

Advies aanvraag herleidbaarheid onderwijs datasets

Ministerie van Onderwijs, Cultuur en Wetenschap

Floor Terra, Sjoera Nas en Machiel van der Wal

8 april 2026

INHOUDSOPGAVE

<i>Managementsamenvatting</i>	4
1 <i>Inleiding</i>	6
2 <i>Over PIAAC</i>	6
3 <i>Over TALIS</i>	7
4 <i>Methodologie Privacy Company</i>	7
5 <i>Getroffen maatregelen TALIS</i>	10
6 <i>Voorgenomen maatregelen PIAAC</i>	11
7 <i>Antwoorden op de vragen van OCW</i>	13
7.1 Kunnen de 2500+ toetsscores (deelnemende individuen) in de twee onderzoeken zelfstandig worden beschouwd als persoonsgegevens in de zin van de AVG?	13
Zijn individuen, op basis van (een combinatie van) de 14 resterende onderdrukte variabelen in combinatie met de toetsscores, (PIAAC) of op basis van de gepubliceerde data (TALIS) identificeerbaar?	13
7.2	13
7.3 Vereist identificatie van een individu een excessieve inspanning voor OCW of eventuele (elke) derde?	14
7.4 Bij welke combinaties van variabelen bestaat er geen mogelijkheid tot identificatie (of vereist dit een excessieve inspanning)?	15
7.5 Welke juridische criteria zijn van toepassing om te bepalen of toetsscores als persoonsgegevens moeten worden behandeld?	15
7.6 Kunnen jullie de opvatting van het CBS volgen dat toetsscores sowieso als gevoelige gegevens zijn aangemerkt?	16
7.7 Zijn alle toetsscores herleidbare persoonsgegevens?	21
7.8 Is er volgens de wet geen sprake van een persoonsgegeven als identificatie van een individu een onevenredige inspanning vereist en in de praktijk ondoenlijk is? Is dat hier van toepassing? Wat betekent de overweging 26 uit de AVG voor deze casus?	21
7.9 Gegeven de voorgaande uitkomsten: Kunnen de uitkomsten van beide onderzoeken in de huidige vorm conform AVG worden gepubliceerd?	23
7.10 Indien nee, welke aanpassingen zijn minimaal noodzakelijk om publicatie mogelijk te maken, met behoud van de grootst mogelijke wetenschappelijke waarde?	24
7.11 Welke maatregelen, in lijn met artikel 89 AVG, kunnen worden genomen om voor toekomstige onderzoeken hergebruik voor wetenschappelijk, historisch of statistisch onderzoek mogelijk te maken?	24

7.12	Zijn er omstandigheden waarbij onderzoeksgegevens, bestaande uit persoonsgegevens, beschikbaar kunnen worden gemaakt?	25
7.13	Is hergebruik voor wetenschappelijk, historisch of statistisch onderzoek wettelijk mogelijk wanneer de oorspronkelijke verwerking op toestemming van de betrokkene(n) berust?	26
7.14	Is hergebruik voor wetenschappelijk, historisch of statistisch onderzoek wettelijk mogelijk wanneer de oorspronkelijke verwerking onrechtmatig tot stand is gekomen, bijvoorbeeld wanneer er geen grondslag was voor de verwerking of niet is voldaan aan de informatieverplichting (artikel 13/14 AVG) tijdens de oorspronkelijke verwerking?	33
7.15	Hoe kunnen we een afwegingskader ontwikkelen voor het evenwicht tussen privacy-bescherming en wetenschappelijke bruikbaarheid?	33
7.16	Welke best practices van andere (academische) instellingen kunnen worden overgenomen?	34
7.17	Wat moeten we procedureel wijzigen om dit soort kwesties in de toekomst voor te zijn?	35
7.18	Hoe verhoudt de kleine kans van identificatie zich tot de statistisch kleine kans in de CABR casus? En waarom voldoet publicatie in een PUF aan het subsidiariteitsvereiste, kijkend naar de CABR casus?	35
<i>Bijlage A – TALIS-velden met een hoog niveau van koppelbaarheid</i>		37
<i>Bijlage B – PIAAC-velden met een hoog niveau van koppelbaarheid</i>		41

Lijst van tabellen

Tabel 1: Eerste 20 rijen van alle TALIS-velden met een hoog niveau van koppelbaarheid	10
Tabel 2: Voorbeelden van datavelden met een koppelbaarheidsniveau 2	23

Managementsamenvatting

Het ministerie van OCW heeft advies gevraagd aan Privacy Company over de herleidbaarheid van deelnemers in (opgeschoonde) datasets van de twee onderwijs-gerelateerde onderzoeken PIAAC en TALIS. De OESO organiseert deze onderzoeken om internationale vergelijkingen mogelijk te maken. De OESO maakt niet alleen rapporten openbaar met algemene vergelijkingen, maar ook de onderliggende (opgeschoonde) datasets met vragen en antwoorden, in de vorm van een *Public Use File* (PUF).

Het ministerie heeft al een aantal specifieke maatregelen en waarborgen getroffen om identificatie van de deelnemers in deze datasets te voorkomen, maar wilde advies of deze maatregelen voldoende zijn, of dat er aanvullende maatregelen en waarborgen noodzakelijk zijn, ook met het oog op toekomstige onderzoeken op onderwijsgebied.

Dit advies beantwoordt puntsgewijs de 18 vragen die het ministerie heeft voorgelegd, specifiek over PIAAC en TALIS, maar ook met algemene kaders en uitleg zodat het ministerie de beoordeling in de toekomst zelf kan doen, en een beleidskader hiervoor op kan stellen.

Privacy Company heeft alle variabelen in de beide datasets beoordeeld op mate van koppelbaarheid aan externe datasets, en entropie (de hoeveelheid bits in een antwoord). Het Kohnstamm Instituut van de UvA (dat de analyses verricht voor het ministerie) heeft deze analyse gebruikt om de datasets verder op te schonen van gegevens met een koppelbaarheid van 2 (met een paar uitzonderingen). De conclusie is dat de pseudonieme data in deze datasets na deze opschoning redelijkerwijs niet meer te herleiden zijn naar individuele deelnemers in de zin van de AVG. De data zijn niet 'anoniem' of 'statisch veilig', maar zijn wel zodanig bewerkt dat de kans op mogelijke herleidbaarheid is geminimaliseerd.

Privacy Company heeft ook de omstandigheden bestudeerd waaronder de gegevens zijn verzameld (zoals grondslag en informatie). Dit leidt niet tot een onherstelbaar onrechtmatige gegevensverwerking. Voor de publicatie van de opgeschoonde TALIS dataset in een PUF kan het ministerie een beroep doen op evenredigheid van de verdere verwerking als vrijwel onherleidbare pseudonieme persoonsgegevens. Voor de opgeschoonde PIAAC dataset kan het ministerie hier geen beroep op doen, omdat de toepasselijkheid van het openbaarmakingregime uit de Wet CBS aan publicatie in een PUF in de weg staat, zie paragraaf 1.7.6.

Om het risico op herleidbaarheid structureel te mitigeren, adviseert Privacy Company het ministerie bij toekomstige onderzoeken de volgende 9 maatregelen te nemen:

1. Ontwikkel een afwegingskader om voorgenomen onderzoeksverwerkingen te kunnen toetsen aan de principes van vertrouwelijkheid, legitimiteit en proportionaliteit, data-minimalisatie, goed fatsoen, transparantie en de juiste grondslag. Zie paragraaf 1.7.15.
2. Dring bij de OESO aan op publicatie van opgeschoonde datasets in de vorm van een RUF, een Restricted Use File, in plaats van een PUF, een Public Use File. Belangstellenden moeten dan bij de OESO een aanvraag doen voor toegang tot de opgeschoonde datasets. Verbindt in samenspraak met de OESO aan het gebruik van de RUF de voorwaarde dat dit gebruik beperkt moet zijn tot verantwoorde onderzoeksdoeleinden. Dit levert extra doorlopende administratieve lasten op.¹
3. Dring bij de OESO aan om de vragenlijsten te laten controleren op entropie en koppelbaarheid. Vermijd vragen die naar verwachting antwoorden met een hoge entropie en koppelbaarheid opleveren. Vervang die vragen door vragen naar bandbreedtes (van bijvoorbeeld leeftijd of werkervaring).

¹ Privacy Company heeft geen onderzoek gedaan naar de omvang van deze extra administratieve lasten.

4. Verwerk in toekomstige OESO-onderzoeken persoonsgegevens op basis van de grondslag van de noodzaak voor het vervullen van een taak van algemeen belang, ga niet uit van toestemming. Maak onderscheid tussen vrijwillige deelname aan onderzoek, en de grondslag voor de verwerking van vervulling van een taak van algemeen belang.
5. Informeer de deelnemers adequaat, op een voor hun leeftijd en begrip begrijpelijk niveau, over het doel van de vragenlijsten en de grondslag (géén toestemming). Persoonsgegevens zijn “alle informatie over een geïdentificeerde of identificeerbare natuurlijke persoon”, dus zijn de data in de dataset niet anoniem in de juridische zin van het woord. Gebruik daarom niet meer het woord ‘anoniem’, maar leg uit dat de resultaten zoveel mogelijk zijn ontdaan van gegevens die tot herleidbaarheid kunnen leiden. Privacy Company adviseert om een flyer te maken met verwijzing naar de website van de OESO met de datasets en de onderzoeksresultaten zodat de deelnemers ook na afloop van het onderzoek nog eens kunnen opzoeken wat er met hun antwoorden is gebeurd. Geef informatie over de toegepaste waarborgen om herleidbaarheid te voorkomen.
6. Zorg bij de OESO voor een ethische en privacytechnische toets nadat de expert commissies de vragenlijsten hebben opgesteld, om te voorkomen dat de onderzoeken vragen bevatten naar bijzondere persoonsgegevens of gegevens van strafrechtelijke aard, of vragen waaruit indirect dit soort bijzondere karakteristieken kunnen worden afgeleid.
7. Pas een nog strengere toets toe bij vragen aan minderjarigen (zoals het PISA onderzoek onder 15-jarigen).
8. Geef opgeschoonde brondata pas vrij voor openbare publicatie nadat eventuele koppeltabellen naar ID's van identificeerbare deelnemers zijn vernietigd. Voorkom sowieso het bijhouden van koppeltabellen als het enige doel is om mensen hun toestemming in te kunnen laten trekken.
9. Werk samen met SURF² om *best practices* voor onderzoek op te halen bij de hoger onderwijsinstellingen.

² SURF is de ict-coöperatie van onderzoeks- en onderwijsinstellingen, en is in de ogen van Privacy Company de meest aangewezen organisatie om de huidige *best practices* op halen bij haar leden.

1 Inleiding

Nederland doet mee aan meerdere onderwijs-gerelateerde onderzoeken zoals PIAAC en TALIS die vanuit de OESO worden georganiseerd om internationale vergelijkingen mogelijk te maken. De OESO maakt niet alleen rapporten openbaar met algemene vergelijkingen, maar ook de onderliggende datasets met vragen en antwoorden.

Het ministerie van OCW heeft advies gevraagd over de herleidbaarheid van de data in deze grote bestanden, met inachtneming van een aantal specifieke maatregelen en waarborgen die Nederland al heeft getroffen, of zou moeten treffen voor de publicatie. Die maatregelen en waarborgen worden hieronder toegelicht, na een korte uitleg over de inhoud van de PIAAC en TALIS onderzoeken en toelichting op de methodologie van Privacy Company.

Dit advies beantwoordt puntsgewijs de 18 vragen die het ministerie heeft voorgelegd om in de toekomst zelf de beoordeling te kunnen doen, en een beleidskader hiervoor op te kunnen stellen.

Onderzoeker Floor Terra heeft technisch onderzoek uitgevoerd naar de feitelijke datasets.

De input en feedback van het CBS (en haar FG) op de vorige versie van 29 september 2025 zijn in dit advies verwerkt.³

2 Over PIAAC

Nederland is één van de 31 deelnemende landen aan onderzoek naar de taal- en rekenvaardigheden en het probleemoplossend vermogen van volwassenen (van 16 tot 65 jaar). Dit onderzoek heet PIAAC (*Programme for the International Assessment of Adult Competencies*), wordt elke 10 jaar uitgevoerd, en wordt internationaal gecoördineerd door de OESO. De meest recente cyclus is in 2022/23 uitgevoerd. Het rapport met de uitkomsten van dit vergelijkende onderzoek is aangeboden aan de Tweede Kamer.⁴ De OESO publiceert de data van alle deelnemende landen in een Public Use File (PUF). Hierdoor heeft in principe iedereen ter wereld toegang tot de data. Op aanraden van het CBS en de Autoriteit Persoonsgegevens heeft het ministerie de Nederlandse data wel aan OESO verstrekt, maar nog geen toestemming gegeven voor publicatie in de PUF.

Het onderzoek bestaat uit een interview van circa een uur, en een reeks toetsen die deelnemers op een tablet moeten uitvoeren om hun vaardigheden te meten (circa 45 minuten). Het CBS heeft voor het PIAAC-onderzoek een steekproef getrokken van NAW-gegevens (Naam, Adres en Woonplaats) in Nederland, met leeftijd en geslacht van de deelnemers. Deze laatste 2 gegevens waren noodzakelijk om aan de deur te kunnen controleren of de deelnemer ook de beoogde participant was.

³ De input en feedback betrof een methodologische reactie, een juridische reactie en een reactie van de Functionaris Gegevensbescherming (FG) van het CBS. Privacy Company heeft hier vervolgens op 10 februari 2026 op gereageerd, waarna op 26 februari 2026 een door het ministerie van OCW georganiseerd gesprek heeft plaatsgevonden tussen het CBS en Privacy Company. De FG van het CBS heeft hierna nog een onthullende casus (combinatie van variabelen die in de praktijk leiden tot herleidbaarheid van een uniek individu) gedeeld met Privacy Company. Deze casus is gebaseerd op een eerdere versie van de dataset. Een aanzienlijk deel van deze variabelen is reeds aangemerkt als een variabele met een hoog risico en is vervolgens geschrapt of verder geaggregeerd.

⁴ OECD Skills Studies, Do Adults Have the Skills They Need to Thrive in a Changing World? SURVEY OF ADULT SKILLS 2023, URL: <https://open.overheid.nl/documenten/dpc-e971ce7e9b60130e601e5da1ed21d35630a2cb0e/pdf>.

Onderzoeksbureau Kantar heeft deze gegevens gebruikt om langs deze adressen te gaan en de geselecteerde deelnemers gevraagd om op een iPad een vragenlijst te beantwoorden, met toetsen om de taalvaardigheid te meten. De antwoorden op de vragen zijn niet met het CBS gedeeld (en maken dus geen deel uit van de door het CBS verstrekte gegevens), maar het CBS heeft wel een controle gedaan op de representativiteit van de onderzoekspopulatie, en een wegingsfactor gemaakt voor Kantar om te compenseren voor ontbrekende deelnemers naar locatie, leeftijd en geslacht.

Kantar heeft de direct identificerende gegevens van de deelnemers verwijderd, en aan elke deelnemer een uniek random gegenereerd nummer toegekend. Kantar heeft de resulterende gepseudonimiseerde scores per deelnemer aan het internationale onderzoek consortium verstrekt voor schaling en analyse. Het gepseudonimiseerde analysebestand is daarna door het internationale consortium gedeeld met Kohnstamm.

De Nederlandse PIAAC dataset bevat ruim 2.960 variabelen.

3 Over TALIS

TALIS (*Teaching and Learning International Survey*) is een vijfjaarlijks onderzoek naar de werkomstandigheden en leeromgevingen van leraren en bestuurders. Het bestaat uit twee vragenlijsten: één voor leerkrachten en één voor schoolbesturen, met ieder respectievelijk 58 en 45 vragen (met ieder mogelijk meerdere subvragen). Deze vragenlijsten zijn het meest recent in 2024 afgenomen bij 251 random geselecteerde scholen. De OESO gaat de resultaten en datasets publiceren 7 oktober 2025.⁵

De onderzoeksmethodologie en datasets zijn beschikbaar op de website van OESO.⁶ Net als bij PIAAC worden de vragenlijsten en resultaten online gepubliceerd door de OESO in een PUF.

De Nederlandse dataset bevat 2.375 variabelen, verdeeld in vier bestanden, inclusief variabelen die in meerdere bestanden voorkomen, over leraren en bestuurders in het basis- en voortgezet onderwijs.

4 Methodologie Privacy Company

Privacy Company heeft verwerkersovereenkomsten gesloten met het ministerie en met het Kohnstamm Instituut om de PIAAC bestanden te ontvangen, met specifiek geheimhoudingsbeding. Privacy Company heeft onder geheimhoudingsbeding ook de TALIS bestanden ontvangen.

Privacy Company heeft de beide databestanden geanalyseerd op volgende twee factoren:

1. Een inschatting van de entropie per veld binnen de dataset (hoeveel informatie bevat de variabele)
2. Een inschatting van de herleidbaarheid (koppelbaarheid) van de variabelen met externe bronnen.

Privacy Company heeft (in samenwerking met het Kohnstamm instituut) per variabele een inschatting gegeven van de Shannon entropie.⁷ Dat is een schatting van de hoeveelheid informatie die een variabele bevat (hoeveel 'bits').

⁵ OECD, TALIS, URL: <https://www.oecd.org/en/about/programmes/talis.html>

⁶ OECD TALIS database 2018, URL: <https://www.oecd.org/en/data/datasets/talis-2018-database.html>

⁷ Wikipedia: Entropy (information theory), URL: [https://en.wikipedia.org/wiki/Entropy_\(information_theory\)](https://en.wikipedia.org/wiki/Entropy_(information_theory))

De entropie geeft aan hoeveel 'bits' aan informatie het antwoord bevat. Wanneer een variabele maar 1 mogelijk antwoord bevat, bijvoorbeeld in de dataset met alleen maar Nederlandse respondenten, is elke 'CountryID' hetzelfde. Dat levert nul bits aan entropie op, en is daarmee niet bruikbaar om aan externe datasets te linken.

