Reactie op het "Wetsvoorstel pseudonimiseren van leerlinggegevens ten behoeve van de toegang tot en het gebruik van digitale leermiddelen" d.d. 18-05-2016 door prof. dr. E.R. Verheul CISSP CISA[1]

Onderstaande reactie is mijn bijdrage aan de internetconsultatie geïnitieerd vanuit het ministerie van Onderwijs, Cultuur en Wetenschap, zie https://www.internetconsultatie.nl/wetpseudonimiseren.

Mijn reactie bestaat uit een conclusie en vijf onderbouwde delen.

Conclusie

Het wetsvoorstel beoogt pseudoniemen ("ketenpseudoniem") te introduceren die worden gedeeld met (publiek/private) partijen in het onderwijs: scholen, distributeurs en uitgevers. Het woord pseudoniem is daarmee een eufemisme voor persoonsnummer. Het fundamentele beginsel dat het wetvoorstel zou moeten hanteren is dat iedere partij zijn *eigen* pseudoniem van een leerling heeft. Betoogd wordt dat dit beginsel ook het gangbare is in een on-line context.

Het niet volgen van dit beginsel in het wetsvoorstel introduceert risico's in het onderwijs omdat partijen kunnen samenspannen om leerling gegevens te combineren, te verrijken en zo ook leerlingen te kunnen identificeren. Partijen kunnen ook worden gehackt waarbij anderen deze koppeling uitvoeren. Het afwijken van het beginsel is ook niet nodig indien gebruik wordt gemaakt van *Privacy Enhancing Technology* (PET).

Het gebruik van PET is echter onvoldoende onderzocht door de opstellers van de wet en in de Privacy Impact Assessment (PIA). Daarbij zijn reeds voor de Nederlandse overheid *Privacy Enhancing Technology* ontwikkeld op basis waarvan het genoemde fundamentele beginsel kan worden gerealiseerd. Om die reden kan gesteld worden dat het wetsvoorstel in strijd is met twee basis privacy beginselen: data minimalisatie en beveiliging volgens stand der techniek. Tot slot kijkt het wetsvoorstel niet naar de samenhang met het authenticatievraagstuk op korte of langere termijn in bredere onderwijs context. Dit is een gemiste kans is en werkt desinvestering in de hand.

Onderbouwende delen

1. *Het wetvoorstel werpt een rookgordijn op door persoonsnummers eufemistisch pseudoniemen te noemen.*
2. *Risico´s van sector brede persoonsnummers in het onderwijs*
3. *Een privacy vriendelijk alternatief is mogelijk*
4. *Er is onvoldoende naar privacy vriendelijke alternatieven gekeken*
5. *De problematiek wordt niet in relatie gebracht met on-line identificatie binnen het onderwijs*

1. Het wetvoorstel werpt een rookgordijn op door persoonsnummers eufemistisch pseudoniemen te noemen.

Het wetsvoorstel [1.] werpt een rookgordijn op en beoogt feitelijk de introductie van nieuwe persoonsnummers in het onderwijs die worden gedeeld met alle betrokken partijen: scholen, distributeurs en uitgevers. Met andere woorden: het wetsvoorstel beoogt de invoering van een persoonsnummer dat zowel wordt gedeeld met publieke als met private deelnemers.

[1] Deze reactie staat op https://www.cs.ru.nl/E.Verheul/presentations/outline.htm

Dit gegeven staat enigszins verborgen in het wetsvoorstel in de definiërende zin: “Een pseudoniem is een unieke identiteit voor leerlingen die door elke leverancier kan worden gebruikt, zonder dat direct te herleiden is om welke specifieke leerling het gaat”. De uitgevoerde Privacy Impact Assessment (PIA) [2.] waarnaar wordt verwezen in het wetsvoorstel (zonder overigens aan te geven waar deze kan worden gevonden!) is explicieter. Daar staat bijvoorbeeld op pagina 8:

“Wat is een juiste reikwijdte van het ketenpseudoniem? Dit uit zich bijvoorbeeld in de vraag waarom hetzelfde ketenpseudoniem voor alle distributeurs en uitgevers van leermiddelen gebruikt wordt.”

