Schriftelijke inbreng Prof. mr. A. Berlee t.b.v. deskundigenbijeenkomst over het wetsvoorstel Wet implementatie Open data richtlijn (36.382) inzake 'Openbaarheid versus privacy: (toekomstige) technische mogelijkheden' van de commissie voor Digitalisering (DIGI) van de Eerste Kamer van 21 mei 2024

Geachte leden van de commissie voor Digitalisering,

Het algemene uitgangspunt van de richtlijn en de voorliggende wet is dat overheidsgegevens "zo open als mogelijk, zo gesloten als nodig" moeten zijn.[1] Hiermee wordt de wens uitgesproken om hergebruik van overheidsinformatie zoveel waar mogelijk te faciliteren met eerbiediging voor de persoonlijke levenssfeer van degene over wie informatie openbaar wordt gemaakt en in lijn met de bescherming van diens persoonsgegevens. Dit is een goed streven, maar in de praktijk gaat dit gepaard met een uitvoeringscomplexiteit die kan leiden tot onderproductie van overheidsinformatie geschikt voor hergebruik, maar (juist) ook tot het onbedoeld toegankelijk maken van persoonsgegevens, wat ernstige risico's voor de persoonlijke levenssfeer van betrokkenen met zich mee kan brengen.

Openbaarheid en toegankelijkheid

Hoewel Who zelf geen grondslag vormt voor het openbaar maken van (persoonsg)gegevens, zorgt deze wet er wel voor dat gegevens die op grond van een andere wet al openbaar zijn, (zeer) toegankelijk worden gemaakt. Het openbaar maken van persoonsgegevens is terecht met vele waarborgen omgeven, in het bijzonder moet deze voldoen aan de vereisten die de Algemene verordening gegevensbescherming (AVG) daaraan stelt.

In de praktijk wordt aan de uitwerking van de waarborgen omtrent openbaarheid van persoonsgegevens op uiteenlopende wijze vormgegeven. Het is belangrijk te onthouden dat (persoons)gegevens openbaar kunnen zijn op grond van de wet, maar toch niet gemakkelijk te vinden, op te vragen of verder te gebruiken zijn. Zo kunnen er vereisten worden gesteld aan zowel de vorm van het *opvragen* van de informatie,[2] als de vorm van *verstrekking* van de (persoons)gegevens.[3] Het stellen van veel van deze vereisten kan ook met behoud van de 'openbaarheid' van de gegevens. Dat betekent dat er een verschil bestaat tussen het 'openbaar' zijn van gegevens en de 'toegankelijkheid' ervan en de mogelijkheid om de persoonsgegevens verder te gebruiken. Het wetsvoorstel beoogt niet de openbaarheid te veranderen, maar juist de toegankelijkheid en de mogelijkheden om de persoonsgegevens verder te verwerken.[4]

De actieve verplichting tot het klaarmaken voor hergebruik van overheidsinformatie brengt met zich mee dat het opvragen van gegevens ten behoeve van hergebruik niet langer het uitgangspunt is. Daarnaast beoogd de Who en het wetsvoorstel de wijze van

[1] KST II 2022–2023, 36 382, nr. 3, blz. 22.
[2] Is informatie decentraal of centraal beschikbaar, is de aanvraag per post/e-mail/online of alleen in persoon opvraagbaar. Moet er worden betaald, worden er verificatievragen gesteld, kan informatie in bulk worden verzocht, worden er beperkingen opgelegd ten aanzien van de doelen waarvoor de gegevens mogen worden opgevraagd, etc.?
[3] Worden de gegevens enkel mondeling verstrekt, op papier, per e-mail, per PDF of in machineleesbaar formaat?
[4] Hoewel, zie in deze ook 2024Z01868 Position paper d.d. 6 februari 2024 - Position paper J. Wolswinkel t.b.v. rondetafelgesprek over het wetsvoorstel Wet Implementatie Open Data richtlijn d.d. 13 februari 2024.