Een vraag naar geslacht levert in de TALIS-set twee antwoorden op, man of vrouw, waarvan de antwoorden bijna 50/50 verdeeld zijn. In dat geval ligt de entropie ongeveer op 1 bit. De herleidbaarheid hiervan is beperkt. Wanneer het antwoord op een vraag zeldzame attributen bevat of een bereik heeft van bijvoorbeeld 16 tot 65 (zoals bij leeftijd) dan is de entropie hoog omdat er veel mogelijke uitkomsten zijn. De kans op herleidbaarheid is hoger bij een hogere entropie.

De entropie is ingeschat onder de aanname dat alle variabelen onafhankelijk zijn. In de praktijk zal dit waarschijnlijk niet altijd waar zijn waardoor de entropie te hoog is ingeschat.

Het CBS merkt in reactie op dit advies terecht op dat, kort samengevat, bij individuele uitschieters een hoger risico op herleiding kan ontstaan. Privacy Company heeft bewust voor deze benadering gekozen, op basis van twee overwegingen.

Ten eerste heeft Privacy Company op verzoek van het ministerie van OCW advies gegeven over de herleidbaarheid van gegevens conform de AVG, niet over de statistische veiligheid zoals die door het CBS is gedefinieerd op grond van de Wet CBS. De AVG bevat geen norm voor het (statistisch) aantoonbaar uitsluiten van onthullingsrisico's om te bepalen of gegevens wel of niet herleidbaar kunnen zijn naar natuurlijke personen. De AVG vraagt om het meewegen van de context, zoals de middelen die verwerkingsverantwoordelijken redelijkerwijs voor handen kunnen hebben.

Ten tweede heeft Privacy Company het advies in een beperkte tijdsspanne geschreven. Een methode om een grote hoeveelheid combinaties van variabelen in samenhang te analyseren kost meer tijd. Daarnaast zou Privacy Company de analyse herhaaldelijk opnieuw uit moeten voeren op basis van voorgenomen (voorlopige) keuzes in het nemen van maatregelen. Privacy Company heeft er daarom voor gekozen om alle variabelen als onafhankelijk te behandelen. Bij vragen die door sommige respondenten niet beantwoord zijn heeft Privacy Company de ontbrekende informatie niet meegewogen in de inschatting van de entropie. Er is bij elke vraag opnieuw geteld hoe vaak de vraag door Nederlandse respondenten is beantwoord en hoe vaak elk antwoord voorkomt.

Kort samengevat heeft Privacy Company twee aspecten beoordeeld:

- Een (niet exacte) inschatting voor de gemiddelde hoeveelheid informatie per variabele;
- Een inschatting (score 0, 1 of 2) op basis van ervaring of verschillende categorieën verwerkingsverantwoordelijken over informatie beschikken die eventueel gekoppeld kunnen worden aan de PIAAC-data waardoor betrokkenen herleid zouden kunnen worden.

Onderzoekers die persoonsgegevens verwerken, moeten in de praktijk altijd een afweging maken tussen de behoefte aan specifieke informatie en de risico's die het opnemen van die informatie met zich brengt. Privacy Company kan niet beoordelen hoe belangrijk de publicatie van specifieke variabelen is. Privacy Company heeft zich daarom gericht op een analyse per variabele zodat het ministerie die afweging zelf kan maken. Door alle mogelijke combinaties te analyseren, van alle keuzes, voorafgaand aan de publicatie van een subset van de variabelen, zou de analyse enorm complex worden. Omdat Privacy Company niet vooraf weet welke afwegingen het ministerie wil maken en omdat die afwegingen ook kunnen afhangen van de analyse van Privacy Company heeft Privacy Company ervoor gekozen om een advies te geven op basis van een conservatieve risicobeoordeling.

Door per variabele de entropie te bekijken geeft Privacy Company een worst-case beoordeling van het risico bij de opeenstapeling van variabelen voor herleidbaarheid/koppelbaarheid. Neem het volgende praktijkvoorbeeld, met een opeenstapeling van 6 variabelen:

1. Man
2. Met kinderen
3. Met 3 kinderen
4. Met 3 onvolwassen kinderen
5. Werkzaam als docent in Nederland
6. 30 jaar werkervaring als docent

De entropie van de variabelen 3, 4 en 6 is hoog. De koppelbaarheid van de zesde variabele schatten we ook in als hoog, omdat deze man misschien een LinkedIn profiel heeft waarin deze werkervaring staat. Dus variabele 6 wordt hoe dan ook onderdrukt. De koppelbaarheid van variabele 3 bestaat alleen bij het CBS, en krijgt in verband met een wettelijk verbod op onthulling van persoonsgegevens een 1. De koppelbaarheid van de variabelen 4 en 5 is ruimer. Er bestaan externe bronnen die over deze gegevens beschikken, zoals het consultatiebureau, het GBA, de school waar deze man werkt, familieleden en vrienden. Toch krijgen deze variabelen een koppelbaarheid van 1, omdat het gaat om kennis die je niet in openbare bronnen kunt opzoeken, en omdat deze kennis al bestaat bij deze partijen.

Hetzelfde voorbeeld zou wel tot een hogere kans van herleidbaarheid leiden als de data niet meer zijn geaggregeerd naar ranges, maar veel specifiekere zijn gemaakt, zodat er meer kans is op koppelbaarheid via externe bronnen

1. Gehuwde man van 58 jaar geboren in Ethiopië
2. Met kinderen
3. Met 2 jongens en 1 meisje
4. De jongste jongen is 8 maanden oud, de middelste jongen is 2.3 jaar oud en het oudste meisje is 4.8 jaar oud.
5. Werkzaam als docent op een basisschool in het dorp Goes.
6. 30 jaar werkervaring als docent in het basisonderwijs in de provincie Arnhem.

In dit geval hebben alle variabelen een hoge entropie, behalve de tweede, en krijgen de variabelen 1, 4, 5 en 6 een hoge koppelbaarheid van 2. Dan zou Privacy Company adviseren om die vier variabelen te onderdrukken, of te reduceren tot 'man', 'onvolwassen kinderen', 'docent in Nederland' en 'meer of minder dan 10 jaar werkervaring'. We zouden ook adviseren om de leeftijden van de kinderen weg te laten om koppelbaarheid te voorkomen, en er 'drie kinderen' van te maken, zonder onderscheid naar geslacht.

Daarnaast geeft Privacy Company een inschatting van de kans dat er externe bronnen zijn waaraan de variabele gekoppeld kan worden.

Het CBS vroeg om toelichting op het onderscheid tussen de koppelbaarheid met publieke externe bronnen en mogelijke niet-publieke externe bronnen. Privacy Company maakt dit onderscheid omdat het risico dat de gegevens in de praktijk daadwerkelijk kunnen worden herleid kleiner is als eventuele koppelingen alleen mogelijk zijn met niet-publieke bronnen. Ook zijn de waarschijnlijke negatieve gevolgen voor betrokkenen lager als zij herleid worden door partijen die al toegang hebben tot overlappende vertrouwelijke informatie over de betrokkenen.

De herleidbaarheid (koppelbaarheid) van de variabelen met externe bronnen is ingedeeld in drie niveaus:

0. Niet herleidbaar: er zijn geen externe bronnen waarin deze gegevens kunnen worden opgezocht.

1. Identificeerbare gegevens zijn nog op de school beschikbaar (voor OCW of een derde is niveau 1 niet herleidbaar).
2. De gegevens zijn mogelijk terug te vinden in een openbaar register of externe dataset.

Op basis van die analyse heeft Privacy Company in samenwerking met het Kohnstamm instituut doorgenomen welke variabelen het Kohnstamm instituut weg zou moeten laten uit de beide publicaties van datasets op de OESO website.

Het Kohnstamm instituut heeft ervoor gekozen om zoveel mogelijk variabelen te onderdrukken met een koppelbaarheidsniveau van 2, op variabelen na met een entropie van nul, en twee korte lijstjes specifieke uitzonderingen die het Kohnstamm graag zou willen behouden om internationale vergelijkingen mogelijk blijven te maken. Die uitzonderingen worden hieronder in paragrafen 5 en 6 toegelicht.

5 Getroffen maatregelen TALIS

De OESO had het ministerie een deadline opgelegd voor aanlevering van de onderzoeksdata van maandag 17 augustus 2025. Privacy Company heeft samengewerkt met het Kohnstamm instituut om die deadline te kunnen halen.

Het Kohnstamm instituut heeft toegelicht dat in de TALIS set sommige velden bijvoorbeeld een uniek ID hebben, maar alleen tijdelijk door de school herleidbaar zijn. Alleen de school beschikt over een koppeltabel tussen de identiteit van de deelnemer en de identiteit van de school, en de random ID's die hiervoor in de datasets zijn opgenomen. De school mag deze koppeltabel maar gedurende 1 jaar bewaren. Omdat het onderzoek meer dan een jaar geleden is afgerond, hebben deze velden een koppelbaarheid van 0 (nul) gekregen.

Bijlage A bevat de tabel met alle rijen met een koppelbaarheid van 2. Dit zijn er in totaal 146. [Tabel 1](#) geeft ter illustratie de betekenis weer van alle veldnamen van vragen met een koppelbaarheid van 2 voor de bestuurders in het primair onderwijs. Het Kohnstamm instituut heeft ervoor gekozen om al deze velden te onderdrukken (op 2 na) uit de dataset die ze op 17 augustus heeft aangeleverd aan de OESO, met uitzondering van de velden met een entropie van nul (groen gekleurd in de tabel). Dat waren er 22. In totaal zijn dus 124 variabelen onderdrukt.

De twee velden met een hoge koppelbaarheid die het Kohnstamm niettemin toch wil behouden (alleen in de lerarenbestanden (po en vo), niet in de bestuurder-bestanden, zijn:

- TT4G01 Gender
- T4TAGEGR Leeftijd in categorieën

Deze twee variabelen leveren een verwaarloosbaar identificatierisico op, ook wanneer beschouwd in combinatie met de resterende andere variabelen.

Tabel 1: Eerste 20 rijen van alle TALIS-velden met een hoog niveau van koppelbaarheid

Veldnaam	Vraag	Entropie	Koppelbaarheid
CNTRY	Vragenlijst bestuurders primair onderwijs - Country (Nederland)	0	2
IDCNTRY	Country ID	0	2
IDCNTPOP	Country Alpha Code and ISCED Level	0	2
IDLANG_CQ	Language questionnaire	0	2
TC4G01	Gender	0.91053306	2
TC4G02	Age	4.906317583	2
TC4G03	Highest level of formal education completed	1.139285534	2

	Years of work experience; regardless of full-time or part-time		
TC4Go4A	As a principal at this school	3.667422333	2
TC4Go4B	As a principal in total	4.33003027	2
TC4Go4C	In other school management roles (not principal)	3.649321792	2
TC4Go4D	As a teacher in total	4.78675565	2
TC4Go4E	In other non education jobs	2.40632579	2
TC4G11	School location description	1.92707992	2
TC4G12A	Percentage of school's total funding: - Government	2.503777382	2
TC4G12B	Student fees or school charges paid by parents or guardians	2.269906138	2
TC4G12C	Other (e.g. donations; grants; fundraising)	1.544675301	2
TC4G16A	ISCED levels and/or programmes taught: - ISCED 2011 Level 0	0	2
TC4G16B	ISCED 2011 Level 1	0	2
TC4G16C	ISCED 2011 Level 2: general education prgs	0	2
TC4G16D	ISCED 2011 Level 2: vocational or technical education prgs	0	2
TC4G16E	ISCED 2011 Level 3: general education prgs	0	2
TC4G16F	ISCED 2011 Level 3: vocational or technical education prgs	0	2
TC4G17	Current school enrolment	6.779984928	2
TC4G18A	Student admission factors: Residence in a particular area	1.703254408	2
TC4G18B	Students' record of academic performance	1.481008791	2
TC4G18C	Recommendation of feeder schools	1.947692557	2
TC4G18D	Parents' endorsement of instructional or religious philosophy of school	1.950766896	2
TC4G18E	Student requires or is interested in a special prog.	1.746472048	2
TC4G18F	Family members of current student	1.529266795	2
TC4G18G	Emergency situations	1.845151003	2
TC4G21A	School composition: Students with difficulties understanding instr. language	1.48610762	2
TC4G21B	Students [first language] diff. from instr. language	1.825475781	2
TC4G21C	Students with special education needs	1.665933271	2
TC4G21D	Students from socio-economically disadvantaged homes	1.831593377	2
TC4G21E	Students belonging to ethnic national minorities or Indigenous comm.	1.721127673	2
TC4G21G	Students who are immigrants or with migrant background	1.297471127	2
TC4G21H	Students who are refugees	1.159057012	2
T4Pa-qEGR	Principal age (grouped)	1.893801963	2
T4PHEDAT	Principal's highest level of formal education completed (grouped)	1.139285534	2
T4PEMPWH	Principal's employment status in terms of working hours (grouped)	0.999337504	2
T4SCHLOC	School location (grouped)	1.307472969	2
T4STRATIO	Student-teacher ratio	6.745444736	2
T4TPRATIO	Teacher-pedagogical support personnel ratio	5.712421875	2
T4TARATIO	Teacher-administrative or management personnel ratio	5.329572838	2
T4NUMSTUD	Number of enrolled students (grouped)	0.95583987	2
T4SCHISCN	Number of ISCED levels taught in the school	0	2
T4TTNRATIO	New teacher-teacher ratio	4.98094518	2

6 Voorgenomen maatregelen PIAAC

Het Kohnstamm instituut heeft in 2024 samen met collega onderzoeksinstituten uit de andere deelnemende landen gekeken naar de herleidbaarheid van variabelen, en het ministerie geadviseerd om 280 variabelen te onderdrukken die de deelnemers individueel herleidbaar zouden kunnen maken.

Het gaat het om de volgende variabelen:

- Herkomstland of -regio;
- Thuis taal;
- Inkomen;
- Geografische informatie (woonregio/provincie);
- Gezinssamenstelling;
- Gedetailleerde informatie over het werk dat een respondent doet en de sector (ISCO en ISIC 2 digit level);
- Specifieke opleidingsvariabelen met kleine aantallen respondenten (samenvoegen van ISCED 0 en ISCED 1 en/of ISCED 7 en 8);
- Alle variabelen die verwijzen naar specifieke cijfers/aantallen (zoals de leeftijd van respondenten, het aantal jaar dat een respondent in Nederland woont of gestopt is met werken, het aantal cursussen dat in het afgelopen jaar is gevolgd) zijn of onderdrukt, of ingedikt naar bredere categorieën. Vaak bevatte de set ook beide gegevens: zowel leeftijd naar bandbreedte van tien jaar, en exacte leeftijd.

Het CBS heeft daarna een analyse gedaan op de statistische veiligheid (conform het CBS handboek statistische beveiliging⁸) van de resulterende dataset, en voorwaarden gesteld om twee maatregelen te treffen: het aantal variabelen over de deelnemers terugbrengen tot maximaal 15 identificerende variabelen voor een Public Use File. Het CBS heeft specifiek gekeken naar zeldzame (combinaties van) categorieën, en geëist dat het ministerie alle gevoelige variabelen zou verwijderen. Daaronder vallen volgens het CBS ook de testscores.

Als het ministerie van OCW dit advies zou volgen, zou de Nederlandse bijdrage aan de internationale PUF nutteloos zijn: dan bevat de set immers geen data meer over de vaardigheden van volwassenen. Er is dan niks meer te vergelijken met andere landen. In ieder geval niet wat PIAAC gemeten heeft, want dat zijn juist die testscores.

De FG van het CBS heeft een signaal afgegeven bij de Autoriteit Persoonsgegevens dat OCW voornemens was een bestand openbaar te publiceren met gevoelige gegevens en dat dat niet conform de AVG zou zijn. De FG van OCW was het eens met de analyse van de FG van het CBS. De AP heeft het ministerie van OCW telefonisch geïnformeerd dat dit signaal ontvangen was, en dat de AP zich aansloot bij de analyse van het CBS. Daarop heeft de AP aan OCW gevraagd maatregelen te treffen om de publicatie met gevoelige gegevens tegen te gaan. Het ministerie van OCW heeft daarop de publicatie van de PUF bij de OESO uitgesteld om de gerezen vragen nader te kunnen uitzoeken, en ook met de OESO in internationaal verband te bespreken.

Privacy Company heeft de PIAAC dataset op dezelfde manier geanalyseerd als de TALIS dataset, dus voor elke variabele gekeken naar entropie en koppelbaarheid. Daar komt een lijst uit naar voren van 204 variabelen met een koppelbaarheid van 2. Deze lijst is opgenomen in Bijlage B. De lijst bevat ook nog 25 variabelen met een entropie van 0 (nul). Die voegen geen extra risico op herleidbaarheid toe, en zijn daarom groen gekleurd.

Het Kohnstamm instituut heeft aangegeven dat ze graag tien variabelen wil behouden met een koppelbaarheid van 2. Dit zijn:

- GENDER_R (gender)
- AGE10LFS (leeftijd in categorieën)

⁸ CBS Handboek Statistische Beveiliging, URL: <https://www.cbs.nl/nl-nl/corporate/2024/42/cbs-publiceert-hernieuwde-handboek-statistische-beveiliging>.

- EDCAT6_TC1 (opleiding)
- A2_Q03a_T (geboren in NL ja/nee)
- C2_D05 (employment status)
- ISCOSKIL4 (werkskills)
- E2_Q03 (sector)
- E2_Q04 (zelfstandige)
- ICTEVER (ooit ICT gebruikt)
- B2_Q11 (job related training)

Behoud van deze 10 variabelen leidt tot een iets hoger herleidbaarheidsrisico. Statistisch is het mogelijk om met deze 10 variabelen 1 persoon binnen een groep van 2.000 personen te herkennen. Dit heeft Privacy Company gebaseerd op de entropie van de variabelen. De variabelen bevatten samen ongeveer 11 bits aan entropie. Dat is gemiddeld genomen voldoende om specifieke personen in een populatie van 2.000 (2^{11} afgerond) uniek te identificeren. Privacy Company heeft daarbij geen algemeen populatie-onderzoek uitgevoerd. Omdat PIAAC een steekproef onder de volwassen Nederlanders betreft (en niet een bekende groep van 2.000 personen) is deze kans op herleidbaarheid nog steeds heel erg laag.