Als onderdeel van het rookgordijn misbruikt het wetsvoorstel ook de term ‘pseudoniem’ omdat een pseudoniem in een on-line context (en standaarden zoals de SAML standaard [2.]) altijd partij specifiek is. Dat wil zeggen, dat verschillende partijen dezelfde leerling kennen onder verschillende pseudoniemen. Dit beginsel is binnen het wetsvoorstel verlaten omdat met een naïeve toepassing van dergelijke pseudoniemen bepaalde functionaliteit niet mogelijk is. Een voorbeeld van dergelijke functionaliteit is de terugkoppeling van leerling test resultaten vanuit een uitgever naar de school: als school en uitgever verschillende pseudoniemen gebruiken is dat lastig. Betoogd zal worden dat het gebruik van verschillende pseudoniemen in deze context evenwel toch mogelijk is middels *Privacy Enhancing Technology* (PET).

2. Risico´s van sector brede persoonsnummers in het onderwijs
Door de introductie van sector brede persoonsnummers in het onderwijs ontstaan koppelrisico’s in het onderwijs. Het is ook niet voor niets dat Artikel 31 van de Wet bescherming persoonsgegevens voorschrijft dat de Autoriteit Persoonsgegevens een “voorafgaand onderzoek” uitvoert voorafgaande aan de introductie van persoonsnummers.

Het is mij overigens onduidelijk of de kennelijk voorziene (maar nog niet ingevulde) “reactie” van de Autoriteit Persoonsgegevens in Sectie 5 van het wetsvoorstel [1.] als een dergelijk “voorafgaand onderzoek” moet worden bestempeld. Ook omdat een “reactie” wat vrijblijvend klinkt waarbij Artikel 31 van de Wet bescherming persoonsgegevens spreekt over een “verklaring omtrent de rechtmatigheid van de gegevensverwerking”. Ik vraag mij af of een private partij op deze manier zou kunnen acteren richting de Autoriteit Persoonsgegevens.

Als nadere duiding van dergelijke koppelrisico’s: distributeurs en uitgevers kunnen gaan samenwerken om hun leerling gegevens te combineren, te verrijken en zo ook leerlingen te kunnen identificeren. Meer zorgwekkend is dat partijen (scholen, distributeurs, uitgevers) ook gehackt kunnen worden en waarbij de koppeling en identificatie door de hackers wordt uitgevoerd. Het resultaat kan worden misbruikt, verkocht of zelfs worden gepubliceerd als onderdeel van een chantage. Twee min of meer willekeurige incidenten onderstrepen dit gevaar. De hack in de zomer van 2015 op het Amerikaanse *Office of Personnel Management* [4.], toch niet de eerste beste partij, leidde tot een compromittatie van meer dan 22 miljoen persoonsgegevens waaronder van CIA medewerkers. De hack op de ‘overspelsite’ Ashley Madison [5.], ook in de zomer van 2015, leidde tot de compromittatie van meer dan 37 miljoen persoonsgegevens. Hierbij werden de persoonsgegevens ook gepubliceerd nadat een chantage poging mislukte.

Daarbij is mijn indruk dat uitgevers van elektronische leermiddelen nog onvoldoende volwassen zijn op het terrein van informatiebeveiliging. De uitgever van mijn eigen dochter (eerste klas middelbare school) maakt op dit moment niet eens gebruik van TLS of SSL voor de bescherming van de

communicatie. Dit betekent dat zij en al haar medeleerlingen bijvoorbeeld kwetsbaar zijn voor digitale kinderlokkers [6.]. Vanuit de onderwijs sector begrijp dat ik de betreffende uitgever nog een van de betere is qua beveiliging. Kortom, kunnen wij van dergelijke partijen wel verwachten dat zij leerling gegevens adequaat beveiligen? Dit onderstreept in ieder geval het belang van de data minimalisatie: gegevens (waaronder persoonsnummers) die uitgevers en distributeurs niet hebben, kunnen zij ook niet verliezen!

Overigens vinden koppelingen tussen school, distributeur en uitgevers in de huidige praktijk veelal plaats op de volledige naam van de leerling, diens school en klas. Dit introduceert nog veel meer risico's dan persoonsnummers. Mogelijk dat de beweegredenen van de opstellers van het wetsvoorstel de mitigatie van de huidige risico's is. Hoewel ik deze beweegredenen begrijpelijk zou vinden, denk ik dat het tijd is voor structurele oplossingen en niet voor tijdelijke, suboptimale oplossingen.

