, de N5 regel, waarbij N staat voor populatie. Deze zegt dat de populatie (dus niet alleen de respondenten) in een groep binnen een "sleutel" minimaal 5 moet zijn om indirecte herleidbaarheid te voorkomen. Hierbij wordt de "sleutel" net als in het BM gevormd door de combinatie van *alle* identificerende variabelen in het record van de respondent. Records van kleinere groepen worden verwijderd, evenals alle overige identificerende variabelen (zoals leeftijd) die geen deel uitmaken van de gekozen sleutel.

Gehanteerde beveiligingssleutel voor IB

Studiekeuze123 hanteert in 2020 standaard voor het IB dezelfde O5-sleutel als voor het BM. Het door Studiekeuze123 aldus beveiligde IB bestand bevat dan voor alle resultaten in 2020 niet de detailinformatie die de instellingen in voorgaande jaren wel kregen. Met name voor de labels is dit een beperking. Studiekeuze123 presenteert in paragraaf 4 hoe informatie op het niveau van labels toch op een veilige manier kunnen worden teruggeleverd. Met name sommige labels (bv. opleiding "Litouws") zijn sterk identificerend en moeten dus samen met de O5-sleutel op indirecte herleidbaarheid getest worden met de N5 regel.

De open vragen (OV)

Alle open antwoorden behorende bij het IB kunnen worden geleverd zonder beveiliging. Sommige open antwoorden/opmerkingen (bv. "ik ervaar etnische discriminatie") zijn weliswaar identificerend, maar bij het invullen ervan wordt de student vooraf geïnformeerd dat de antwoorden op deze vragen 1 op 1 zullen worden teruggeleverd aan de (eigen) onderwijsinstelling. Studiekeuze123 zal deze antwoorden in een apart bestand aan de instelling te leveren, zonder record ID en zonder sleutel. Hierdoor is koppeling met de overige IB-records (en daardoor onthulling) onmogelijk.

Respons met betrekking FAZ

In de NSE2020 zijn net als in voorgaande jaren vragen opgenomen voor studenten met een handicap. Deze worden vooraf gegaan door een eigen toestemmingsvraag. Studenten die geen toestemming geven beantwoorden deze vragen niet. De (uitgebreide) toestemmingsvraag impliceert dat bij toestemming de instelling alle antwoorden van de student van dit vragenblok mag hebben, ook nu de NSE2020 door alle studenten "anoniem" wordt ingevuld. Het betreft hier Bijzondere Persoonsgegevens in de zin van de AVG, die tevens (indirect) identificerend zijn. De juridische grondslag voor het mogen terug leveren aan de (eigen) onderwijsinstelling is 'toestemming'.

Studiekeuze123 zal deze antwoorden in een apart bestand aan de instelling te leveren, zonder record ID en zonder sleutel. Hierdoor is koppeling met de overige IB-records (en daardoor onthulling) onmogelijk, terwijl de informatie over Functiebeperking, Aandoening en Ziekte in het IB toch volledig geleverd kan worden.

4. DE LABELS

Optioneel kan een onderwijsinstelling labels aanleveren. Zoals beschreven in de GUO mogen de labels geen direct herleidbare informatie naar studenten bevatten. De instelling is verantwoordelijk voor de juridische grondslag van de aanlevering van de labels.

Instellingen kunnen Studiekeuze123 verzoeken de 5-dimensionale beveiligingssleutel uit te breiden met Studiejaar (wel/niet 1e jaars) en of enkele van de zelf aangeleverde labels. Op deze manier krijgen instellingen in het IB met de respons meer variabelen terug. Echter, de N5 regel wordt toegepast op deze verlengde "sleutel" waarbij ook sneller kleinere populatiegroepjes kunnen ontstaan.

Samengevat: een met de N5 regel beveiligd IB bevat dus geen persoonsgegevens in de zin van de AVG. En met een "beveiligingssleutel" bestaande uit meer dan de 5-dimensionale combinatie kunnen door Studiekeuze123 meer variabelen worden teruggeleverd maar zullen meer records moeten worden verwijderd vanwege de toegenomen kans op identificatie. Onderwijsinstellingen kunnen voorafgaand aan de keuze welke variabelen zij teruggeleverd willen zien in het IB een maatwerkberekening laten doen door Studiekeuze123 om te onderzoeken hoeveel procent van de respons op deze manier verloren gaat.

5. BEGRIPPENLIJST

Benchmarkbestand	Openbaar bestand ontdaan van persoonsgegevens en beveiligd met de n5 regel.
Instellingsbestand	Besloten bestand met alleen de respons van 1 instelling, alleen toegankelijk voor enkele medewerkers binnen 1 instelling.
FAZ (Functie, Aandoening, Ziekte)	Apart blok vragen binnen de NSE voorafgegaan door een toestemmingsvraag.
O5-sleutel	Combinatie van Brin, Brinvolgnummer, Crohocode, TypeStudent en opleidingsvorm3 die wordt gebruikt voor de beveiliging van de bestanden. (zie n5 regel en N5 regel)
n5-regel	Regel die zegt dat op de combinatie van de O5 sleutel minimaal 5 respondenten aanwezig moeten zijn in het benchmarkbestand. De kleine n slaat op respons.
N5-regel	Regel die zegt dat op de combinatie van de O5 sleutel minimaal 5 studenten aanwezig moeten zijn in het instellingsbestand. De grote N slaat op populatie.
Directe persoonsgegevens	Zaken als geslacht en geboortedatum.
Indirecte persoonsgegevens	Zaken als welke opleiding een student volgt.
Bijzondere persoonsgegevens	Gegevens over FAZ.
Labels	Kenmerken die een instelling mee kan geven bij het aanleveren van de populatiegegevens.