7 Antwoorden op de vragen van OCW

7.1 Kunnen de 2500+ toetsscores (deelnemende individuen) in de twee onderzoeken zelfstandig worden beschouwd als persoonsgegevens in de zin van de AVG?

Ja. Zoals uitgebreider toegelicht hierboven in de paragraaf over de methodologie heeft Privacy Company (in samenwerking met het Kohnstamm instituut) per variabele een inschatting gegeven van de entropie (hoeveel informatie de variabele bevat) en een (veel minder exacte) inschatting van de kans dat de variabele gekoppeld kan worden aan overige bronnen. Privacy Company heeft de dataset bekeken vanuit het uitgangspunt dat alle scores pseudonieme persoonsgegevens zijn als 1 van de variabelen in de set tot herleidbaarheid leidt. Daarmee gaat Privacy Company uit van het hoogst mogelijke risico, en heeft ze geen berekening gedaan van combinaties van datavelden die tot herleidbaarheid kunnen leiden.

Op basis van de analyse heeft (TALIS) of kan (PIAAC) het ministerie in de openbare set alle variabelen onderdrukken die kunnen worden gekoppeld aan externe gegevens. Door die maatregel is de kans op herleidbaarheid van de persoonsgegevens tot vrijwel nul teruggebracht. Maar er is dus géén sprake van statistisch veilige anonieme gegevens in de PUF, de standaard waar het CBS vanuit gaat.⁹

7.2 Zijn individuen, op basis van (een combinatie van) de 14 resterende onderdrukte variabelen in combinatie met de toetsscores, (PIAAC) of op basis van de gepubliceerde data (TALIS) identificeerbaar?

Het gaat om pseudonieme persoonsgegevens waarvan de herleidbaarheid vrijwel is uitgesloten na onderdrukking van velden met een hoge entropie en hoge koppelbaarheid.

⁹ CBS, 'Handboek Statistische Beveiliging: Info voor externen' (2^e editie), juni 2024, [URL:://www.cbs.nl/-/media/cbs/over-ons/organisatie/handboekstatistischebeveiliging2024.pdf](https://www.cbs.nl/-/media/cbs/over-ons/organisatie/handboekstatistischebeveiliging2024.pdf).

Zoals hierboven toegelicht beschouwt Privacy Company de beide datasets als pseudonieme persoonsgegevens. Om de kans op herleidbaarheid te minimaliseren, heeft Privacy Company van alle velden in kaart gebracht wat de entropie was, en de velden ingedeeld in drie vormen van herleidbaarheid: 0, 1 of 2. Na onderdrukking van de variabelen met een koppelbaarheidsscore van 2 acht Privacy Company identificatie van individuele deelnemers in hoge mate onwaarschijnlijk.

Privacy Company heeft geen vergelijking gedaan van alle Nederlands bestanden met de internationale gepubliceerde resultaten en kan dus niet zien welke variabelen in de Nederlandse set mogelijk onderdrukt zijn. Privacy Company weet ook niet wat de 14 resterende onderdrukte variabelen waren. Privacy Company heeft een benadering gekozen die het ministerie helpt om de risico's van publicatie van de datasets te minimaliseren. Dat is een andere benadering dan het CBS. Het CBS gaat uit van statistisch veilige anonimiteit.¹⁰ Privacy Company heeft de dataset niet getoetst aan anonimiteit, omdat die norm vrijwel niet uitvoerbaar is bij een openbare publicatie met zoveel variabelen, een dataset die bovendien onbeperkt beschikbaar blijft.

Privacy Company heeft onderzoek gedaan naar de twee meest recente PIAAC en TALIS datasets, beiden gebaseerd op onderzoek uitgevoerd in 2024. Het ministerie had de opgeschoonde PIAAC data al gedeeld met de OESO, maar het Kohnstamm instituut heeft nu een nieuwe opschoning doorgevoerd, aan de hand van de analyse van entropie en koppelbaarheid.

Het Kohnstamm instituut heeft op dezelfde wijze de TALIS data opgeschoond, en eind augustus gedeeld met de OESO.

Het ministerie heeft voor de PIAAC dataset géén toestemming gegeven om ze op te nemen in de PUF. De TALIS dataset is inmiddels wel opgenomen in de PUF.

7.3 Vereist identificatie van een individu een excessieve inspanning voor OCW of eventuele (elke) derde?

Ja. Na opschoning zijn de gegevens met een herleidbaarheidsniveau van 0 of 1 niet herleidbaar voor OCW en derde partijen.

De variabelen zijn beoordeeld op hun mate van herleidbaarheid (koppelbaarheid) en ingedeeld in drie niveaus (0, 1, 2), afhankelijk van in hoeverre zij deelnemers individueel identificeerbaar maken. Bij het niveau 0 vergt herleidbaarheid een buitensporige, excessieve inspanning of is het feitelijk onmogelijk. Dat is bijvoorbeeld het geval bij de variabele land: die is altijd Nederland, omdat het om een exclusief Nederlandse dataset gaat. Bij niveau 1 is herleidbaarheid (gedurende een jaar) mogelijk bij de bron. Bij TALIS is dat de school, bij PIAAC zou dat niet mogelijk moeten zijn, omdat het onderzoeksbureau geen direct identificerende gegevens verzamelt. De dataset bevat wel identifiers, zoals het *Person operational identification number (with check digit)*, *Sequential ID (randomly derived)*, *Interviewer ID* en *Sampling ID: Primary sampling unit (PSU) identification number*. Het Kohnstamm Instituut had in algemene zin aangegeven dat er geen koppeltabellen voor deze gegevens bestaan, maar het is de vraag of dat ook voor het interviewer ID geldt (ook een persoonsgegeven).

Voor OCW en derde partijen die de dataset gebruiken, zijn deze variabelen (ingedeeld in niveau 1) niet herleidbaar. Niveau 2 is koppelbaar met externe bronnen. Er is een aanzienlijk risico op terugvindbaarheid in openbare bronnen zoals het register van de Kamer van Koophandel of openbare overzichten van OCW.

¹⁰ Ibid.

7.4 Bij welke combinaties van variabelen bestaat er geen mogelijkheid tot identificatie (of vereist dit een excessieve inspanning)?

Variabelen met een koppelbaarheidsscore van 0 zijn volgens Privacy Company hoogstwaarschijnlijk niet te koppelen aan externe bronnen. Denk daarbij bijvoorbeeld aan een vraag waarbij een leraar een subjectieve mening of ervaring aangeeft over de leeromgeving. Variabelen met een score van 1 zijn mogelijk te herleiden met behulp van een paar besloten bronnen, zoals in een besloten administratie van de betreffende school zelf. Variabelen met een score van 2 zijn mogelijk terug te vinden in externe of zelfs openbare bronnen. Denk daarbij aan informatie die DUO als open data over scholen publiceert, of informatie die sommige leraren of bestuurders mogelijk op hun LinkedIn profiel publiceren.

Privacy Company heeft een selectie gemaakt van alle variabelen met een score van 2. Het Kohnstamm instituut heeft vervolgens een afweging kunnen maken tussen de waarde van het opnemen van de variabelen voor onderzoeksdoeleinden en hoeveel de variabele bijdraagt aan het herleidbaarheidsrisico op basis van de entropie.

Let wel, de gegevens in de opgeschoonde dataset zijn niet anoniem. De gegevens in de dataset zijn pseudonieme persoonsgegevens, waarbij de kans op mogelijke herleidbaarheid is geminimaliseerd. Het risico op herleidbaarheid door bijvoorbeeld directe collega's van respondenten is niet geheel weggenomen, maar herleidbaarheid door individuen buiten de school is volgens Privacy Company in hoge mate onwaarschijnlijk.

7.5 Welke juridische criteria zijn van toepassing om te bepalen of toetsscores als persoonsgegeven moeten worden behandeld?

Een persoonsgegeven in de zin van de AVG is alle informatie die een natuurlijk persoon identificeert of identificeerbaar maakt (artikel 4, lid 1 AVG). Om te bepalen of een natuurlijk persoon identificeerbaar is (en of er dus sprake is van persoonsgegevens), wordt gekeken naar alle middelen die de verwerkingsverantwoordelijke of derden kunnen gebruiken om die persoon direct of indirect te identificeren. Bij de beoordeling of deze middelen redelijkerwijs kunnen worden gebruikt om een persoon te identificeren moeten objectieve factoren, zoals kosten, beschikbare technologie en technologische ontwikkeling in aanmerking worden genomen (overweging 26 AVG).

Anonieme gegevens zijn geen persoonsgegevens. Het gaat dan om gegevens die niet verwijzen naar een geïdentificeerde of identificeerbare natuurlijke persoon of om persoonsgegevens die zodanig zijn geanonimiseerd dat de betrokkene niet of niet meer identificeerbaar is. De AVG is niet van toepassing op anonieme gegevens. Als de bron datasets gewist zouden zijn (maar die blijven juist expres beschikbaar), zouden de rapporten van de OESO alleen statistisch veilige anonieme gegevens bevatten.

Gepseudonimiseerde persoonsgegevens zijn gegevens die door het gebruik van aanvullende gegevens aan een natuurlijk persoon kunnen worden gekoppeld (waarna die persoon kan worden geïdentificeerd). Gepseudonimiseerde gegevens zijn wel persoonsgegevens (overweging 26 AVG). De AVG is om die reden van toepassing.

De EDPB legt uit dat pseudonisering (artikel 4, lid 5 AVG) de koppelbaarheid beperkt van een dataset aan de oorspronkelijke identiteit van een deelnemer. Bij het beoordelen van de effectiviteit van pseudonisering moet de verwerkingsverantwoordelijke (of verwerker) ook rekening houden met aanvullende informatie waar zij zelf geen directe controle over heeft, zolang het redelijkerwijs te verwachten is dat die informatie beschikbaar is. Zo kan informatie uit openbaar toegankelijke bronnen, zoals berichten op sociale media of openbare registers, bijdragen

aan de heridentificatie van gepseudonimiseerde gegevens aan betrokkenen.¹¹ Privacy Company heeft hier in dit advies uitdrukkelijk rekening mee gehouden.

7.6 Kunnen jullie de opvatting van het CBS volgen dat toetsscores sowieso als gevoelige gegevens zijn aangemerkt?

Ja, die opvatting kan Privacy Company volgen wanneer het regime uit de Wet op het Centraal bureau voor de statistiek (Wet CBS) van toepassing is op de toetsscores.

Privacy Company heeft gekeken naar de praktische beschikbaarheid van koppelingsinformatie om deelnemers te kunnen (her)identificeren in de zin van de AVG. Er zijn geen andere bronnen, anders dan de antwoorden die de deelnemer op de tablet heeft ingevuld (PIAAC) of antwoorden van leerkrachten en bestuurders op de vragen uit de enquête (TALIS), die de subjectieve toetsscores of subjectieve meningen of ervaringen over de leeromgeving bevatten. Na het onderdrukken van alle variabelen die alsnog tot identificatie van een deelnemer kunnen leiden (en daarmee, van herleidbaarheid van zijn of haar toetsscores), acht Privacy Company het in hoge mate onwaarschijnlijk dat iemand de toetsscores kan herleiden naar een specifieke deelnemer.

7.6.1 Het verstrekken en openbaar maken van gegevens op grond van de Wet CBS

Het CBS is een zelfstandig bestuursorgaan met een eigen wettelijk kader, waaronder art. 37, 41 en 42 Wet op het Centraal bureau voor de statistiek (Wet CBS).

Op grond van art. 37 Wet CBS mag het CBS gegevens die zij heeft ontvangen in het kader van de uitoefening van CBS-taken alleen verstrekken aan anderen die belast zijn met de uitvoering van de taak van het CBS (tweede lid) en slechts zodanig openbaar maken dat daaraan geen 'herkenbare gegevens over een afzonderlijk persoon' kunnen worden ontleend (derde lid).¹² Uit de Memorie van Toelichting bij dit artikel blijkt dat bewust voor het begrip 'herkenbare gegevens' is gekozen ten opzichte van het striktere 'herleidbaarheid'.¹³

Art. 41 Wet CBS bevat een uitzondering op art. 37(2) Wet CBS en maakt het mogelijk dat het CBS op verzoek een verzameling van gegevens verstrekt of toegang verleent aan een in de wet genoemde dienst, organisatie of instelling. Instellingen die niet expliciet in de wet worden genoemd kunnen op grond van de 'Beleidsregel toegang instellingen tot microdata CBS' onder voorwaarden een machtiging krijgen.¹⁴ Dit verstrekken of toegang verlenen mag alleen ten behoeve van statistisch of wetenschappelijk onderzoek. Met betrekking tot het gebruik van deze gegevens moeten passende maatregelen zijn genomen om de herkenning van afzonderlijke personen¹⁵ te

¹¹ EDPB, Guidelines 01/2025 on Pseudonymisation, Adopted on 16 January 2025, nr. 21, URL: https://www.edpb.europa.eu/system/files/2025-01/edpb_guidelines_202501_pseudonymisation_en.pdf.

¹² Dit herkenbaarheids criterium geldt ook voor een afzonderlijk huishouden, onderneming of instelling. Er bestaat een uitzondering voor de laatste twee categorieën als er een gegronde reden is om aan te nemen dat bij de betreffende onderneming of instelling geen bedenkingen tegen openbaarmaking bestaan. Voor personen en huishoudens is het verbod absoluut wat betreft openbaarmaking.

¹³ Kamerstukken II, 28 277, nr. 3 ([Memorie van Toelichting bij de Wet CBS](#)), p. 36.

¹⁴ 'Beleidsregel toegang instellingen tot microdata CBS' (Stort. 2021 36083).

¹⁵ En huishoudens, ondernemingen of instellingen. De 'Beleidsregel publiceren op instellingsniveau ex art. 41 Wet CBS' (Stort. 2023 12517) beschrijft de voorwaarden waaronder een onderzoeksinstantie op het niveau van een afzonderlijke onderneming of instelling mag publiceren. Daarbij gaat het ex art. 1(1) om onderzoek 'op de gegevens van het CBS op basis van artikel 41 van de Wet CBS'.

voorkomen. In dit artikel wordt opnieuw het begrip ‘herkenning’ gebruikt, wat in het licht van artikel 37 Wet CBS en de bijbehorende toelichting moet worden geïnterpreteerd. De Memorie van Toelichting vermeldt dat het CBS ruimte heeft om de mate van beveiliging van bestanden en de passende maatregelen te bepalen.¹⁶

De ‘Beleidsregel publiceren op instellingsniveau ex art. 41 Wet CBS’¹⁷ beschrijft de voorwaarden waaronder een onderzoeksinstituting vervolgens ook op het niveau van een afzonderlijke onderneming of instelling mag publiceren. Het doel van deze beleidsregel is om *“eenduidig beleid op [te] stellen ten aanzien van het publiceren op instellingsniveau en tegelijkertijd de positie van onderzoeksinstitutingen waar mogelijk gelijk trekken met die van het CBS. Hiermee vergroot het CBS de mogelijkheden van andere aanbieders van statistische diensten om gebruik te kunnen maken van de data waarover het CBS beschikt.”*¹⁸ Een publicatie is ex art. 2(1) van de beleidsregel in ieder geval niet toegestaan als publicatie leidt tot onthulling op persoons- of huishoudensniveau, net als voor het CBS geldt op grond van art. 37(3) Wet CBS.

Het CBS willicht op grond van art. 42 Wet CBS een verzoek om een verzameling van gegevens slechts in als de verzoeker naar het oordeel van het CBS voldoende maatregelen heeft getroffen om te voorkomen dat deze gegevens voor een ander doeleinde dan statistisch of wetenschappelijk onderzoek wordt gebruikt. Deze maatregelen worden vastgelegd in een contract. De Memorie van Toelichting benoemt enkele in dat contract op te nemen bedingen.¹⁹

In het ‘Handboek Statistische Beveiliging’²⁰ beschrijft het CBS de beveiligingsregels waaraan alle CBS-output in principe moet voldoen om gepubliceerd te kunnen worden. In dit handboek worden regels gesteld over statistische beveiliging, waarbij de grondregel is dat moet worden voorkomen ‘dat er conclusies over herkenbare eenheden kunnen worden getrokken op basis van CBS-output’²¹. Dit bepaalt dus de interpretatie van het herkenbaarheids criterium in de bovengenoemde artikelen uit de Wet CBS.

Dit regime gaat verder dan de vraag naar de herleidbaarheid van de identiteit van een natuurlijk persoon in de zin van de AVG en of de gegevens als persoonsgegevens kunnen (moeten) worden aangemerkt. De AVG is wat dat betreft genuanceerder dan het regime van de Wet CBS.

7.6.2 Toepasbaarheid op de PIAAC casus

Het CBS stelt in haar juridische reactie op dit advies, net als de de FG van het CBS, dat waar het PIAAC betreft getoetst moet worden aan het bovenstaande kader uit de Wet CBS. De FG stelt: *“omdat PIAAC op basis van een CBS-steekproef wordt uitgevoerd, moet het CBS de eigen regels ook opleggen aan derden”*. Het CBS geeft in de

¹⁶ Kamerstukken II, 28 277, nr. 3 ([Memorie van Toelichting bij de Wet CBS](#)), p. 38.

¹⁷ ‘Beleidsregel publiceren op instellingsniveau ex art. 41 Wet CBS’ (Stcrt. 2023 12517).

¹⁸ Ibid., p. 2.