verstrekking van data voor hergebruik ook grotendeels te stroomlijnen. Wanneer het gaat om overheidsinformatie waarin persoonsgegevens zijn vervat biedt het wetsvoorstel enkele waarborgen. Immers, ook het hergebruik is een verwerking van persoonsgegevens en deze moet voldoen aan de AVG.[5] In veel gevallen zal het actief beschikbaar stellen van informatie ten behoeve van hergebruik van persoonsgegevens dus niet plaatsvinden omdat dit een verdere verwerking van persoonsgegevens is die niet verenigbaar is met het doel waarvoor de gegevens ooit zijn verzameld.[6] Zie in dit geval uitgebreid het debat in de Tweede Kamer inzake de uitzonderingspositie van openbare registers ingesteld bij wet, zoals is ingevoerd na bezwaar van de Autoriteit Persoonsgegevens.[7]

Pseudonimiseren en anonimiseren

In het wetsvoorstel en de toelichting wordt er mede daardoor ook veel nadruk gelegd op de mogelijkheden die pseudonimiseren en anonimiseren met zich mee zouden kunnen brengen. Pseudonimiseren is een verwerking van persoonsgegevens op een zodanige wijze dat de persoonsgegevens niet meer aan een specifieke betrokkene kunnen worden gekoppeld zonder dat er aanvullende gegevens gebruikt worden, die apart en beveiligd bewaard worden.[8] De verdere verwerking van gespeudonimiseerde gegevens blijft echter een verwerking van persoonsgegevens. Pseudonimiseren van persoonsgegevens is daarmee wezenlijk anders dan anonimiseren. Het anonimiseren van een dataset met persoonsgegevens leidt tot een dataset zonder persoonsgegevens. Het voordeel van een geanonimiseerde dataset lijkt me helder. Waar persoonsgegevens veelal niet voor hergebruik beschikbaar mogen worden gesteld, geldt deze beperking niet voor geanonimiseerde datasets,[9] omdat deze geen persoonsgegevens bevatten. Om hergebruik te stimuleren en persoonsgegevens beschermen wordt er dus veel verwacht van anonimiseren.

Anonimiseren wordt niet gedefinieerd in de AVG, maar behelst een verwerking van persoonsgegevens waarvan het resultaat een dataset is waaruit de verwerkingsverantwoordelijke (de overheidsinstantie) of een derde (een afnemer) in redelijkheid geen natuurlijk persoon meer kan identificeren. Om te bepalen of een natuurlijk persoon identificeerbaar is, ‘moet rekening worden gehouden met alle middelen waarvan redelijkerwijs valt te verwachten dat zij worden gebruikt door de verwerkingsverantwoordelijke of door een andere persoon om de natuurlijke persoon direct of indirect te identificeren’.[10] Het proces van anonimiseren is dus pas geslaagd wanneer het niet meer redelijkerwijs mogelijk is om uit de dataset zelf een natuurlijk

---

[5] De richtlijn doet geen afbreuk aan de AVG. Zie Overweging 52 van Richtlijn 2019/1024 van het Europees Parlement en de Raad van 20 juni 2019 inzake open data en het hergebruik van overheidsinformatie (herschikking), Pb. L. 172/56.

[6] Artikel 5, lid 1, sub b) *jo* artikel 6, lid 4, AVG.

[7] KST II 2023–2024, 36 382, nr. 6, blz. 23-25 en de verschillende voorgestelde amendementen, m.n. KST II 2023–2024, 36 382, nr. 7, 12, 13 en 16. Alsmede Handelingen TK 2023/2024, nr. 47, item 22, vanaf blz. 4.

