Artikel

De strafvorderlijke normering van het geautomatiseerd overnemen van persoonsgegevens uit publiek toegankelijke bronnen met behulp van webcrawlers

Mr. R.J.A. Klaar*

1. Inleiding en plan van behandeling

Gegevens uit publiek toegankelijke bronnen, in het bijzonder het world wide web, zijn voor opsporingsdiensten in toenemende mate relevant voor het bestendigen en (verder) versterken van de informatiepositie. Een veelgebruikte term in dit verband is '*Open Source Intelligence*' (hierna: Osint).[1] Osint kan in een digitale context worden omschreven als methoden van onderzoek en (technische) hulpmiddelen waarmee in publiek toegankelijke (internet)bronnen, al dan niet op geautomatiseerde wijze, veelal online persoonsgegevens[2] kunnen worden verzameld die kunnen worden geanalyseerd teneinde bepaalde beslissingen te nemen of bepaalde handelingen te verrichten.[3]

Het kan gaan om het handmatig verzamelen van persoonsgegevens op het internet, het (stelselmatig) waarnemen van online gedrag van een persoon, het heimelijk onder dekmantel online interacteren met een persoon, maar ook om het *op geautomatiseerde wijze* vergaren van persoonsgegevens. Binnen de politiële context worden dergelijke online persoonsgegevens niet slechts gebruikt ter afbakening van het opsporingsonderzoek (start- en sturingsinformatie), maar steeds vaker ook als onderdeel van de (digitale) bewijsvoering.[4] Bovendien zullen gegevens uit publiek toegankelijke internetbronnen naar verwachting in de toekomst steeds vaker worden gebruikt voor het voorspellen van misdrijven.[5]

Publiek toegankelijke gegevens kunnen op verschillende manieren via geautomatiseerde systemen worden

* Mr. R.J.A. (Roel) Klaar is werkzaam als stafjurist/onderzoeker bij het Kenniscentrum Cybercrime voor de Rechtspraak. De auteur dankt mr. J.W. van den Hurk (als raadsheer verbonden aan het Gerechtshof Den Haag, en tevens als onderzoeker verbonden aan het Kenniscentrum Cybercrime voor de Rechtspraak), en Prof. mr. dr. L. Stevens, Hoogleraar straf- en strafprocesrecht aan de Vrije Universiteit Amsterdam, voor hun waardevolle commentaar bij eerdere versies van dit artikel.

1 De term Osint verwijst niet uitsluitend naar *digitale* publiek toegankelijke gegevens, maar ook naar *analoge* publiek toegankelijke gegevens, zoals nieuwsbladen.

2 De moderniseringswetgever heeft voor de uitleg van de term 'persoonsgegevens' aangesloten bij art. 4 lid 1 Algemene verordening gegevensbescherming (AVG). Deze definitie wordt als bekend verondersteld bij de lezer en noopt binnen het verband van dit artikel niet tot problematisering. Omwille van de leesbaarheid van dit artikel worden de termen 'persoonsgegevens' en 'gegevens' door elkaar gebruikt, nu duidelijk is dat telkens wordt verwezen naar 'persoonsgegevens'.

3 Deze omschrijving is ontleend aan www.computerweekly.com/tip/Using-open-source-intelligence-software-for-cybersecurity-intelligence.

4 Actuele voorbeelden zijn o.a.: Hof Amsterdam 9 november 2021, ECLI:NL:GHAMS:2021:3454 (Osint-rapport waarin wordt gerelateerd dat vanaf een Twitter-account vijf tweets waren verzonden waarin met een terroristisch misdrijf werd gedreigd. Het e-mailadres waarmee het Twitter-account was aangemaakt kon na een rechtshulpverzoek aan de e-mailprovider worden herleid tot verdachte) en Rb. Noord-Nederland 26 april 2021, ECLI:NL:RBNNE:2021:1583 (Osint-rapportage met informatie over een openbaar VK-profiel gebruikt als bewijs voor diverse uitingsdelicten).