¹⁹ Kamerstukken II, 28 277, nr. 3 ([Memorie van Toelichting bij de Wet CBS](#)), p. 38-39: *“onder meer het uitsluitend gebruik voor statistische of wetenschappelijk onderzoek, het verbod om de gegevens aan derden te verstrekken en het verbod om de gegevens op individueel niveau te koppelen”*.

²⁰ CBS, ‘Handboek Statistische Beveiliging: Info voor externen’ (2^e editie), juni 2024, [URL: //www.cbs.nl/-/media/cbs/over-ons/organisatie/handboekstatistischebeveiliging2024.pdf](https://www.cbs.nl/-/media/cbs/over-ons/organisatie/handboekstatistischebeveiliging2024.pdf).

²¹ Ibid., p. 7.

juridische reactie aan dat het contract en de algemene voorwaarden die van toepassing zijn bij het PIAAC onderzoek op verschillende plekken de toepasselijkheid van (eisen uit) de Wet CBS benoemen.²³

Wat betreft de verstrekking van de NAW-gegevens (plus leeftijd en geslacht) van de steekproef door het CBS is de Wet CBS in ieder geval van toepassing. Het CBS heeft aangegeven dat de verstrekking van de steekproefgegevens voor het PIAAC-onderzoek aan Kantar op grond van art. 37(2) Wet CBS heeft plaatsgevonden (dus niet op grond van het uitzonderingsregime van art. 41 en 42 Wet CBS). Art. 37(2) bepaalt dat de gegevens die het CBS ter uitvoering van de Wet CBS heeft ontvangen alleen mag verstrekken aan anderen die belast zijn met de uitvoering van de taak van het CBS, zie de voorgaande paragraaf. Om deze reden is het regime van de Wet CBS, specifiek de voorwaarden voor openbaarmaking en statistische veiligheid,²⁴ ook van toepassing op de publicatie van de aan de hand van de steekproef verzamelde en vervolgens opgeschoonde dataset.

Het CBS heeft de steekproef uitgevoerd en de NAW-gegevens van steekproefpersonen verstrekt, maar de gegevens zijn vervolgens verzameld door Kantar, en na pseudonimisering verder verwerkt door het Kohnstamm Instituut. Het gaat bij deze gegevens dus niet zonder meer over microdata van het CBS *waarop* onderzoek is uitgevoerd. Uit het projectvoorstel blijkt dat het CBS onder meer betrokken is geweest bij de steekproef, de weging en de non-responsanalyse.²⁵ Hoewel in de overeenkomst staat dat het CBS ook een analyse doet op de vragenlijst, begrijpt Privacy Company dat het CBS dit toch niet heeft gedaan, omdat het een internationaal gestandaardiseerde vragenlijst betrof.

Uit de Wet CBS en het overeengekomen projectvoorstel volgt volgens Privacy Company niet duidelijk dat het openbaarmakingsregime uit de Wet CBS ook van toepassing is op de publicatie van de resultaten van een onderzoek waarbij slechts gebruik is gemaakt van een steekproef van herkenbare personen en de weging van de representativiteit van de respondenten door het CBS. Privacy Company adviseert dan ook om in de toekomst duidelijke voorwaarden in de projectovereenkomst op te nemen over de publicatie van onderzoeksresultaten, met verwijzingen naar de relevant wettelijke kaders. Over het overeengekomen projectvoorstel merkt Privacy Company nog het volgende op.

7.6.3 Contractueel kader

Het contractuele kader bij het PIAAC-onderzoek bestaat uit de volgende documenten:

- Projectvoorstel: Werkzaamheden CBS in Programme for the International Assessment of Adult Competencies (PIAAC) van 26 november 2019 (projectvoorstel)
- Algemene voorwaarden aanvullend statistisch onderzoek van het CBS 1 oktober 2019 (algemene voorwaarden CBS)
- De Verwerkersovereenkomst tussen CBS en Kantar ten behoeve van het gebruik mogen maken van de gegevens van steekproefpersonen (verwerkersovereenkomst)

In deze documenten staan enkele passages die ingaan op de herleidbaarheid en geheimhouding in de PIAAC casus.

7.6.3.1 Projectvoorstel

²³ Dit is een contract tussen het NRO (sinds 1 januari NKO) en het CBS, 'Projectvoorstel: Werkzaamheden CBS in Programme for the International Assessment of Adult Competencies (PIAAC)', 26 november 2019.

²⁴ CBS, 'Handboek Statistische Beveiliging: Info voor externen' (2^e editie), juni 2024, [URL://www.cbs.nl/-/media/cbs/over-ons/organisatie/handboekstatistischebeveiliging2024.pdf](https://www.cbs.nl/-/media/cbs/over-ons/organisatie/handboekstatistischebeveiliging2024.pdf).

²⁵ Ibid, p. 5.

- *“Het CBS kan alleen als samenwerkingspartner aan onderzoek meedoen wanneer dit onderzoek voldoet aan de normen die het voor zijn eigen onderzoek stelt. Voor het PIAAC-onderzoek betekent dit dat het CBS een review uitvoert op de vragenlijst waarbij het kijkt of er geen vragen aan respondenten worden gesteld die een (te) hoge mate van gevoeligheid hebben.”²⁶*
- *“Publiceren. Van alle landen worden de uitkomsten van het PIAAC-onderzoek door OECD gepubliceerd in een Public Use File. In het kader van de privacy wetgeving en data beveiliging is het mogelijk dat landen sommige variabelen als “missing” invullen. Het CBS heeft het recht vanuit zijn publiekrechtelijke taak (het openbaar maken van op grond van overheidswege verricht statistisch onderzoek) na oplevering het PIAAC-databestand aan te wenden voor eigen publicaties. Het CBS deelt in dat geval in een vroegtijdig stadium maar minimaal twee weken van te voren aan OCW mee welke informatie het CBS beoogt te publiceren en zorgt ervoor dat een correcte bronvermelding wordt toegepast bij alle publicaties en media-uitingen van het CBS. OCW, Kohnstamm Instituut en NRO stellen CBS minimaal 2 weken van te voren op de hoogte van eventuele geplande publicitaire uitingen die (mede) op basis van het PIAAC-onderzoek zijn opgesteld.”²⁷*
- *“Alle door het CBS verstrekte gegevens van steekproefpersonen mogen door het uitvoerende veldwerkbureau uitsluitend voor de gegevensverzameling ten behoeve van PIAAC worden gebruikt en dienen na afloop te worden vernietigd.”²⁸*

Het Projectvoorstel bevat een bepaling over een voorwaarde voor deelname vanuit het CBS, namelijk dat het onderzoek moet voldoen aan de voorwaarden die het CBS stelt aan haar onderzoek. Daaronder kan het regime uit de Wet CBS worden begrepen, zoals hiervoor beschreven, al is dit niet expliciet zo geformuleerd.

Wel bevat het Projectvoorstel een kopie over publicatie, met daarbij een bepaling over de publicatie van uitkomsten van het PIAAC-onderzoek in een Public Use File (PUF), waarbij het mogelijk is dat landen *sommige* variabelen invullen als “missing”. Het projectvoorstel bevat geen expliciete verwijzing naar het regime van openbaarmaking en verstrekking uit de Wet CBS en maakt dus niet expliciet duidelijk dat dit regime van toepassing is.

7.6.3.2 Algemene voorwaarden

- *“Microdata: data ten behoeve van statistisch of wetenschappelijk onderzoek door het CBS vervaardigd met betrekking waartoe passende maatregelen zijn genomen om herkenning van afzonderlijke personen, huishoudens, ondernemingen of instellingen te voorkomen, zoals bedoeld in artikel 41 van de Wet op het Centraal Bureau voor de Statistiek, nader te noemen; “CBS-wet”.²⁹*
- *“Partijen kunnen in de Overeenkomst van deze Algemene voorwaarden afwijken. In geval van strijd tussen de Overeenkomst en de Algemene voorwaarden prevaleert de Overeenkomst.”³⁰*
- *“Het CBS verzamelt gegevens om statistieken te maken en het CBS maakt gebruik van deze gegevens/statistieken om aanvullend statistisch onderzoek uit te voeren. Het CBS beschikt daarom over*

²⁶ Projectvoorstel, p. 5. Zoals hierboven toegelicht, heeft het CBS deze beoordeling niet gedaan.

²⁷ Ibid.

²⁸ Art. 4 Projectvoorstel.

²⁹ Art. 1.6 algemene voorwaarden CBS. Dit artikel wordt ook geciteerd in de juridische reactie van het CBS. In de algemene voorwaarden komt dit begrip alleen terug in art. 10.6 en 10.7. Daarbij gaat het om de toegang tot door het CBS samengestelde of verrijkte Microdata. Dit mag uitsluitend in een beveiligde netwerkomgeving van het CBS. In het geval van remote access of on site moet een apart contract worden afgesloten. Van beide is in de PIAAC-casus geen sprake, nu het om de verstrekking van NAW-gegevens van steekproefpersonen gaat.

³⁰ Art. 3.3 algemene voorwaarden CBS.

informatie van burgers, bedrijven en organisaties. De juridische kaders waarbinnen het CBS met privacygevoelige informatie om mag gaan zijn strikt geregeld in de CBS-wet en de Algemene Verordening gegevensbescherming (AVG). Aanvullend heeft het CBS een eigen gedragscode. Het CBS heeft een functionaris voor de gegevensbescherming. Deze persoon waakt over de naleving door het CBS van de wettelijke bepalingen inzake de privacybescherming en houdt daarnaast een meldingenregister bij van de verwerkingen van persoonsgegevens.”³¹

- *“Uit de verstrekte onderzoeksresultaten mag geen informatie kunnen worden herleid over individuele personen, bedrijven of organisaties. Dit betekent dat de resultaten kunnen worden afgerond en/of geheimgehouden.”³²*

De algemene voorwaarden zien op het uitvoeren van aanvullend statistisch onderzoek door het CBS. Daaruit blijkt niet zonder meer dat ‘de verstrekte onderzoeksresultaten’ in de zin van art. 10.5 van de algemene voorwaarden CBS ziet op de publicatie van de (opgeschoonde) resultaten van het PIAAC-onderzoek. Gezien het Projectvoorstel lijkt het eerder te gaan om de verstrekking van de NAW-gegevens van steekproefpersonen, al heeft dit plaatsgevonden op grond van art. 37(2) Wet CBS. Privacy Company begrijpt in ieder geval dat Kantar deze steekproefgegevens op grond van art. 4.2 Projectvoorstel heeft vernietigd.

7.6.3.3 Verwerkersovereenkomst

Privacy Company heeft de verwerkersovereenkomst tussen het CBS en Kantar niet in kunnen zien. De verwerkersovereenkomst zou als onderdeel van het contractuele kader bepalingen kunnen bevatten over de herleidbaarheid en geheimhouding van de onderzoeksresultaten in de PIAAC casus.

Het CBS heeft laten weten dat in de verwerkersovereenkomst niets staat over de Wet CBS, dat Kantar slechts ging over de verzameling van de respons (en niet over de publicatie) en dat de verwerkersovereenkomst alleen de volgende verwerking regelt: *“Verwerker gebruikt de van het Verantwoordelijke gegevens uitsluitend voor de communicatie met potentiële respondenten voor PIAAC 2020/2021/2022”*.³³

7.6.4 Overige opmerkingen en conclusie

Voor zover het CBS bezwaar heeft gemaakt dat een openbare publicatie van een dataset die het gevolg is van wetenschappelijk onderzoek nooit onder de noemer ‘wetenschappelijk onderzoek’ zou kunnen vallen, citeert Privacy Company de EDPB dat de definitie van wetenschappelijk onderzoek onder de AVG ruim moet worden opgevat:

“De definitie van wetenschappelijke onderzoeksdoeleinden heeft aanzienlijke gevolgen voor het bereik van verwerkingsactiviteiten die een verwerkingsverantwoordelijke mag uitvoeren. Het begrip “wetenschappelijk onderzoek” wordt in de AVG niet gedefinieerd. Overweging 159 stelt: “(...) Voor de toepassing van deze verordening moet de verwerking van persoonsgegevens met het oog op wetenschappelijk onderzoek ruim worden opgevat. (...)”; de EDPB is echter van mening dat het begrip niet ruimer mag worden opgevat dan de gebruikelijke betekenis en verstaat onder “wetenschappelijk onderzoek” in deze context een onderzoeksproject dat opgezet wordt volgens de relevante methodologische en ethische normen van de sector, en conform goede praktijken.”³⁴

³¹ Art. 5.5 algemene voorwaarden CBS.

³² Art. 10.5 algemene voorwaarden CBS.

³³ CBS, reactie op verzoek om inzage verwerkingsovereenkomst (via OCW), 9 februari 2026.

³⁴ EDPB Richtsnoeren 5/2019 betreffende de criteria voor het recht om vergeten te worden in de zoekmachinezaken krachtens de AVG (deel 1), versie 2.0, 7 juli 2020, par. 152, URL: https://www.edpb.europa.eu/our-work-tools/our-documents/guidelines/guidelines-52019-criteria-right-be-forgotten-search-engines_nl.

Het is aannemelijk dat OESO, via de commissie met expert-onderzoekers uit de ruim 30 deelnemende landen, de juiste methodologische en ethische normen heeft gehanteerd bij het opstellen van beide types onderzoek.³⁵ Er is alleen een dispuut over de veiligheid van de openbare dataset met vragen en antwoorden. Privacy Company meent dat er na onderdrukking van de variabelen met een koppelbaarheidsscore van 2 geen of alleen een theoretisch risico is op heridentificatie van de deelnemers. Dat is voldoende onder de AVG, maar onvoldoende onder het regime van statistische veiligheid op grond van de Wet CBS.

Er wordt internationaal grote waarde gehecht aan de publicatie van (opgeschoonde) brondata van internationaal vergelijkend onderzoek, vanuit transparantie en om secundair onderzoek te kunnen doen. De AVG staat daaraan niet in de weg, mits het ministerie de variabelen onderdrukt met een koppelbaarheidsscore van 2 en die voldoende informatie bevatten om samen een betrokkene uniek te herleiden. Privacy Company adviseert om in de toekomst bij betrokkenheid van derde partijen duidelijke voorwaarden in de projectovereenkomst op te nemen over de publicatie van onderzoeksresultaten, met verwijzingen naar de relevant wettelijke kaders

Maar Privacy Company ziet wel een algemeen risico in publicatie van de antwoorden in een open dataset, namelijk dat de (vrijwel niet herleidbare) gegevens worden hergebruikt voor doelen die niet voorspelbaar of wenselijk zijn. In dat licht adviseert Privacy Company het ministerie om aan te dringen bij OESO op een procedure van Restricted Use Files (RUF in plaats van PUF), waarbij onderzoekers kunnen motiveren waarom zij toegang tot de dataset willen, en waarin zij tekenen dat zij de set alleen voor verantwoorde onderzoeksdoelen gebruiken. Dat betekent niet dat de toegang beperkt zou moeten zijn tot onderzoekers die aan hogescholen of universiteiten zijn verbonden, in het licht van de ruime opvatting in de AVG van de term 'onderzoek'.

7.7 Zijn alle toetsscores herleidbare persoonsgegevens?

Privacy Company kan niet alle combinaties van variabelen doorrekenen om vast te stellen of een specifieke toetsscore herleidbaar is. Daarom heeft Privacy Company het risicomijdende uitgangspunt gehanteerd dat zodra één variabele herleidbaar is, de volledige dataset als herleidbaar geldt. Vanuit dat uitgangspunt heeft het Kohnstamm instituut de variabelen onderdrukt in TALIS met een score op koppelbaarheid van 2 en een hoge entropie.

7.8 Is er volgens de wet geen sprake van een persoonsgegeven als identificatie van een individu een onevenredige inspanning vereist en in de praktijk ondoenlijk is? Is dat hier van toepassing? Wat betekent de overweging 26 uit de AVG voor deze casus?

Ja, volgens de recente uitspraak van het Hof van Justitie is dit het geval. Dit wordt ook wel de *relatieve leer* genoemd.

In het privacyrecht zijn er twee mogelijkheden om over persoonsgegevens na te denken: vanuit de absolute, of vanuit de relatieve leer. De absolute leer houdt in dat als een persoonsgegeven door iemand herleidbaar is, dat het dan altijd een persoonsgegeven blijft, ook al kan de ontvanger de gegevens niet meer zelf herleiden. De relatieve leer houdt in dat een persoonsgegeven alleen een persoonsgegeven is als de ontvanger de gegevens zelf nog kan herleiden. Dat wil zeggen dat de kwalificatie afhankelijk zou zijn van het perspectief van de verwerker. Als er een niet-onwettige mogelijkheid bestaat om een gegeven aan een persoon te koppelen, moet een organisatie ervan

³⁵ Tegelijkertijd moet het ministerie zich inspannen om zelf en bij de OESO controles uit te voeren op zowel de ethische wenselijkheid van vragen, als op het gebied van entropie en herleidbaarheid.

uitgaan dat zij persoonsgegevens verwerkt, ook al kan zij zelf de gegevens niet aan specifieke personen koppelen.³⁶ De AVG bevat juist een definitie van pseudonieme persoonsgegevens om aan te geven dat er veel mogelijkheden zijn binnen de AVG om met pseudonieme persoonsgegevens te werken.

Privacy Company acht het, op de voet van de argumenten van de EDPB en de EDPS bij het Europees Hof van Justitie, onwenselijk om uit te gaan van een relatieve leer bij de vraag of een pseudoniem gegeven een persoonsgegeven is, maar volgt de redenering van het Europees Hof van Justitie in het recente arrest EDPS vs. Single Resolution Board (SRB) dat het wel degelijk mogelijk is dat pseudonieme persoonsgegevens géén persoonsgegevens meer zijn als de ontvanger ze niet meer kan herleiden.

In de zaak van de EDPS tegen het Single Resolution Board (SRB) over de verwerking van pseudonieme (interview) gegevens door Deloitte adviseerde de advocaat-generaal (a-g) dat de EDPS had moeten bewijzen dat de gepseudonimiseerde, geaggregeerde gegevens nog herleidbaar waren voor Deloitte, ondanks het voortbestaan van de set met identificerende gegevens.³⁷ Het Hof heeft dit advies overgenomen.