De PIA [2.] onderkent op p. 38 ook de beide, bovengenoemde koppelrisico's rond de voorgestelde opzet in het wetsvoorstel. De PIA relateert deze risico's ook aan de risico's in de huidige praktijk en doet daar de volgende opmerkelijke uitspraak over: "Het [koppel] risico wordt met de introductie van het ketenpseudoniem niet groter of kleiner dan in de huidige situatie: distributeurs en uitgevers hebben een unieke identificator voor een leerling nodig om een door een leerling bij de distributeur besteld elektronisch leermiddel, door de uitgever aan diezelfde leerling beschikbaar te laten stellen."

Ik vraag mij af of de opstellers van de PIA [2.] erg blij zijn met de wijze waarop hun analyse in het wetsvoorstel wordt gebruikt in de volgende zin op pagina 15: "Op basis van de PIA kan gesteld worden dat het risico op koppelbaarheid van de gegevens van leerlingen door gebruik van het PGN als basis voor het pseudoniem verwaarloosbaar zijn". In combinatie met bovenstaande PIA citaat kan men dit immers lezen als uit een uitspraak dat de PIA opstellers menen dat de risico's ook in de huidige praktijk verwaarloosbaar zijn. Ik vraag me af of dit een conclusie is die de PIA opstellers delen.

Net zoals op pagina 13 van de PIA wordt op onduidelijke gronden geconcludeerd dat zo een (onderwijs brede) "identificator" noodzakelijk is. In de volgende sectie zal ik betogen dat dit niet correct is.

3. Een privacy vriendelijk alternatief is mogelijk

Structurele oplossingen bestaan. Voor de Nederlandse overheid is reeds in 2014 een *Privacy Enhancing Technology* ontwikkeld waarmee het basis beginsel van "verschillende partijen/verschillende pseudoniemen" zonder verlies van functionaliteit wel gerealiseerd kan worden. Deze technologie heet polymorfe pseudonimisering, zie ondermeer [7.] en [8]. In [8.] Wordt beschreven hoe de polymorfe opzet de privacy problematiek specifiek in het onderwijs kan adresseren met in acht neming van het genoemde basis beginsel.

Door toepassing van polymorfe pseudonimisering ontstaan geen nieuwe persoonsnummers zoals die wel in het wetsvoorstel ontstaan. De kracht van polymorfe pseudonimisering is dat het granulair toelaat dat bepaalde partijen gegevens kunnen delen zoals in het onderwijs de uitgever die testresultaten met de school kan delen. Bij polymorfe pseudonimisering kunnen verschillende uitgevers de gegevens van een leerling evenwel niet koppelen (tenzij dat noodzakelijk zou zijn).

Door het bestaan van deze technologie kan gesteld worden dat het wetvoorstel in strijd is met twee privacy basis beginselen: data minimalisatie en beveiliging.

Het eerste beginsel is opgenomen in Artikel 5 in de aankomende Europese privacy verordening [9.]. Dit beginsel behelst dat alleen de minimaal benodigde hoeveelheid persoonsgegevens mag worden verwerkt. Omdat het wetsvoorstel onnodig persoonsnummers introduceert, worden meer persoonsgegevens verwerkt dan wat minimaal nodig is.

Het tweede beginsel (opgenomen als Artikel 13 in de Nederlandse wet bescherming persoonsgegevens) stelt dat de beveiliging van persoonsgegevens volgens de stand der techniek moet geschieden. Polymorfe pseudonimisering beschermt de persoonsgegevens van leerlingen beter dan de techniek gesuggereerd in het wetsvoorstel zodat dit niet volgens de stand der techniek is.

De opvatting in het wetsvoorstel dat sector brede pseudoniemen (persoonsnummers) noodzakelijk zijn komt mogelijk voort uit de uitgevoerde Privacy Impact Assessment. Daar wordt namelijk op onder meer pagina's 13 en 38 op onduidelijke gronden geconcludeerd dat een (onderwijs brede) "identificator" noodzakelijk is. Bij deze conclusie is echter onvoldoende gekeken naar de toepassing van *Privacy Enhancing Technology* zoals polymorfe pseudonimisering.

4. Er is onvoldoende naar privacy vriendelijke alternatieven gekeken
Curieus daarbij is dat de PIA [2.] op pagina 57 wel verwijst naar een voor Kennisnet uitgevoerd onderzoek naar de toepassing van polymorfe pseudonimisering maar dit verder niet bespreekt. Een kwaaddenkend persoon zou kunnen denken dat de privacy gunstige eigenschappen de opstellers van de PIA niet uitkwamen en in de definitieve versie van het rapport zijn geschrapt waarbij men alleen vergeten is de referentie te schrappen.