5 Zie kritisch over het gebruik van voorspellingen als startinformatie in opsporingsonderzoeken bijvoorbeeld R.A. Hoving, 'Verdacht door een algoritme. Kan predictive policing leiden tot een redelijke verdenking?', *DD* 2019/41.

doi: 10.5553/PMSv/258950952022005003001

verzameld. Kenmerkend voor dergelijke geautomatiseerde systemen is dat zij automatisch, op basis van trefwoorden, informatie van het internet verzamelen, ordenen en rangschikken, ook wel *datamining* genoemd. Datamining maakt (stelselmatig) handmatig zoeken op internet overbodig.

Een voorbeeld is het sinds 2016 door de politie gebruikte geautomatiseerde zoeksysteem iColumbo.[6] Dergelijke geautomatiseerde zoeksystemen maken vaak gebruik van, al dan niet in overleg met de politie ontworpen, *webcrawlers* en *webscrapers*. De werking van deze software berust op algoritmen die bepalen op welke wijze zoekresultaten voor de gebruiker worden gegenereerd, geordend en gerangschikt.

Het huidige Wetboek van Strafvordering (Sv) bevat geen voorschrift(en) waarin expliciet wordt geregeld op welke manier publiek toegankelijke internetbronnen mogen worden geraadpleegd, en op welke manier het verzamelen en analyseren van daardoor verkregen informatie mag plaatsvinden.

Voor zover geen sprake is van een (stelselmatig) onderzoek dat meer dan een beperkte inbreuk maakt op de persoonlijke levenssfeer van de betrokkene, vormt de algemene taakstellende bevoegdheid van de politie (art. 3 Politiewet 2012 en art. 141 en 142 Sv) een toereikende wettelijke grondslag.[7] Maar voor zover dergelijk onderzoek tot een meer dan beperkte inbreuk op de persoonlijke levenssfeer leidt, ontbreekt vooralsnog een expliciet daarop toegesneden wettelijke grondslag. Algemeen wordt wel gesteld dat geautomatiseerd onderzoek aan publiek toegankelijke internetbronnen op de regeling van de stelselmatige observatie (art. 126g Sv) of de regeling van het stelselmatig inwinnen van informatie (art. 126j Sv) kan worden gebaseerd.[8] Deze constructies bieden echter slechts een tijdelijke noodoplossing. De moderniseringswetgever onderkent dit probleem en meent dat een afzonderlijke wettelijke grondslag voor geautomatiseerd onderzoek aan publiek toegankelijke internetbronnen noodzakelijk en gewenst is, omdat het verzamelen en analyseren van gegevens van de betrokkene die reeds aanwezig en beschikbaar zijn op het world wide web wezenlijk verschilt van het (al dan niet met behulp van een technisch hulpmiddel) rechtstreeks volgen of waarnemen van het gedrag van de betrokkene.[9] De ambtelijke versie van het Wetsvoorstel tot vaststelling van het nieuwe Wetboek van Strafvordering (juli 2020) (hierna: het wetsvoorstel Modernisering Sv) introduceert in artikel 2.8.8 Sv (nieuw) een nieuwe, bijzondere opsporingsbevoegdheid waaronder geautomatiseerd onderzoek aan publiek toegankelijke internetbronnen kan worden begrepen, namelijk het stelselmatig overnemen van persoonsgegevens uit publiek toegankelijke bronnen. Deze nieuwe bepaling houdt (voor zover hier relevant) in dat in geval van het *stelselmatig* overnemen van persoonsgegevens uit publiek toegankelijke bronnen een bevel van de officier van justitie is vereist.

Het centrale thema van dit artikel is de strafvorderlijke normering van de inzet van webcrawlers in de opsporing. De normering van de inzet van webscrapers blijft in dit artikel buiten beschouwing. De reden is praktisch van aard: hoewel denkbaar is dat webscrapers in het kader van de opsporing worden ingezet, zijn er vooralsnog geen aanwijzingen dat dit in de praktijk (al dan niet op beperkte schaal) plaatsvindt.[10] Getracht wordt de volgende onderzoeksvraag te beantwoorden:

> 'in hoeverre kunnen opsporingsautoriteiten op basis van de relevante gezichtspunten/factoren voor stelselmatigheid en bezien in het licht van de huidige technische mogelijkheden van een webcrawler, de mate van inbreuk op de persoonlijke levenssfeer inschatten en de aard en ernst van die inbreuk zoveel mogelijk beperken?'