De a-g gaf in zijn advies tegelijkertijd wel een strenge uitleg aan het begrip onevenredige inspanning uit overweging 26, althans, de gelijklopende overweging 16 uit de verordening gegevensbescherming voor Europese instellingen, (EU) 2018/1725.³⁸

Het moet voor de derde partij/ontvanger echt onmogelijk zijn om de betrokken personen te identificeren uit de pseudonieme dataset, ook al bestaat er nog een dataset met identificerende gegevens, het gevaar voor identificatie moet onbestaand of onbeduidend zijn wil er geen sprake meer zijn van (gepseudonimiseerde) persoonsgegevens. Er is dus niet snel sprake van een onevenredige inspanning.

De a-g schreef:

³⁶ Zie de jurisprudentie van het Europees Hof van Justitie over identificeerbaarheid, zoals samengevat in het advies van a-g Spielmann van 6 februari 2025 in voetnoot 25 (voetnoot hieronder). “For example, in the judgment in **Breyer**, a dynamic IP address in the hands of an online media services provider, although dissociated from the identification data held by the internet service provider, is classified as ‘personal data’, since the online media services provider had the means which may likely reasonably be used in order to identify the data subject of that IP address. Similarly, the judgment of 9 November 2023, **Gesamtverband Autoteile-Handel** (Access to vehicle information) (C-319/22, EU:C:2023:837), concerned the VIN of a vehicle, defined as the alphanumeric code assigned to a vehicle by its manufacturer in order to ensure proper identification of that vehicle. Although it is not in itself personal data, that VIN becomes personal as regards someone who reasonably has means enabling that datum to be associated with a specific person (paragraph 46) and thus to be linked to an identified or identifiable natural person (paragraph 49). Again in the same vein, a press release of the European Anti-Fraud Office (OLAF) containing identifiers allowing the appellant to be identified, either on the basis of a simple, objective reading of that press release or by means ‘reasonably likely to be used’ by one of its readers, constitutes personal data (judgment of 7 March 2024, **OC v Commission**, C-479/22 P, EU:C:2024:215). Similarly, in the judgment of 7 March 2024, **IAB Europe** (C-604/22, EU:C:2024:214), the association IAB, which represents undertakings in the digital advertising and marketing sector, had set up a framework for recording the preferences of website users, encoded by means of a ‘TC String’ (a combination of letters and characters). That TC String was regarded as personal data since, when associated with an identifier, it allows the internet user concerned to be identified, including for IAB, which did not hold the identifying data but could indirectly have access to it by reasonable means (paragraphs 48 to 50 of the judgment).”

³⁷ EHvJ, advies a-g Spielmann in zaak C 413/23 P, European Data Protection Supervisor v Single Resolution Board, 6 februari 2025, ECLI:EU:C:2025:59.

³⁸ Regulation (EU) 2018/1725 of the European Parliament and of the Council of 23 October 2018 on the protection of natural persons with regard to the processing of personal data by the Union institutions, bodies, offices and agencies and on the free movement of such data, and repealing Regulation (EC) No 45/2001 and Decision No 1247/2002/EC (Text with EEA relevance.), URL: <https://eur-lex.europa.eu/eli/reg/2018/1725/oj/eng>.

“51. Uit de bewoordingen van deze bepalingen leid ik af dat pseudonimisering de mogelijkheid openlaat dat de betrokkenen niet identificeerbaar zijn, anders zouden de bewoordingen van deze overweging 16 geen bestaansrecht hebben. Ik voeg hieraan toe dat de laatste volzinnen van deze overweging over anonimisering deze uitlegging bevestigen: zij sluiten geanonimiseerde (of anoniem gemaakte) gegevens uit van de werkingssfeer van verordening 2018/1725, maar sluiten gepseudonimiseerde gegevens alleen ervan uit voor zover de betrokkenen niet identificeerbaar zijn. Als het onmogelijk is om de betrokken personen te identificeren, worden zij derhalve wettelijk geacht voldoende beschermd te zijn door het pseudonimiseringsproces, ondanks het feit dat de aanvullende identificatiegegevens niet volledig zijn gewist.

52. Met andere woorden, er is geen sprake van dat gepseudonimiseerde gegevens automatisch uit de werkingssfeer van deze verordening worden gehaald. In het licht van overweging 16 ervan kan echter niet worden uitgesloten dat dergelijke gegevens onder bepaalde voorwaarden niet onder het begrip „persoonsgegevens” vallen.[onderstrepingen toegevoegd door Privacy Company].”

En de a-g concludeerde: “Alleen als het gevaar voor identificatie onbestaand of onbeduidend is, kunnen gegevens dus wettelijk ontsnappen aan de kwalificatie als „persoonsgegevens [onderstreping toegevoegd door Privacy Company].” Daarbij verwijst de a-g in een voetnoot naar het Breyer-arrest: “Uit punt 46 van het arrest Breyer volgt dat dit het geval is indien de identificatie van de betrokkene bij de wet verboden wordt of in de praktijk ondoenlijk is, bijvoorbeeld omdat zij – gelet op de vereiste tijd, kosten en mankracht – een excessieve inspanning vergt.”³⁹

In het arrest volgt het Hof de relatieve leer, haalt expliciet paragraaf 51 van de a-g aan⁴⁰ en stelt verder: “Hieruit volgt dat, anders dan de EDPS stelt, gepseudonimiseerde gegevens niet in alle gevallen en voor eenieder mogen worden beschouwd als persoonsgegevens voor de toepassing van verordening 2018/1725, aangezien pseudonimisering, afhankelijk van de omstandigheden van het geval, daadwerkelijk kan verhinderen dat andere personen dan de verwerkingsverantwoordelijke de betrokkene op zodanige wijze identificeren dat deze betrokkene voor hen niet of niet meer identificeerbaar is.”⁴¹

7.9 Gegeven de voorgaande uitkomsten: Kunnen de uitkomsten van beide onderzoeken in de huidige vorm conform AVG worden gepubliceerd?

Nee. De uitkomsten van beide onderzoeken kunnen in de huidige vorm (zonder onderdrukken van set variabelen met een hoge entropie en koppelbaarheidsniveau 2) niet conform de AVG worden gepubliceerd. In [Tabel 3](#) hieronder staan een paar voorbeelden van gegevens die een schooldirecteur bijvoorbeeld op LinkedIn zou kunnen publiceren, en waarmee die directeur dus potentieel herleidbaar wordt in de dataset. Om die reden heeft het Kohnstamm Instituut de variabelen met een hoge entropie onderdrukt in de TALIS set, en zal zij dit doen voor de PIAAC-set.

Tabel 2: Voorbeelden van datavelden met een koppelbaarheidsniveau 2

Label	Naam	Entropie	Koppelbaarheidsniveau
Personal Background		0	
Gender	TC4G01	0.91053306	2
Age	TC4G02	4.906317583	2

³⁹ HvJ, advies a-g Spielmann in zaak C 413/23 P, par. 57 en voetnoot 26.

⁴⁰ HvJ, uitspraak van 4 september 2025, EDPS v SRB, C 413/23 P, ECLI:EU:C:2025:645, para. 80

⁴¹ Ibid., para. 86.

Highest level of formal education completed	TC4G03	1.139285534	2
Years of work experience; regardless of full-time or part-time		0	
As a principal at this school	TC4G04A	3.667422333	2
As a principal in total	TC4G04B	4.33003027	2
In other school management roles (not principal)	TC4G04C	3.649321792	2
As a teacher in total	TC4G04D	4.78675565	2
In other non education jobs	TC4G04E	2.40632579	2

7.10 Indien nee, welke aanpassingen zijn minimaal noodzakelijk om publicatie mogelijk te maken, met behoud van de grootst mogelijke wetenschappelijke waarde?

Het ministerie kan de uitkomsten van het TALIS onderzoek laten publiceren door de OESO als PUF, indien het ministerie het gedane voorstel volgt over de te onderdrukken set van variabelen. Dit is expliciet afgestemd tussen Privacy Company en het Kohnstamm Instituut. Zoals toegelicht in paragraaf 7.6 verzet de Wet CBS zich tegen publicatie van PIAAC als PUF, zelfs na het volgen van het advies over de te onderdrukken set van variabelen.

Een tweede noodzakelijke aanpassing onder de AVG betreft de grondslag. De grondslag voor de huidige publicaties (in ieder geval bij PIAAC in verband met de onjuiste informatie aan betrokkenen over 'anonieme gegevens') blijft toestemming, maar dan met een nadere toets van de verenigbaarheid van de verdere verwerking in de vorm van publicatie van nauwelijks herleidbare opgeschoonde brondata.

In de toekomst zou het ministerie met een andere grondslag moeten werken, namelijk publicatie die noodzakelijk is voor vervulling van een taak van algemeen belang (artikel 6(1) e AVG). Beide grondslagen worden in antwoord op [vraag 13](#) apart toegelicht.

7.11 Welke maatregelen, in lijn met artikel 89 AVG, kunnen worden genomen om voor toekomstige onderzoeken hergebruik voor wetenschappelijk, historisch of statistisch onderzoek mogelijk te maken?

Artikel 89 AVG bepaalt dat de verwerking van persoonsgegevens met het oog op archivering in het algemeen belang, wetenschappelijk of historisch onderzoek of statistische doeleinden, onderworpen dient te zijn aan passende waarborgen voor de rechten en vrijheden van de betrokkenen. In ieder geval moet het ministerie (samen met het Kohnstamm Instituut) technische en organisatorische maatregelen treffen om dataminimalisatie te garanderen. Als het doel van het onderzoek kan worden bereikt met gepseudonimiseerde persoonsgegevens, dan moeten persoonsgegevens worden gepseudonimiseerd (overweging 156 AVG).

Om problemen met publicatie van datasets te voorkomen, adviseert Privacy Company het ministerie zeven maatregelen te treffen:

1. Dring bij de OESO aan om van de PUF een RUF te maken, een dataset met beperkte toegang voor onderzoeksdoeleinden, in ieder geval als het gaat om gegevens van minderjarigen (PISA).
2. Geef de (opgeschoonde) brondata pas vrij voor openbare publicatie nadat eventuele koppeltabellen naar ID's van identificeerbare deelnemers zijn vernietigd. Voorkom sowieso het bijhouden van koppeltabellen als het enige doel is om mensen hun toestemming in te kunnen laten trekken. Werk in plaats daarvan met de grondslag van noodzaak voor vervulling van een taak van algemeen belang.

3. Informeer de deelnemers adequaat, op een voor hun leeftijd en begrip begrijpelijk niveau, over wat het doel is van de vragenlijsten. Het ministerie moet uitleggen dat deelname vrijwillig is, maar dat de verwerking gebeurt in het publiek belang (van Nederland en van Europa), en dus niet op basis van toestemming. Het ministerie moet ook uitleggen dat de resultaten ontdaan zijn van alle gegevens die tot herleidbaarheid kunnen leiden, maar dat dat niet hetzelfde is als anoniem. Privacy Company adviseert om een flyer te maken met verwijzing naar de website van de OESO met de datasets en de onderzoeksresultaten zodat de deelnemers ook na afloop van het onderzoek nog eens kunnen opzoeken wat er met hun antwoorden is gebeurd.
4. Dring bij OESO aan op een extra controle of de onderzoeken geen vragen bevatten naar bijzondere persoonsgegevens of gegevens van strafrechtelijke aard, of vragen waaruit indirect dit soort bijzondere karakteristieken kunnen worden afgeleid.
5. Voer bij vragen aan minderjarigen (in casu het PISA onderzoek onder 15-jarigen) een extra ethische toetsing uit of er geen onbetamelijke vragen aan kinderen worden gesteld, of de antwoorden daadwerkelijk noodzakelijk zijn voor het behalen van legitieme onderzoeksdoelen. Daarbij moet het ministerie zich rekenschap geven van het feit dat het begrip 'noodzakelijk' een autonoom begrip is van het Unierecht. Concreet betekent dat controleren of het wel proportioneel is om de persoonsgegevens te verzamelen, in verhouding tot de inbreuk die de vragen maken op het privéleven van de betrokkenen. Een vraag stellen aan 15-jarigen of ze wel eens zwanger zijn geweest, is onbehoorlijk en disproportioneel.⁴² Dat soort vertrouwelijke en intieme informatie hoort alleen onder medisch beroepsgeheim te worden gedeeld. Denk bijvoorbeeld ook aan de ophef over de ruim duizend vragen die consultatiebureaus kunnen stellen voor het kinddossier, waaronder de vraag naar schaamhaar.⁴³
6. Dring bij de OESO aan om de vragenlijsten (nadat die door de commissie met onderzoek-experts zijn opgesteld vanuit inhoudelijke overwegingen) te laten controleren op entropie en koppelbaarheid. Vermijd vragen die naar verwachting antwoorden met een hoge entropie en koppelbaarheid opleveren. Vervang die vragen door vragen naar bandbreedtes (van bijvoorbeeld leeftijd of werkervaring).
7. Maak op begrijpelijke en goed toegankelijke wijze informatie openbaar over de toegepaste waarborgen om herleidbaarheid te voorkomen, om ervoor te zorgen dat deelnemers uit het verleden kunnen terugvinden wat er met hun gegevens is gebeurd. Dat kan bijvoorbeeld via de website van het ministerie, maar ook via de OESO, in een toelichting per land.

7.12 Zijn er omstandigheden waarbij onderzoeksgegevens, bestaande uit persoonsgegevens, beschikbaar kunnen worden gemaakt?

De term 'beschikbaar' is breed en kan verwijzen naar zowel openbare publicatie als gecontroleerde of beperkte toegang tot datasets. Het is onder de AVG verdedigbaar om de PIAAC en TALIS datasets na het verwijderen van variabelen met een koppelbaarheid van 2 te publiceren als open data, maar zoals toegelicht in antwoord op [vraag 6](#) en in punt 6 van [vraag 11](#) hierboven adviseert Privacy Company het ministerie om bij de OESO aan te dringen op publicatie als Restricted Use File (RUF). Bovendien staat de toepasselijkheid van de Wet CBS door de betrokkenheid van het CBS in de weg aan publicatie van de PIAAC datasets als open data, zoals beschreven in paragraaf 7.6.

In het algemeen is bij publicaties van brondata van belang om te kijken naar entropie en herleidbaarheidsniveaus van de variabelen om te waarborgen dat de pseudonimisering voldoende sterk is en het risico op heridentificatie van de betrokkene minimaal is, ook als er in de toekomst nieuwe (nu nog onbekende) datasets beschikbaar komen.

⁴² Privacy Company begrijpt uit reactie van het Ministerie van OCW dat de vraag over zwangerschap in PISA in Nederland nooit in de vragenlijst is opgenomen.

⁴³ Parool, Rouvoet: 'Kinddossier al veertig jaar hetzelfde', 27 december 2008, URL: <https://www.parool.nl/nieuws/rouvoet-kinddossier-al-veertig-jaar-hetzelfde>.

7.13 Is hergebruik voor wetenschappelijk, historisch of statistisch onderzoek wettelijk mogelijk wanneer de oorspronkelijke verwerking op toestemming van de betrokkene(n) berust?

Ja, dat is mogelijk, als de 'verdere' verwerking voor dit onderzoeksdoel evenredig is met het verzameldoel van de gegevens, conform artikel 6(4) van de AVG. Privacy Company heeft deze evenredigheidsanalyse uitgevoerd voor de aangeleverde TALIS en PIAAC datasets. Dat levert een positief resultaat op na het onderdrukken van variabelen met een hoge koppelbaarheid en een entropie hoger dan 0 (nul). De uitwerking van deze toets staat in [paragraaf 13.3](#) hieronder, maar eerst gaat Privacy Company in op toestemming ([paragraaf 13.1](#)) en algemeen belang ([paragraaf 13.2](#)).

7.13.1 Toestemming

'Toestemming' als grondslag voor gegevensverwerking en 'geïnformeerde toestemming' als onderdeel van ethische onderzoekspraktijken zijn twee verschillende concepten. Het ministerie heeft een privacy-analyse meegestuurd van de OESO over de publicatie van de PIAAC-resultaten.

De OESO licht dit onderscheid goed toe: (in de vertaling van Privacy Company): *"Goede ethische onderzoekspraktijken vereisen de geïnformeerde toestemming van de onderzoeksdeelnemers. De eis dat alle respondenten van PIAAC voorafgaand aan hun deelname aan het onderzoek toestemming moeten geven, maakt deel uit van de technische normen en richtsnoeren van PIAAC (norm 2.2 en richtsnoer 2.2C). Dit geldt ook voor respondenten van de zogenaamde huis-aan-huisinterviews. Deze eis geldt ongeacht de rechtsgrondslag voor het verzamelen van gegevens."*⁴⁴

De EDPB heeft deze visie in 2020 bevestigd in de richtlijnen over toestemming: *"Wanneer toestemming de rechtsgrond is voor het uitvoeren van onderzoek conform de AVG, moet deze toestemming voor het gebruik van persoonsgegevens worden onderscheiden van andere eisen inzake toestemming, die dienen als een ethische norm of een procedurele verplichting. Een voorbeeld van een dergelijke procedurele verplichting, waarbij verwerking niet gebaseerd is op toestemming maar op een andere rechtsgrond, is te vinden in de verordening inzake klinische proeven. In het kader van de gegevensbeschermingswetgeving kan deze vorm van toestemming worden beschouwd als een extra waarborg. Tegelijkertijd beperkt de AVG met betrekking tot de verwerking van gegevens voor onderzoeksdoeleinden de toepassing van artikel 6 niet tot alleen toestemming. Zolang passende waarborgen zijn ingesteld, zoals de eisen op grond van artikel 89, lid 1, en de verwerking behoorlijk, rechtmatig en transparant is en in overeenstemming met de normen voor gegevensminimalisering en individuele rechten, kunnen andere rechtsgronden, zoals artikel 6, lid 1, onder e) of f), worden gebruikt. Op grond van de uitzondering van artikel 9, lid 2, onder j), geldt dit ook voor bijzondere categorieën gegevens [onderstrepingen toegevoegd door Privacy Company]."*⁴⁵

Privacy Company weet niet of dit verschil tussen twee vormen van toestemming duidelijk was bij het uitvoeren van het onderzoek, en weet ook niet hoe de respondenten van TALIS en PIAAC precies zijn geïnformeerd. Privacy Company weet alleen dat bij het PIAAC onderzoek sprake was van een harde eis van het CBS om te informeren dat de onderzoeksresultaten anoniem zouden worden gepubliceerd. Dit zou verband kunnen houden met de eisen uit de Wet CBS, zoals besproken in paragraaf 7.6. Bovendien heeft het CBS een toestemmingseis opgelegd.