[8] Artikel 4, onder 5) AVG definieert pseudonimisering als: “het verwerken van persoonsgegevens op zodanige wijze dat de persoonsgegevens niet meer aan een specifieke betrokkene kunnen worden gekoppeld zonder dat er aanvullende gegevens worden gebruikt, mits deze aanvullende gegevens apart worden bewaard en technische en organisatorische maatregelen worden genomen om ervoor te zorgen dat de persoonsgegevens niet aan een geïdentificeerde of identificeerbare natuurlijke persoon worden gekoppeld;”

[9] Let wel, ook voor het proces van anonimiseren *zelf* een grondslag nodig is, of deze moet kwalificeren als een verdere verwerking die voldoet aan de vereisten van het doelbindingsbeginsel zoals neergelegd in artikel 5 lid 1 sub b) AVG en onder meer uitgewerkt in artikel 6 lid 4 AVG.

[10] Overweging 26 bij de AVG stelt dat: “Om te bepalen of een natuurlijke persoon identificeerbaar is, moet rekening worden gehouden met alle middelen waarvan redelijkerwijs valt te verwachten dat zij worden gebruikt door de verwerkingsverantwoordelijke of door een andere persoon om de natuurlijke persoon direct of indirect te identificeren”. Hieraan is uitleg gegeven door het Hof van Justitie in meerdere zaken.

persoon in de dataset te kunnen herleiden, maar dus ook niet in combinatie met andere gegevens waar men (de verwerkingsverantwoordelijke of een derde) redelijkerwijs toegang toe heeft. Dat maakt het proces van anonimiseren omvangrijk en ingewikkeld.

Nog één opmerking over die 'redelijkheid'. In dit kader maak ik mij ook wel enige zorgen over de enorme toevlucht die AI heeft genomen. Dit kan ons ongelofelijk veel moois brengen, alleen bezien in het kader van 'redelijk' is, maak ik me daar ook wel zorgen om. We moeten namelijk wel constateren dat de toegankelijkheid van taalmodellen,[11] die almaar grotere hoeveelheden data kunnen verwerken en daar nuttige informatie uit kunnen destilleren, ook met zich meebrengen dat de inzet van dit soort taalmodellen ook moet worden meegenomen als een 'redelijk middel' om natuurlijk personen te kunnen identificeren. Dit kan een enorme druk kan leggen op anonimiseringstechnieken.[12]

Het uitvoeren van de actieve verplichting tot hergebruik die in het wetsvoorstel is vervat, vereist van veel (ook kleinere) instellingen dat zij nagaan of anonimiseren een middel is dat kan en mag worden ingezet om gegevens voor hergebruik beschikbaar te maken. Dit vereist *per dataset* een afweging in het licht van de stand van de techniek met betrekking tot anonimiseringstechnieken die geschikt kunnen zijn voor dit type dataset. Daarnaast moet ook een inventarisatie plaatsvinden van de al bestaande aanvullende gegevens die beschikbaar zijn, en een afweging of het redelijk is dat deze aanvullende gegevens eventueel worden gebruikt om tot herleiding te kunnen komen. Daarbij moet ook in ogenschouw genomen worden dat hergebruik in beginsel niet is beperkt tot een afgebakende groep afnemers. Kortom, dat vergt nogal wat. Het is niet onmogelijk, maar zeker niet gemakkelijk.

Anonimiseren vergt daarnaast dat met regelmatige tussenpozen wordt gekeken of de gegevens nog afdoende zijn geanonimiseerd. Een dataset, die drie jaar geleden als 'geanonimiseerd' gold volgens de stand van de techniek op dat moment, is dat heden ten dage wellicht niet meer en dat heeft dus ook gevolgen voor de mogelijkheid om deze gegevens actief voor hergebruik beschikbaar te stellen. Deze constatering dat een eenmaal geanonimiseerde dataset dus niet voor altijd dat ook hoeft te blijven vereist een regelmatige controle en kennis van de laatste stand van de techniek en de (nieuwe) aanvullende gegevens die beschikbaar zijn gekomen sinds laatste controle. Dat brengt mij tot amendement 10.