Het niet bespreken van polymorfe pseudonimisering in de PIA is verbonden aan de uitspraak op pagina 11 van de Memorie van Toelichting binnen het wetsvoorstel [1.]. Daar wordt namelijk gesteld: "Er is ook zorgvuldig gekeken naar alternatieven." Dit is dus onjuist omdat er onvoldoende gekeken is naar de toepassing van *Privacy Enhancing Technology*.

5. De problematiek wordt niet in relatie gebracht met on-line identificatie binnen het onderwijs
Het wetsvoorstel poogt alleen een prangend privacy probleem in het onderwijs op te lossen (vergelijk Sectie 2) maar kijkt niet naar de samenhang met het authenticatievraagstuk op korte of langere termijn. Scholen zijn conventionele leerboeken aan het vervangen door hun elektronische evenknieën, elektronische leerboeken, of kortweg e-leerboeken. De term is daarbij wat misleidend, omdat een e-leerboek veel meer functionaliteit omvat dan een gewoon leerboek. Een e-leerboek maakt rijkere inhoudsvormen vormen mogelijk zoals audio en video. Door het toenemende gebruik van e-leerboeken wordt het leerling on-line identificatie vraagstuk steeds belangrijker niet alleen vanuit privacy maar ook vanuit beveiliging.

In de huidige praktijk zijn het de scholen die leerlingen voorzien van on-line identificatiemiddelen waarvan de consensus toch is dat dit niet hun kern competentie is. De vraag is aldus hoe leerlingen zich betrouwbaar on-line kunnen identificeren. Het wetsvoorstel zou ook deze problematiek in overweging hebben moeten nemen, bijvoorbeeld door de aanhaking op het Nederlandse eID stelsel in wording (Idensys).

In het voorstel wordt weliswaar Idensys genoemd en dat de onderwijssector daarin participeert maar er wordt niet inhoudelijk ingegaan op wat dat zou kunnen betekenen, waaronder rond de toepassing van polymorfe pseudonimisering daarbinnen [7.].

Het is een gemiste kans dat het wetsvoorstel niet de samenhang met het authenticatievraagstuk op korte of langere termijn in bredere onderwijs context beziet. Dit werkt ook disinvestering in de hand, met name rond de nummervoorziening.

#	Referentie/Noten
[1.]	Voorstel van wet tot wijziging van de Wet op het primair onderwijs, de Wet primair onderwijs BES, de Wet op de expertisecentra, de Wet op het voortgezet onderwijs, de Wet voortgezet onderwijs BES, de Wet educatie en beroepsonderwijs en de Wet educatie en beroepsonderwijs BES in verband met het pseudonimiseren van leerling- of deelnemergegevens ten behoeve van de toegang tot en het gebruik van digitale leermiddelen. Zie https://www.internetconsultatie.nl/wetpseudonimiseren Cache hier.
[2.]	PBLQ HEC, Privacy Impact Assessment Nummervoorziening in de Leermiddelenketen, versie 1.0, 27 mei 2015. Zie https://www.rijksoverheid.nl/documenten/rapporten/2015/05/27/privacy-impact-assessment-nummervoorziening-in-de-leermiddelenketen Cache hier.
[3.]	https://www.oasis-open.org/committees/download.php/21111/saml-glossary-2.0-os.html
[4.]	Washington Post, Hacks of OPM databases compromised 22.1 million people, federal authorities say, 9 juli 2015. Zie http://www.washingtonpost.com.
[5.]	KrebsonSecurity, Online Cheating Site AshleyMadison Hacked , 19 juli 2015. Zie http://krebsonsecurity.com.
[6.]	Dit is door mij gedemonstreerd tijdens de IRMA meeting op 6 november 6, 2015. Zie https://www.cs.ru.nl/E.Verheul/presentations/outline.htm
[7.]	Programma eID, Polymorphic Pseudonymization, versie 0.91, 07 July 2014 Zie https://www.idensys.nl/fileadmin/bestanden/idensys/documenten/basisdocumentatie/documentatieset/PP_Scheme_091.pdf
[8.]	E. Verheul, Polymorphic pseudonyms in the education sector, Manuscript, 2015. Zie http://eprint.iacr.org/2015/1228
[9.]	HET EUROPEES PARLEMENT EN DE RAAD, Voorstel voor algemene verordening gegevensbescherming, 2012, COM/2012/011. Zie http://eur-lex.europa.eu.