⁴⁴ OESO, DATA PROTECTION: PROGRESS REPORT, 22nd meeting of the PIAAC Board of Participating Countries, 8 April 2019, OECD Headquarters, Paris, COM/DELSA/EDU/PIAAC(2019)6 nr. 10.

⁴⁵ EDPB Richtsnoeren 5/2019 betreffende de criteria voor het recht om vergeten te worden in de zoekmachinezaken krachtens de AVG (deel 1), versie 2.0, 7 juli 2020, par. 153, URL: https://www.edpb.europa.eu/our-work-tools/our-documents/guidelines/guidelines-52019-criteria-right-be-forgotten-search-engines_nl

Afhankelijk van de verstrekte informatie en context zijn de (historische) onderzoeken dus ofwel uitgevoerd op de grondslag van toestemming, of hebben de onderzoekers alleen toestemming gevraagd om de ethische keuzevrijheid te borgen. In dat tweede geval zou het ministerie zich (nu al) kunnen beroepen op de grondslag van de noodzaak van de vervulling van een taak van algemeen belang (artikel 6(1) onder e AVG). Dat wordt hieronder toegelicht.

Voor geldige toestemming als grondslag gelden de volgende eisen: deze moet vrij, specifiek, geïnformeerd en ondubbelzinnig zijn (artikel 4(11) AVG). Betrokkenen moeten zo goed mogelijk worden geïnformeerd over de gegevensverwerking zodat zij verantwoord en bewust toestemming kunnen geven. De toestemming van de betrokkene moet specifiek zijn, dat wil zeggen dat de toestemming betrekking heeft op de verwerking en alle verwerkingsactiviteiten die hetzelfde doel of dezelfde doeleinden dienen.

De AVG stelt mildere eisen aan toestemming voor wetenschappelijke onderzoeksdoeleinden, maar dat helpt niet bij een openbare publicatie van datasets. De AVG legt uit dat het bij wetenschappelijk onderzoek vaak niet mogelijk is om bij het verzamelen van de persoonsgegevens het doel van de gegevensverwerking volledig te beschrijven. Daarom moeten deelnemers de mogelijkheid krijgen om hun toestemming te geven voor specifieke gebieden van het onderzoek waarbij erkende ethische normen voor wetenschappelijk onderzoek in acht worden genomen. De deelnemers moeten de kans krijgen om hun toestemming alleen te verlenen voor bepaalde onderzoeksterreinen of onderdelen van onderzoeksprojecten, zolang het beoogde doel dat toelaat.⁴⁶ Het ministerie heeft geen controle over de doelen waarvoor externe onderzoekers de publieke datasets hergebruiken, en kan betrokkenen daar dus ook niet over informeren.

De verwerkingsverantwoordelijke moet kunnen aantonen dat de betrokkene toestemming heeft gegeven voor de verwerking van zijn persoonsgegevens (artikel 7(1) AVG). Ook moet de betrokkenen zijn of haar toestemming te allen tijde kunnen intrekken (artikel 7(3) AVG). Bij PIAAC is intrekken niet mogelijk, omdat het onderzoeksbureau geen identificerende gegevens verzamelt, alleen een random ID van de deelnemer. Bij TALIS is gedurende 1 jaar sprake van een koppeltabel, waarmee de school een intrekkingverzoek van een docent of bestuurder zou kunnen doorgeven aan de beheerder van de dataset. Bij PISA wordt gedurende enige tijd een koppelingstabel bewaard tussen het random ID van de (minderjarige) deelnemer en zijn of haar echte identiteit, met als enige doel om (ouders van) de deelnemers hun toestemming in te kunnen laten trekken.

Door met de grondslag van toestemming te werken, ontstaat dus feitelijk een extra identificatiemogelijkheid. Dat brengt risico's met zich, zoals datalekken, terwijl het voor het doel van het onderzoek niet nodig is om de deelnemers na het beantwoorden van de vragen nog te kunnen identificeren.

Kortom, toestemming als grondslag voor wetenschappelijk onderzoek brengt allerlei praktische uitdagingen met zich.

7.13.2 Algemeen belang

De OESO legt uit dat het gebruik van de AVG-grondslag toestemming onwenselijk is, en dat de verwerking veel beter past onder de grondslag van een taak van algemeen belang.⁴⁷

⁴⁶ Overweging 33 AVG.

⁴⁷ OESO, DATA PROTECTION: PROGRESS REPORT, COM/DELSA/EDU/PIAAC(2019)6, par. 6.

Uit de rechtspraak van het HvJ EU volgt dat een taak van algemeen belang alleen niet volstaat als rechtsgrond. Voor rechtmatige gegevensverwerking onder artikel 6 lid 1 onder e AVG moet ook een wettelijke basis bestaan in Europees of nationaal recht.

De OESO schrijft: “*PIAAC is een onderzoek dat door de OESO-lidstaten is goedgekeurd als onderdeel van het werkprogramma en de begroting van de organisatie, dat wordt gefinancierd door nationale overheidsinstanties (bijvoorbeeld ministeries van Onderwijs en Arbeid) en namens de deelnemende landen wordt beheerd door het secretariaat van de OESO.*”⁴⁸

Nederland heeft in 1960 geholpen om de OESO op te richten. In het oprichtingsverdrag staan mooie woorden over onderzoek. Zoals in artikel 2b: “*(lidstaten zullen) in the scientific and technological field, promote the development of their resources, encourage research and promote vocational training.*” De ratificatie van het OESO-verdrag is vastgelegd bij wet.⁴⁹

De OESO neemt ook formele besluiten over de (uitvoering van de) PIAAC en TALIS onderzoeken.⁵⁰ Hoewel deelname aan de onderzoeken vrijwillig is, bevatten de OESO besluiten wel een opdracht over de financiering door de deelnemende landen. Nederland neemt altijd deel aan deze onderzoeken, en legt de begroting ook vast in de beleidsplannen en meerjarenbegrotingen van OCW.⁵¹

De grondslag van artikel 6 onder e van de AVG is niet hetzelfde als de grondslag van wettelijke verplichting. Er moet wel een wettelijke verplichting zijn op grond waarvan de overheidsorganisatie een taak van algemeen belang uitvoert, maar dat verband kan dun zijn. Dat blijkt bijvoorbeeld uit de Proctorio uitspraak van het Gerechtshof Amsterdam van juni 2021.⁵² Het feit dat een universiteit tentamens moet kunnen afnemen, ook tijdens Covid, is voldoende grond voor de inzet van een online tool om examens af te nemen. Het Hof schrijft: “*Niet in geschil is dat de software van Proctorio wordt ingezet ter vervulling van een taak in het kader van de uitoefening van het openbaar gezag dat aan de UvA is opgedragen, zoals bedoeld in artikel 6 lid 1 sub e AVG. Wel in geschil is of de gegevensverwerking die door het inzetten van Proctorio plaatsvindt noodzakelijk is voor het vervullen van die taak.*”⁵³

Daarom is de bij wet vastgelegde deelname aan het OESO verdrag, met de daarin opgenomen onderzoeksdoelen, in combinatie met de beleids- en financieringsintentie van het ministerie van OCW voor het onderzoek voldoende voor een grondslag onder artikel 6 lid e onder e AVG.

Privacy Company adviseert het ministerie om een beleidskader te ontwikkelen waarin dit soort onderzoeken gedaan worden (en ook vormgegeven worden) vanuit de noodzaak voor een helder publiek belang. Zoals hierboven toegelicht (in antwoord op [vraag 7.11](#), in het lijstje met de zeven maatregelen) moet het ministerie dan een balans vinden tussen het belang van publicatie en andere publieke belangen, zoals eerbiediging (door de overheid als wetgever) van het recht op privacy, zeker als het om minderjarigen gaat (PISA). Het ministerie moet bij de OESO aandringen op het voorkomen van vragen naar bijzondere persoonsgegevens, of vragen die misschien wel leuk, handig of nuttig zijn, maar te inbreukmakend zijn of een te groot risico opleveren op koppelbaarheid met externe bronnen.

⁴⁸ Ibid.

⁴⁹ <https://wetten.overheid.nl/BWBV0005057/1961-11-13>.

⁵⁰ Zie bijvoorbeeld <https://oecdgroups.oecd.org/Bodies/ShowBodyView.aspx?BodyID=7416&Lang=en&Book=True>

⁵¹ Zie bijvoorbeeld het jaarverslag OCW over 2023, URL:

<https://www.rijksoverheid.nl/binaries/rijksoverheid/documenten/jaarverslagen/2024/05/15/onderwijs-cultuur-en-wetenschap-2023/ocw-jaarverslag-2023.pdf>

⁵² <https://uitspraken.rechtspraak.nl/details?id=ECLI:NL:GHAMS:2021:1560>

⁵³ Idem, ro. 3.3.2. Het Hof schrijft per ongeluk ‘artikel 5 lid 1 sub e AVG’.

7.13.3 Verdere verwerking

Voor de huidige publicatie van PIAAC en (voor zover nodig) TALIS kan het ministerie onder de AVG een beroep doen op evenredigheid van de verdere verwerking.

In beginsel is het niet de bedoeling om gegevens voor een ander doel te verwerken dan waarvoor toestemming is verleend. Het idee achter toestemming is juist dat een organisatie de betrokkene van tevoren heel specifiek informeert waar hij de gegevens voor gaat gebruiken. Maar in de praktijk kan er altijd iets veranderen, en daar houdt de AVG expliciet rekening mee.

De EDPB legt uit: “Onverminderd de bepalingen inzake compatibiliteit van doeleinden moet toestemming specifiek voor het doel zijn. De betrokkenen verlenen hun toestemming in het besef dat zij controle kunnen uitoefenen en dat hun gegevens uitsluitend worden verwerkt voor die vermelde doeleinden. Indien een verwerkingsverantwoordelijke op basis van toestemming gegevens verwerkt, en de gegevens ook wil verwerken voor een ander doel, moet deze verwerkingsverantwoordelijke vragen om aanvullende toestemming voor dit andere doel, tenzij er een andere rechtsgrond is die beter bij de situatie past [onderstreping toegevoegd door Privacy Company].”⁵⁴

Aannemend dat respondenten van zowel PIAAC als TALIS de indruk hebben gekregen dat de verwerking op toestemming was gebaseerd, en in het geval van PIAAC, dat de antwoorden alleen anoniem zouden worden verwerkt, moet het ministerie afwegen of de verdere verwerking van de onderzoeksgegevens verenigbaar is met het oorspronkelijke verzameldoel. De grondslag blijft dan hetzelfde: toestemming. Daartoe moet het ministerie in ieder de vijf criteria aflopen uit artikel 6(4) van de AVG.

Artikel 6(4) van de AVG luidt als volgt:

“Wanneer de verwerking voor een **ander doel** dan dat waarvoor de persoonsgegevens zijn verzameld **niet berust op toestemming van de betrokkene** (...) houdt de verwerkingsverantwoordelijke bij de beoordeling van de vraag of de verwerking voor een ander doel verenigbaar is met het doel waarvoor de persoonsgegevens aanvankelijk zijn verzameld onder meer rekening met:

- a) ieder **verband** tussen de doeleinden waarvoor de persoonsgegevens zijn verzameld, en de doeleinden van de voorgenomen verdere verwerking;
- b) het kader waarin de persoonsgegevens zijn verzameld, met name wat de **verhouding** tussen de betrokkenen en de verwerkingsverantwoordelijke betreft;
- c) de **aard** van de persoonsgegevens, met name of bijzondere categorieën van persoonsgegevens worden verwerkt, overeenkomstig artikel 9, en of persoonsgegevens over strafrechtelijke veroordelingen en strafbare feiten worden verwerkt, overeenkomstig artikel 10;
- d) de mogelijke **gevolgen** van de voorgenomen verdere verwerking voor de betrokkenen;
- e) het bestaan van passende **waarborgen**, waaronder eventueel versleuteling of pseudonimisering.

Onder a (**verband**) houdt de openbare publicatie van de pseudonieme, van koppelbare gegevens ontdane datasets, rechtstreeks verband met het verzameldoel waarover de deelnemers zijn geïnformeerd, namelijk dat zij op anonieme wijze konden bijdragen aan een internationale vergelijking van onderwijsprestaties en vaardigheden. Noch het ministerie noch derde partijen kunnen de identiteit van deelnemers nog achterhalen na het verwijderen van de koppeltabellen op de deelnemende scholen. Het is ongelukkig dat in de informatie aan deelnemers het woord ‘anoniem’ is gebruikt, maar het is ook niet makkelijk om deelnemers het verschil tussen anoniem en vrijwel niet

⁵⁴ EDPB, Richtsnoeren 05/2020 inzake toestemming overeenkomstig Verordening 2016/679, Versie 1.1, vastgesteld op 4 mei 2020, URL: https://www.edpb.europa.eu/sites/default/files/files/file1/edpb_guidelines_202005_consent_nl.pdf.

herleidbaar pseudoniem uit te leggen. Doorgaans verstaan mensen onder het woord anoniem dat hun NAW-gegevens verwijderd zijn. Dat is ook feitelijk het geval en uit de dataset zijn nog aanzienlijk meer gegevens verwijderd die de betrokkene alsnog identificeerbaar zouden kunnen maken. Daarom sluit de verdere verwerking aan bij de gerechtvaardigde verwachtingen van betrokkenen.

Door de ingrepen in de datasets zijn de data in de PUF vrijwel niet meer te onderscheiden van anonieme gegevens. De kans op herleidbaarheid door een derde partij die de PUF downloadt, is - op dit moment - vrijwel nihil. Toch kiest Privacy Company voor de zekerheid voor de betiteling als 'pseudonieme' gegevens, om er zeker van te zijn dat alle mogelijke risico's voor betrokkenen goed in kaart zijn gebracht.

Onder b (**relatie**) zijn de vragen vrijwillig beantwoord door deelnemers, zowel bij de huis-enquêtes (PIAAC) als bij de vragenlijsten aan de leerkrachten en bestuurders (TALIS). Er is geen sprake van een ongelijke machtsverhouding tussen het ministerie of de OESO of het uitvoerende onderzoeksbureau en de volwassen deelnemers.

Onder c (**aard**) is bij TALIS géén sprake van bijzondere of strafrechtelijke gegevens, maar wel bij PIAAC. Privacy Company heeft alle variabelen bestudeerd. TALIS bevat geen vragen waaruit bijzondere kenmerken zouden kunnen worden afgeleid.

Dat ligt anders bij PIAAC. PIAAC stelt zowel ja/nee vragen over de nationaliteit (wel of niet Nederlands), maar vraagt ook naar het geboorteland van de deelnemer (met zeven antwoordopties: Marokko, Turkije, Suriname, Nederlandse Antillen, Duitsland, België of ander land) en van de ouders van de deelnemer, en vraagt ook naar de primair gesproken taal thuis, en of de deelnemer een 'native speaker' is of niet. Dat zijn gegevens over nationaliteit, waaruit informatie kan worden afgeleid over etniciteit of ras zoals bedoeld in artikel 9 van de AVG. Zoals toegelicht in paragraaf 6, had het Kohnstamm instituut al eerder aan het ministerie geadviseerd om in de 2024 dataset de variabelen over herkomstland of regio, en thuistaal te onderdrukken. Privacy Company neemt aan dat dat ook geldt voor de ouders van de deelnemers.

PIAAC stelt ook vragen over de gezondheid van deelnemers. Vraag (C_Q03) 'Om welke van de volgende redenen hebt u de afgelopen 4 weken niet naar werk gezocht?', geeft als 2 van de (gesloten) antwoorden: 'Ik was tijdelijk ziek of geblesseerd', en, 'Ik heb een langdurige ziekte of aandoening.'

PIAAC stelt nog een andere vraag over de gezondheid van de deelnemer (I_Q08): "De volgende vraag gaat over uw gezondheid. Zou u in het algemeen zeggen dat uw gezondheid uitstekend, heel goed, goed, redelijk of slecht is?" De (gesloten) antwoorden zijn:

01 Uitstekend

02 Heel goed

03 Goed

04 Redelijk

05 Slecht

In het algemeen adviseert Privacy Company om geen vragen te stellen of observaties te doen over bijzondere persoonsgegevens en om variabelen met bijzondere persoonsgegevens te onderdrukken in de PUF.

Als het ministerie toch bijzondere persoonsgegevens wil verzamelen, mag zij dit in principe alleen doen met uitdrukkelijke toestemming van betrokkenen. Uit artikel 24 onder c UAVG blijkt namelijk dat de uitzonderingsgrond voor wetenschappelijk onderzoek en statistische doeleinden (uit artikel 9 lid 2 onder j AVG) in Nederland alleen van toepassing kan zijn als het vragen van uitdrukkelijke toestemming op grond van artikel 9 lid 2 onder a AVG

onmogelijk blijkt of een onevenredige inspanning kost. Dat is bij beide onderzoeken niet het geval, daardoor kan het ministerie noch de verzameling, noch de verdere verwerking baseren op artikel 24 UAVG.