Amendement 10

Artikel I, onderdeel B, waarin Artikel 5, vierde lid, Who komt te luiden dat bij of krachtens amvb regels kunnen worden gesteld over het anonimiseren of pseudonimiseren van documenten die persoonsgegevens bevatten en is bij amendement 10 ingevoegd.[13] De toelichting op het artikel is zeer beperkt en ook in de parlementaire behandeling daarvan is er maar zeer beperkt wat over gezegd.[14] Zoals ik deze begrijp, zal de bepaling de mogelijkheid moeten bieden om bij amvb een reeks technische maatregelen op te stellen en aan te bevelen als de laatste standaarden die door instellingen die onder de Who

[11] Ik beperk me voor nu even daartoe.
[12] De vraag of het taalmodel de *juiste* persoon identificeert is
[13] KST II 2023–2024, 36 382, nr. 10.
[14] Met uitzondering van Handelingen TK 2023/2024, nr. 47, item 22, blz. 3. Een debat daarover bleef uit.

vallen zouden moeten worden toegepast; een soort minimumstandaard. Het lijkt dat hiermee wordt gepoogd om de instellingen te helpen bij het in kaart brengen of de door hen gebezigde technieken nog de stand van de techniek zijn of dat deze geüpdatet moeten worden.

Het artikel ziet daarmee echter alleen op het technische onderdeel van het proces van anonimiseren (en pseudonimiseren) en mist daarmee een belangrijke andere pijler: de (redelijke) beschikbaarheid van aanvullende informatie elders. Ook kan deze amvb in algemene zin helpen, maar zal in de praktijk nog altijd in een concreet geval er moeten worden bezien welke inzet van anonimiseringstechnieken het meest geschikt is en gecontroleerd moeten worden of deze juist is toegepast. Daar kan de amvb naar verwachting niet of maar zeer beperkt bij helpen.

'Zo open als mogelijk, zo gesloten als nodig'
Dat betekent dat zich twee risico's voordoen. Allereerst het risico dat gegevens niet voor hergebruik beschikbaar komen omdat organisaties ten onrechte in sommige gevallen denken niet te kunnen of mogen anonimiseren. Dat kan leiden tot het niet verwezenlijken van het doel van de wet, namelijk dat overheidsinformatie 'zo open als mogelijk' zou moeten zijn. Het tweede risico is dat in de uitvoering ten onrechte wordt gedacht dat persoonsgegevens voldoende geanonimiseerd zijn, terwijl dat niet zo is. Dit kan resulteren in het onrechtmatig toegankelijk maken van grote hoeveelheden persoonsgegevens voor een onbeperkte groep in machineleesbaar formaat. Dit kan bij aanvang zo zijn maar gegevens kunnen na verloop van tijd en na onvoldoende/onjuiste controles toch persoonsgegevens blijken te zijn. Ofwel, dat de wet het uitgangspunt 'zo gesloten als nodig' niet waar kan maken. Beide zijn onwenselijk.

Slot
Anonimiseren vergt voortdurend onderzoek naar de werking van (de-)anonimiseringstechnieken. Daarnaast moet in de uitvoering blijvend aandacht worden besteed aan (overige) openbaar gemaakte gegevens en wat deze dataset daaraan gaat toevoegen. De vraag is of de uitvoering daartoe in staat is. Volgens mij vraagt dat andersoortige maatregelen, bijvoorbeeld investering in een bundeling van kennis op dit gebied, of een voorziening in aanvulling op bijvoorbeeld het reeds op grond van de Data Governance Act (DGA) vereiste 'bevoegde orgaan'.[15] Eén gecentraliseerde plek waar kennis gebundeld is en in specifieke (sectorale)richtsnoeren kan worden omgezet[16] maar vooral een plek waar uitvoerders met concrete vraagstukken omtrent anonimiseren terecht kunnen.

Graag licht ik een en ander nog nader toe.

Hoogachtend,

Anna Berlee
Hoogleraar gegevensbescherming en privacyrecht aan de Open Universiteit

[15] In Nederland het Centraal Bureau voor Statistiek (CBS).
[16] In aanvulling op de algemenere aangekondigde wijziging van de Handleiding Wet hergebruik van overheidsinformatie.