Privacy Company heeft ook gekeken in de openbare PIAAC vragenlijst van de OESO voor Nederland⁵⁵, en is daar nog een vraag tegengekomen die niet in de testresultaten terugkeerde, namelijk een code voor de interviewer om een kwalificatie te geven van de gezondheid/geschiktheid van de deelnemer.

Interviewer Instruction

Selecteer de juiste responscode voor dit deel van het onderzoek

Antwoorden:

- 03 *Gedeeltelijk compleet/afgebroken*
- 04 *Weigering - geselecteerde persoon*
- 05 *Weigering - andere persoon*
- 07 *Taalprobleem*
- 08 *Moeite met lezen en schrijven*
- 09 *Leerstoornis/verstandelijke beperking*
- 12 *Gehoorbeperving*
- 13 *Blindheid/visuele beperking*
- 14 *Spraakbeperking*
- 15 *Lichamelijke beperking*
- 16 *Andere beperking*
- 17 *Anders (niet genoemd), zoals ziekte of buitengewone omstandigheden*
- 18 *Overleden*
- 21 *Maximum aantal contactpogingen*
- 24 *Tijdelijk afwezig/onbeschikbaar tijdens de veldwerkperiode*
- 25 *Komt niet in aanmerking (als geen gebruik wordt gemaakt van een screener)*
- 27 *Dubbel - reeds ondervraagd*
- 90 *Technisch probleem*

Privacy Company acht dit een onbehoorlijke werkwijze, omdat de interviewer (via sommige antwoordopties) een gezondheidskwalificatie geeft van de deelnemer. Het ministerie zou zich bij OESO moeten inzetten om dit type vragen te voorkomen.⁵⁶

Onder d (**gevolgen**) kan de verdere verwerking van de open datasets in theorie leiden tot heridentificatie van gepseudonimiseerde gegevens. Daarvan kan vrijwel geen sprake meer zijn na de onderdrukking van de variabelen met een koppelbaarheid van 2, en door inachtneming van een wachttijd tussen het verzamelen en publiceren van meer dan een jaar (bij TALIS), zodanig dat er geen koppeltabel meer beschikbaar is op de school naar de identiteit van hun leerkrachten of bestuurders.

⁵⁵ OECD, PIAAC first cycle, Netherlands questionnaire, URL:

<https://www.oecd.org/content/dam/oecd/en/about/programmes/edu/piaac/background-questionnaire/cycle-1/netherlands-bq.zip>.

⁵⁶ Het Ministerie van OCW geeft in een reactie aan dat de nonresponscodes onderscheid maken tussen weigeringen en redenen waarom deelname feitelijk niet mogelijk was. Een deel van de nonresponscodes, zoals taalprobleem, telt ook mee in de internationale responsverantwoording.

Omdat bij PIAAC ook bijzondere persoonsgegevens worden verwerkt, is sprake van een zeer hoge impact bij heridentificatie. Zie hieronder, bij e) voor de geadviseerde waarborgen.

In het algemeen kan publicatie van open data leiden tot verlies van controle en vertrouwelijkheid van persoonsgegevens van de deelnemers. Deelnemers die hebben ingestemd met het OESO onderzoeksdoel kunnen niet verwachten, en hoeven er ook niet op bedacht te zijn, dat hun antwoorden voor bijvoorbeeld commerciële doelen verder worden verwerkt. In die zin wordt hun legitieme verwachting van vertrouwelijkheid geschaad.

Juridisch gezien valt het gebruik van de open data niet onder de verantwoordelijkheid van het ministerie of van de OESO, maar onder de eigen verantwoordelijkheid van de nieuwe gebruikers. Zij moeten zelf een grondslag hebben voor hun verwerking van die gegevens.

Onder e (**waarborgen**) zijn alle variabelen onderdrukt met een koppelbaarheid van 2, met uitzondering van 10 variabelen bij PIAAC, zie paragraaf 6. Het behoud van deze 10 variabelen leidt tot een iets hoger herleidbaarheidsrisico. Statistisch is het mogelijk om met deze 10 variabelen 1 persoon binnen een groep van 2.000 personen te herkennen. Omdat PIAAC echter een steekproef onder de volwassen Nederlanders betreft (en niet een bekende groep van 2.000 personen), is deze kans op herleidbaarheid nog steeds heel erg laag. Dat betekent dat het vrijwel onmogelijk is om individuele deelnemers te identificeren, en dat de gepubliceerde data dus spreekwoordelijk anoniem zijn, maar niet in de juridische betekenis van anoniem.

Een van de belangrijkste waarborgen bij verwerkingen op basis van toestemming is de mogelijkheid om toestemming in te kunnen trekken. Dat is bij PIAAC nooit mogelijk geweest, omdat de onderzoekers alleen een random ID aan de deelnemer hebben toegekend. Bij TALIS hebben de deelnemende scholen nog 1 jaar een koppeltabel bewaard om de deelnemers te kunnen koppelen aan het random ID in de resulterende dataset, om hen op die manier hun recht uit te kunnen laten oefenen om hun toestemming in te trekken.

Artikel 17(3)(d) AVG formuleert een beperking op het recht op gegevenswissing “voor zover verwerking nodig is met het oog op wetenschappelijk of historisch onderzoek of statistische doeleinden overeenkomstig artikel 89, lid 1, voor zover het in lid 1 bedoelde recht de verwezenlijking van de doeleinden van die verwerking onmogelijk dreigt te maken of ernstig in het gedrang dreigt te brengen”. Overweging 65 beschrijft bovendien dat “betrokkenen het recht hebben hun persoonsgegevens te laten wissen en niet verder te laten verwerken wanneer [...] de betrokkenen hun toestemming hebben ingetrokken [...]. Het dient echter rechtmatig te zijn persoonsgegevens langer te bewaren wanneer dat noodzakelijk is voor [...] wetenschappelijk of historisch onderzoek of statistische doeleinden.”

Vanwege de beperking tot ‘wetenschappelijk onderzoek of statistiek’ adviseert Privacy Company het ministerie om nog wel een tweede waarborg te treffen, maatregelen om het hergebruik van de data in te perken tot onderzoekers met legitieme onderzoeksdoelen. Dat geldt voor PIAAC in verband met de aanwezigheid van bijzondere persoonsgegevens, en voor PISA in verband met de kwetsbare aard van de 15-jarige deelnemers. Het ministerie kan bij de OESO aandringen op publicatie van de datasets als RUF in plaats van PUF, om ongewenst hergebruik voor andere dan onderzoeksdoelen te voorkomen. Los daarvan kan het ministerie, samen met het Kohnstamm Instituut, nog een keer kijken in hoeverre publicatie van de bijzondere persoonsgegevens strikt noodzakelijk is voor de legitieme onderzoeksdoeleinden.

Samenvattend is de verdere verwerking door het ministerie en door de OESO van de onderzoeksgegevens die in het kader van PIAAC en TALIS zijn verzameld met als doel anonieme verwerking, onder de AVG verenigbaar met de publicatie van de vrijwel niet herleidbare pseudonieme persoonsgegevens als open data. Zie voor PIAAC verder de beantwoording van vraag 7.6 in verband met de relevantie van de Wet CBS.

7.14 Is hergebruik voor wetenschappelijk, historisch of statistisch onderzoek wettelijk mogelijk wanneer de oorspronkelijke verwerking onrechtmatig tot stand is gekomen, bijvoorbeeld wanneer er geen grondslag was voor de verwerking of niet is voldaan aan de informatieverplichting (artikel 13/14 AVG) tijdens de oorspronkelijke verwerking?

Zoals hierboven toegelicht in antwoord op [vraag 13](#) heeft Privacy Company specifiek getoetst of verdere verwerking voor een (iets) ander doel mogelijk was voor TALIS en PIAAC op de gegeven grondslag toestemming conform de vijf criteria uit artikel 6(4) AVG. Dat pakt goed uit, na het toepassen van het advies om variabelen te schrappen met een koppelbaarheid van 2.

Volgens Privacy Company is de oorspronkelijke dataverzameling niet onrechtmatig, in de zin dat mensen vrijwillig hebben meegewerkt aan de beide soorten onderzoeken voor het ministerie en als onderdeel van Europees vergelijkend onderzoek voor de OESO. Wel was de informatie die aan betrokkenen is verstrekt onjuist, namelijk dat er sprake zou zijn van verwerking als anonieme gegevens. Er was dus wel sprake van toestemming voor de verzameling van de gegevens, maar geen sprake van toestemming voor de verdere verwerking als pseudonieme gegevens.

Die onjuiste informatie heeft Privacy Company als omstandigheid meegewogen in de evenredigheidstoets, zie ook [vraag 13](#), onder het kopje **a) verband**.

Uit de evenredigheidstoets blijkt ook dat de impact voor betrokkenen verwaarloosbaar is en dat de verwerking van de opgeschoonde datasets daarom verenigbaar is met het oorspronkelijk vermelde verwerkingsdoel. Zoals toegelicht onder het kopje **d) gevolgen** valt het hergebruik van de open data niet onder de verantwoordelijkheid van het ministerie of van de OESO, maar onder de eigen verantwoordelijkheid van de nieuwe gebruikers. Zij moeten zelf een grondslag hebben voor hun verwerking van die gegevens. Maar Privacy Company adviseert het ministerie wel om bij de OESO aan te dringen op een extra waarborg, door publicatie van de brondata als RUF in plaats van PUF.

In het algemeen (bij het vaststellen van beleid voor wetenschappelijk, historisch of statistisch onderzoek) moet het ministerie de grondslag voor verwerkingen vooraf goed doordenken. Zoals toegelicht in antwoord op [vraag 13](#) zou het ministerie in de toekomst onderscheid moeten maken tussen vrijwillige deelname aan onderzoek, en de grondslag van vervulling van een taak van algemeen belang (artikel 6(1) onder e AVG) voor de verwerking van de antwoorden.

Als de betrokkene niet juist is geïnformeerd (bijvoorbeeld wanneer toestemming is gevraagd terwijl dit niet de juiste grondslag was) kan het ministerie de gegevens alleen (her)gebruiken als de verdere verwerking verenigbaar is met het oorspronkelijke verzameldoel. Het ministerie moet zich dan wel inspannen om de (niet meer herleidbare) deelnemers in algemene zin te informeren over de aard en gevolgen van deze verdere verwerking.

7.15 Hoe kunnen we een afwegingskader ontwikkelen voor het evenwicht tussen privacy-bescherming en wetenschappelijke bruikbaarheid?

Op hoofdlijnen moet het afwegingskader handvatten bieden om de **vertrouwelijkheid** van de antwoorden te waarborgen, door een toetsing op entropie en koppelbaarheid verplicht te stellen (door de OESO en/of door het ministerie). Zie de antwoorden op [vragen 1 tot en met 3](#). Daarnaast moet het afwegingskader regels opstellen wanneer data als PUF en wanneer alleen als RUF kunnen worden gepubliceerd (zie de antwoorden op de [vragen 12 en 13](#)).

Het afwegingskader moet ook vragen bevatten om de **legitimiteit en proportionaliteit** van de vragen te beoordelen. Die beoordeling is een inschatting van de risico's voor betrokkenen dat ze alsnog herleidbaar worden. Een concreet voorbeeld is om op OESO niveau af te spreken om bepaalde vragen met hoge entropie niet meer te stellen, dus niet vragen naar de specifieke leeftijd van betrokkenen, maar alleen naar een bandbreedte van bijvoorbeeld 10 jaar.

Het kader moet ook de privacy-inbreuk afwegen die het verzamelen van bijzondere persoonsgegevens (over nationaliteit of gezondheid) met zich brengt, en het verzamelen van andere soorten vertrouwelijke, intieme informatie. **Data-minimalisatie** is essentieel. Daarom mogen vragenlijsten niet alleen door inhoudelijke onderzoekers worden opgesteld, maar moet een privacy-expert ook het verwachte nut en belang voor de onderzoekspraktijk afwegen tegen de privacy-inbreuk bij betrokkenen. Daarbij hoort **goed fatsoen** voorop te staan, omdat de overheid deze vragen stelt en dus het goede voorbeeld zou moeten geven (als wetgever van de AVG). Casus specifiek: zou je het fair vinden als een onderzoeker aan de deur jouw gezondheid beoordeelde, of, zou je als ouder willen dat je 15-jarige kind de vraag zou krijgen naar eerdere zwangerschap van een onderzoeker? (zie de antwoorden op [vragen 11 en 13](#)).

Transparantie jegens betrokkenen over het doel van de verwerking is verplicht, ook of ze hun rechten kunnen uitoefenen op bijvoorbeeld inzage of correctie. Daarom moet het afwegingskader concrete normen bevatten voor de kwaliteit van de informatie richting betrokkenen, zowel voorafgaand aan deelname, als na afloop (zie de zeven maatregelen in antwoord op [vraag 11](#)).

Om verwarring over de **grondslag** te voorkomen adviseert Privacy Company om helder onderscheid te maken tussen vrijwillige, op informatie berustende deelname (ook wel *informed consent* genaamd) en de verwerking op basis van de noodzaak voor de vervulling van een taak van algemeen belang (artikel 6(1) onder e AVG) (zie het antwoord op [vraag 13](#)).

7.16 Welke best practices van andere (academische) instellingen kunnen worden overgenomen?

Privacy Company heeft geen inzicht in de best practices van andere (academische) instellingen. Het ministerie heeft wel een samenvatting gestuurd van de input van andere EU lidstaten over publicatie van de PUF. Hieruit blijkt dat de meeste lidstaten aangeven dat publicatie als open data mogelijk is als de gegevens niet (meer) tot personen herleidbaar zijn, soms ook door variabelen te onderdrukken. De reagerende lidstaten noemen verschillende waarborgen om de onderzoeken conform de AVG uit te voeren.

- Grote beslissingen eerst voorleggen aan de interne FG van het ministerie. Op de internationale doorgifte naar (en via) de OESO is artikel 49 (1) onder d van de AVG van toepassing ("*de doorgifte is noodzakelijk wegens gewichtige redenen van algemeen belang.*")
- DPIA uitvoeren of een vergelijkbare risico-inventarisatie.
- De FG raadplegen van het uitvoerende onderzoeksbureau.
- Procedures expliciet maken en publiceren bij het onderzoeksrapport als technische bijlage,
- Voorafgaand aan de dataverzameling (PISA) informatie over de dataverwerking geven aan leerlingen, ouders en schoolpersoneel (conform de artikelen 13 en 14 AVG), incl. contactgegevens van de FG van het ministerie of het onderzoeksbureau.
- Technische en organisatorische beveiligingsmaatregelen treffen (inclusief opleiden van personeel) tegen ongeautoriseerde toegang, verlies of vernietiging van data, voor de beveiliging van de gegevens tijdens de verzameling, overdracht en opslag.
- Overeenkomst tussen de nationale projectmanager en het ministerie
- Pseudonimiseren van de gegevens voor verzending, later weer te koppelen met random ID.

- Volgen van de OESO-richtlijnen.
- Variabelen onderdrukken in de PUF.
- Grondslag op orde maken waar nodig.
- Gevoelige informatie alleen op schoolniveau uitvragen, niet op persoonsniveau
- De koppeltabel bewaren om deelnemers in staat te stellen hun toestemming in te trekken.
- Data pseudonimiseren op schoolniveau, voordat het naar het onderzoeksbureau of ministerie wordt gestuurd.

Dit lijstje biedt geen limitatief overzicht van best practices, maar het ministerie kan wel cherry-picken bij het opstellen van haar afwegingskader (zie het antwoord op [vraag 15](#) hierboven). Om een completer beeld te krijgen van best practices, stelt Privacy Company voor dat het ministerie deze vraag aan de leden van SURF⁵⁷ voorlegt. Het is voorstelbaar dat SURF er ook bij gebaat is om een uniform afwegingskader met best practices aan te bieden aan alle onderzoekers die met persoonsgegevens werken.

7.17 Wat moeten we procedureel wijzigen om dit soort kwesties in de toekomst voor te zijn?

Hier zijn enkele voorstellen om dergelijke kwesties in de toekomst te voorkomen:

- Voer geen steekproeven uit op basis van voorwaarden van anderen, ga in ieder geval het gesprek aan over eisen zoals 'anonimiteit'.
- Neem bij betrokkenheid van anderen duidelijke voorwaarden op in de projectovereenkomst over de publicatie van onderzoeksresultaten, met verwijzingen naar de relevant wettelijke kaders.
- Win juridisch advies in (bijvoorbeeld bij het opstellen van informatie voor betrokkenen).
- Vraag de OESO om een expert in te schakelen op het gebied van entropie en herleidbaarheid nadat de onderzoeks-experts de vragen hebben opgesteld, en leidt ook in eigen huis medewerkers op die dit nog een keer kunnen nakijken voordat de onderzoeken worden uitgevoerd.
- Vraag de OESO om vragen op een andere manier te formuleren, bijvoorbeeld door te vragen naar leeftijdscategorieën in plaats van specifieke leeftijden, of naar regio's in plaats van plaatsnamen.

7.18 Hoe verhoudt de kleine kans van identificatie zich tot de statistisch kleine kans in de CABR casus? En waarom voldoet publicatie in een PUF aan het subsidiariteitsvereiste, kijkend naar de CABR casus?

De omstandigheden bij publicatie van het Centraal Archief Bijzondere Rechtspleging waren ingrijpend anders dan bij de datasets PIAAC en TALIS. In de eerste plaats was de inhoud van de dossiers grotendeels onbekend, en was dus niet duidelijk in hoeverre de personen identificeerbaar waren. In de tweede plaats heeft de AP nadrukkelijk gewaarschuwd dat de dossiers bijzondere persoonsgegevens bevatten en gegevens van strafrechtelijke aard. In de derde plaats waarschuwde de AP het ministerie om ook maatregelen te nemen om familieleden/nabestaanden van de betrokken personen te beschermen. Tenslotte is het specifieke openbaarheidsregime van de Archiefwet van toepassing.

Deze vier omstandigheden doen zich niet voor bij de PIAAC en TALIS datasets. Alle variabelen zijn bekend, alle variabelen met bijzondere of gevoelige persoonsgegevens zijn onderdrukt (door Kohnstamm), er is geen sprake van herleidbaarheid van familieleden of nabestaanden, en de Archiefwet is niet van toepassing omdat het hier om

⁵⁷ SURF is de ict-coöperatie van onderzoeks- en onderwijsinstellingen, en is in de ogen van Privacy Company de meest aangewezen organisatie om de huidige *best practices* op halen bij haar leden.

interviews/onderzoek gaat. De AP geeft aan de 'openbaar' op grond van artikel 14 AW niet hetzelfde is als via internet doorzoekbaar maken, dat daarvoor een proportionaliteitstoets vereist is. Die proportionaliteitstoets pakt verkeerd uit voor het CABR, juist omdat de dossiers veel gevoelige informatie bevatten, maar ook, omdat het mogelijk is om de gegevens, ook de gewone persoonsgegevens, op een minder inbreuk makende manier te ontsluiten.

De analyse van Privacy Company langs de assen van entropie en koppelbaarheid van alle variabelen leidt ertoe dat de opgeschoonde datasets alleen nog pseudonieme persoonsgegevens bevatten waarvan de kans op herleidbaarheid uiterst minimaal is. Door die werkwijze past het ministerie al een effectieve maatregel toe om de proportionaliteit van publicatie in een PUF te waarborgen voor TALIS.

Wat betreft het subsidiariteitsvereiste is de vraag of hetzelfde legitieme onderzoeksdoel niet op een andere, voor de bij de verwerking van persoonsgegevens betrokken personen, minder nadelige wijze kunnen worden verwezenlijkt. Het ministerie en de OESO hebben andere argumenten voor openbaarmaking dan het CABR. Het gaat bij de opgeschoonde PIAAC en TALIS dataset om vrijwel niet herleidbare onderzoeksgegevens om vergelijkend onderzoek te kunnen doen naar onderwijsprestaties en werkdruk in het grote aantal landen dat deelneemt aan het onderzoek. Het gaat dus niet om het vertellen/ontsluiten van individuele dossiers en verhalen met mogelijk veel gevoelige gegevens. Het feit dat het mogelijk is om beperkingen op te leggen aan de openbare toegankelijkheid, via een RUF, betekent niet dat het ministerie die route ook moet kiezen. Door de getroffen maatregelen is de impact op betrokkenen van publicatie in een PUF al vrijwel nihil. De mogelijkheid om die impact verder te verlagen door beperkingen te stellen aan de toegang, draagt weinig bij aan verlaging van het gegevensbeschermingsrisico, maar maakt de toegang wel moeilijker voor geïnteresseerden. Immers, het ministerie moet de OESO overtuigen van het optuigen van de RUF, met bijbehorende administratieve lasten, en al die tijd zouden de Nederlandse data ontbreken voor onderwijsvergelijkingen. Het ligt daarom meer voor de hand dat het ministerie inzet op publicatie in de vorm van een RUF van de volgende versie van de TALIS onderzoeken. Zoals toegelicht in paragraaf 7.6 verzet andere wetgeving dan de AVG, namelijk de Wet CBS, zich wel tegen publicatie in de vorm van een PUF voor de opgeschoonde dataset van het PIAAC onderzoek.

Bijlage A – TALIS-velden met een hoog niveau van koppelbaarheid

In deze tabel zijn alle velden gemarkeerd met een koppelbaarheid van 2. Dit zijn er 146. Het Kohnstamm instituut heeft ervoor gekozen om al deze velden te onderdrukken uit de TALIS dataset die ze op 17 augustus heeft aangeleverd aan de OESO, met uitzondering van de velden met een entropie van nul (groen gekleurd in de tabel). Dat waren er 22. In totaal zijn dus 124 variabelen onderdrukt.

Veldnaam	Entropie	Koppelbaarheid
CNTRY	0	2
IDCNTRY	0	2
IDCNTPOP	0	2
IDLANG_CQ	0	2
TC4G01	0.91053306	2
TC4G02	4.906317583	2
TC4G03	1.139285534	2
TC4G04A	3.667422333	2
TC4G04B	4.33003027	2
TC4G04C	3.649321792	2
TC4G04D	4.78675565	2
TC4G04E	2.40632579	2
TC4G11	1.92707992	2
TC4G12A	2.503777382	2
TC4G12B	2.269906138	2
TC4G12C	1.544675301	2
TC4G16A	0	2
TC4G16B	0	2
TC4G16C	0	2
TC4G16D	0	2
TC4G16E	0	2
TC4G16F	0	2
TC4G17	6.779984928	2
TC4G18A	1.703254408	2
TC4G18B	1.481008791	2
TC4G18C	1.947692557	2
TC4G18D	1.950766896	2
TC4G18E	1.746472048	2
TC4G18F	1.529266795	2
TC4G18G	1.845151003	2
TC4G21A	1.48610762	2
TC4G21B	1.825475781	2
TC4G21C	1.665933271	2
TC4G21D	1.831593377	2
TC4G21E	1.721127673	2
TC4G21G	1.297471127	2

TC4G21H	1.159057012	2
T4Pa-gEGR	1.893801963	2
T4PHEDAT	1.139285534	2
T4PEMPWH	0.999337504	2
T4SCHLOC	1.307472969	2
T4STRATIO	6.745444736	2
T4TPRATIO	5.712421875	2
T4TARATIO	5.329572838	2
T4NUMSTUD	0.95583987	2
T4SCHISCN	0	2
T4TTNRATIO	4.98094518	2
CNTRY	0	2
IDCNTRY	0	2
IDCNTPOP	0	2
TT4G01	0.686762141	2
TT4G02	5.423754278	2
TT4G03	1.171047829	2
TT4G05	5.348836708	2
TT4G09	0.548600088	2
TT4G10	0.381585872	2
TT4G11	1.320263511	2
TT4G12	1.896087538	2
TT4G13A	4.56698429	2
TT4G13B	5.19003263	2
TT4G13C	1.869396073	2
TT4G13D	2.742900339	2
TT4G20A	1.71586603	2
TT4G48BA	0.184685303	2
TT4G48BB	0.232350998	2
TT4G48BC	0.869492844	2
TT4G48BD	0.770769074	2
TT4G48BE	0.783567338	2
TT4G48BF	0.013576964	2
TT4G48BG	0.82315083	2
TT4G48BH	0.770769074	2
TT4G48BI	0.562539375	2
TT4G48BJ	0.985730507	2
TT4G48BK	0.91079496	2
TT4G48BL	0.531882603	2
TT4G49	1.201496273	2
T4Ta-gEGR	2.190409641	2
T4TEMPWH	0.89662077	2
CNTRY	0	2
IDCNTRY	0	2
IDCNTPOP	0	2
IDLANG_CQ	0	2

TC4G01	0.995727452	2
TC4G02	4.682713361	2
TC4G03	0.99730873	2
TC4G04A	3.636360908	2
TC4G04B	4.527179905	2
TC4G04C	3.538518786	2
TC4G04D	4.520079809	2
TC4G04E	2.86703399	2
TC4G11	1.647304223	2
TC4G12A	2.679450431	2
TC4G12B	2.576320295	2
TC4G12C	1.219365264	2
TC4G16A	0.111909909	2
TC4G16B	0.113274303	2
TC4G16C	0.177389453	2
TC4G16D	0.404448386	2
TC4G16E	0.918295834	2
TC4G16F	0	2
TC4G17	6.155399312	2
TC4G18A	1.013241047	2
TC4G18B	1.417571267	2
TC4G18C	1.537445963	2
TC4G18D	1.32457049	2
TC4G18E	1.869344438	2
TC4G18F	1.423168674	2
TC4G18G	1.824516389	2
TC4G21A	1.965584593	2
TC4G21B	1.727859082	2
TC4G21C	1.980062757	2
TC4G21D	1.911974368	2
TC4G21E	1.818333058	2
TC4G21G	1.388063166	2
TC4G21H	1.348921281	2
T4Pa-gEGR	1.859866039	2
T4PHEDAT	0.99730873	2
T4PEMPWH	0.552495114	2
T4SCHLOC	1.222751588	2
T4STRATIO	6.182426339	2
T4TPRATIO	5.692363332	2
T4TARATIO	5.631684616	2
T4NUMSTUD	1.225478813	2
T4SCHISCN	0.942163169	2
T4SCHPRGN	0.418684311	2
T4TTNRATIO	5.866404826	2
CNTRY	0	2
IDCNTRY	0	2

IDCNTPOP	0	2
TT4G01	1.016900639	2
TT4G02	5.480291013	2
TT4G03	1.278812661	2
TT4G05	5.227058075	2
TT4G09	0.849401235	2
TT4G10	0.596483417	2
TT4G11	1.257338771	2
TT4G12	1.803109567	2
TT4G13A	4.532448128	2
TT4G13B	5.091263796	2
TT4G13C	1.810577936	2
TT4G13D	3.688875675	2
TT4G20A	1.476319644	2
TT4G48A	3.218082378	2
TT4G49	1.100452078	2
T4Ta-gEGR	2.229201617	2
T4TEMPWH	0.924632772	2

Bijlage B – PIAAC-velden met een hoog niveau van koppelbaarheid

Privacy Company heeft de PIAAC dataset op dezelfde manier geanalyseerd als de TALIS dataset, dus voor elke variabele gekeken naar entropie en koppelbaarheid. Daar komt een lijst uit naar voren van 204 variabelen met een koppelbaarheid van 2. De lijst bevat ook nog 25 variabelen met een entropie van 0 (nul). Die voegen geen extra risico op herleidbaarheid toe, en zijn daarom groen gekleurd.

De 10 variabelen die het Kohnstamm Instituut graag wil behouden, leiden tot een iets hoger herleidbaarheidsrisico. Statistisch is het mogelijk om met deze 10 variabelen 1 persoon binnen een groep van 2.000 personen te herkennen. Omdat PIAAC een steekproef onder de volwassen Nederlanders betreft (en niet een bekende groep van 2.000 personen) is deze kans op herleidbaarheid nog steeds heel erg laag.

Veldnaam	Entropie	Koppelbaarheid
CNTRYID	0	2
CNTRYID_E	0	2
CNTRY	0	2
CNTRY_E	0	2
AGE_R	5.627409071	2
CI_AGE	5.628035936	2
CALCAGE	5.648306014	2
DIAGERANGE	2.120508034	2
DOBY	5.626944944	2
DOBMM	3.58396237	2
GENDER_R	0.999870879	2
CI_GENDER	0.999901742	2
GENDER	0.999870289	2
ID_PSU	7.656423366	2
REGION	3.31097603	2
URBRUR	0.992950544	2
DI_Q02	0.996079991	2
DI_Q03	2.120508034	2
DI_Q04	1.972511828	2
DI_Q05	2.077909133	2
DI_Q06a	0.295181743	2
A2_Q01a	5.626944944	2
A2_D01b	5.645847509	2
A2_Q01b	3.58396237	2
A2_N02	0.999819099	2
A2_Q03a	0.506766352	2
A2_S03b	1.005286779	2
A2_S04a1	1.174996993	2
A2_S04a2	0.936670699	2
B2_Q01	3.511213022	2
B2_D01a	3.511213022	2

EDLEVEL3	1.52403073	2
B2_Q02b01	0.585767613	2
B2_Q02b02	0	2
B2_Q02b03	0.647007129	2
B2_Q02b04	0.800097376	2
B2_Q02b05	0	2
B2_Q02b06	0	2
B2_Q02b07	0	2
B2_Q02b08	0	2
B2_Q02b09	0	2
B2_Q02b10	0.340264044	2
B2_Q02b11	0	2
B2_Q02b12	0.324235016	2
B2_Q02b13	0.989929607	2
B2_Q02b14	0	2
B2_Q02b15	0	2
B2_Q02b16	0	2
B2_Q02b17	0.337380397	2
B2_Q02b18	0	2
B2_Q02b19	0	2
B2_Q02b20	0.124353719	2
B2_Q02b21	0	2
B2_Q02b22	0	2
B2_Q02b23	0.325709528	2
B2_Q02b24	0.307776634	2
B2_Q02b25	0	2
B2_Q02b26	0.041087108	2
B2_Q02b27	0.151582562	2
B2_Q02b28	0.016044355	2
B2_Q02b29	0	2
B2_Q02b30	0	2
B2_Q02bDK	0.032199927	2
B2_Q02bRF	0.025927257	2
B2_D02b	2.982506635	2
B2_Q03a	0.375664899	2
B2_S03b	0.91515547	2
B2_Q04a	3.64089127	2
B2_Q04b1	4.513202064	2
B2_Q04b2	5.512333991	2
B2_Q04c	2.72746993	2
B2_Q05a	0.722396918	2
B2_Q05b	3.159992978	2
B2_D05b	3.159992978	2
B2_Q05c	3.715230941	2
B2_Q06a	0.919949139	2
B2_Q06b	3.219414882	2

B2_D06b	3.219414882	2
B2_Q06c1	4.350493932	2
B2_Q06c2	5.504363959	2
B2_Q06d	3.406648196	2
B2_Q11	0.737554693	2
B2_Q15	1.710248221	2
B2_Q18	1.369275424	2
C2_Q04	3.029460205	2
C2_D05	0.751909194	2
C2_D06	1.119736535	2
C2_Q09a	0.570296507	2
C2_D09c	0.994984828	2
C2_Q10	5.486575454	2
C2_D10	0.890124362	2
D2_Q01a	9.912485389	2
D2_Q01b	10.06138504	2
D2_Q02a	9.591270028	2
D2_Q02b	9.973703133	2
D2_Q03	1.035926636	2
D2_Q04	0.55426904	2
D2_Q05a1	5.414947097	2
D2_Q05a2	4.516015012	2
D2_Q05a3	3.527910967	2
D2_Q05b1	5.364417452	2
D2_Q05b2	5.009207512	2
D2_Q05b3	3.262069985	2
D2_Q06a	1.067702042	2
D2_Q06b	0.999727567	2
D2_Q06c	0.872199497	2
D2_Q07a	2.311858164	2
D2_Q07b	1.465863123	2
D2_Q07c	0.870082617	2
E2_Q01a	1.344661069	2
E2_Q01b	1.350569093	2
E2_Q02a	1.317824197	2
E2_Q02b	1.345799697	2
E2_Q03	0.874931276	2
E2_Q04	0.368677465	2
E2_Q05a1	5.136886985	2
E2_Q05a2	4.267718117	2
E2_Q05b1	3.823067982	2
E2_Q05b2	4.095795255	2
E2_Q06	2.231700438	2
F2_Q05e	1.856838937	2
F2_Q05f	0.896488182	2
H2_Q02a	0.645967117	2

H2_Q02b	0.55937669	2
H2_Q03a	1.379419982	2
H2_Q03b	2.256093417	2
H2_Q03c	1.904096643	2
H2_Q05b	2.209173663	2
H2_Q07a	1.802506177	2
H2_Q07b	1.725501511	2
J2_Q01	2.412057575	2
J2_Q02a	0.85500807	2
J2_Q02b	1.79439559	2
J2_Q03a	0.982237369	2
J2_Q03b	1.843508912	2
J2_Q03c	4.962416115	2
J2_Q03d1	5.219421583	2
J2_Q03d2	5.366489691	2
TIMEZONE	0.997143333	2
PDSCountry	0	2
BQLANG	0.241756543	2
DALANG	0	2
CNT_BRTH	1.13686134	2
REG_TL2	1.867664684	2
AGE_R_ORG	5.648310867	2
AGEG10LFS	2.311834215	2
AGEG5LFS	3.307012474	2
AGE_BQDI	2.340436332	2
GENDER_BQDI	0.999870289	2
ISCO1C	3.265208139	2
ISCO2C	4.885462621	2
ISCO1L	1.251362985	2
ISCO2L	1.411590605	2
ISIC1C	3.997835484	2
ISIC2C	5.212542586	2
ISIC1L	1.312885603	2
ISIC2L	1.423962341	2
ISCO1_MOTH	2.781591668	2
ISCO2_MOTH	3.680908692	2
ISCO1_FATH	3.41998306	2
ISCO2_FATH	4.943482915	2
ISCOSKIL4	1.611890322	2
ISCOSKIL4_MOTH	1.645769213	2
ISCOSKIL4_FATH	1.560945774	2
BIRTHRGN	0.655230887	2
CTRYRGN	0	2
CRGN_UN	0	2
REG_TL2_new	1.867664684	2
BIRTHCNTRY_BQDI	0.571416888	2

EDU_BQDI	1.740061978	2
CTRYQUAL	0.392980157	2
EDCAT6_TC1	2.379481157	2
EDCAT7_TC1	2.55050798	2
EDCAT8_TC1	2.61977217	2
EDA_PISCED	3.511213022	2
EDA_ISCED123	3.107011048	2
EDA_ISCED1	2.647463776	2
EDE_PISCED	3.159992978	2
EDE_ISCED123	2.665738772	2
EDE_ISCED1	2.127112566	2
EDI_PISCED	3.219414882	2
EDI_ISCED123	2.738338527	2
EDI_ISCED1	2.252179102	2
A2_Q03a_T	0.506766352	2
B2_Q01_TC1	2.920359017	2
B2_Q04b1_C	2.192226465	2
B2_Q05b_TC1	2.432639608	2
C2_Q07_T	1.341929521	2
D2_Q05a1_C	3.049491682	2
D2_Q05b1_C	3.08038812	2
E2_Q05a1_C	2.789488497	2
E2_Q05b1_C	2.703559324	2
EDU_BQDI_C	1.704071182	2
YRSQUALC2_C	2.94024751	2
EDA_ISCED1_C	2.55050798	2
EDE_ISCED1_C	2.06185497	2
EDI_ISCED1_C	2.222448452	2
EDFIELD	3.658218449	2
EDCURFIELD	3.715163643	2
ICTEVER	0.00715537	2
PARTNER	0.954703738	2
CHILDNR	2.048508493	2
ETSAGEG5	3.312799605	2
ETSAGEG10	2.316907982	2