Ter beantwoording van deze vraag worden kanttekeningen geplaatst bij de toepassing van het stelselmatigheidscriterium in relatie tot het geautomatiseerd overnemen van persoonsgegevens uit publiek toegankelijke internetbronnen. Geconcludeerd wordt dat deze kanttekeningen aanleiding vormen om nader te onderzoeken in hoeverre een verantwoorde inzet van webcrawlers in de opsporing in de (nabije) toekomst voldoende gewaarborgd is.

De opbouw van dit artikel is als volgt. In paragraaf 2 wordt de in het kader van de Modernisering Strafvordering voorgestelde strafvorderlijke bepaling van artikel 2.8.8 Sv (nieuw) besproken, waarin een regeling voor het stelselmatig overnemen van persoonsgegevens uit publiek toegankelijke bronnen wordt gegeven. Daarbij wordt de blik gericht op de betekenis die volgens de moderniseringswetgever aan 'stelselmatig overnemen' toekomt, en wordt in het bijzonder aandacht besteed aan

6 Zie algemeen over het iColumbo-platform M. Hildebrandt, 'Data-gestuurde intelligentie in het strafrecht', in: E.M.L. Moerel e.a., *Homo Digitalis* (Preadviezen NJV 2016-1), Deventer: Wolters Kluwer 2016, par. 2.3.2. Hildebrandt omschrijft iColumbo als: 'een platform gebouwd met als doel het analyseren van "open en big data" ten einde inzicht te verkrijgen in specifieke criminaliteitspatronen (…), dan wel het monitoren en profileren van een specifiek persoon of specifieke groep personen, in verband met gepleegde of mogelijk te plegen strafbare feiten'.

7 Zie uitvoerig M.J. Borgers, 'Normering van "lichte" opsporingshandelingen', *DD* 2015/15. Daarbij wordt aangenomen dat onderzoek aan publiek toegankelijke internetbronnen als zodanig *niet* zeer risicovol is voor de integriteit en beheersbaarheid van de opsporing.

8 Vgl. M.J. Borgers & T. Kooijmans, *Het Nederlands Strafprocesrecht*, Deventer: Wolters Kluwer 2021, p. 353.

9 Conceptmemorie van toelichting bij de Ambtelijke versie van het wetsvoorstel tot vaststelling van het nieuwe Wetboek van Strafvordering (juli 2020) (MvT), p. 496-497. Hoewel de moderniseringswetgever 'stelselmatige inwinning van informatie' niet benoemt, geldt ook daarvoor dat die opsporingsmethode gericht is op het (door een informant) rechtstreeks volgen en/of waarnemen van het gedrag van de betrokkene.

10 In bestuursrechtelijke context dient te worden vermeld dat de afdeling Handel & Toezicht van de Nederlandse Voedsel- en Warenautoriteit (NVWA) zich bezighoudt met de ontwikkeling van webscrapers om haar informatiepositie in het kader van het toezicht op internethandel, bijvoorbeeld op het gebied van dierenwelzijn, te verstevigen. Lijst van vragen en antwoorden aan de minister van Landbouw, Natuur en Voedselkwaliteit, over de brief van 23 december 2019 inzake Jaarplan 2020 van de NVWA, *Kamerstukken II* 2019/20, 33835, nr. 136, specifiek het antwoord op vraag 27. Vgl. ook deze blog op toezine.nl: www.toezine.nl/artikel/401/met-een-druk-op-de-knop-het-online-aanbod-in-kaart/ (26 januari 2021).

doi: 10.5553/PMSv/258950952022005003001

het problematische karakter van de termen 'publiek toegankelijke bron' en 'overnemen van gegevens'.
In paragraaf 3 wordt uitgelegd op welke wijze webcrawlers in technische zin functioneren, en welke technische aspecten van webcrawlers vanuit strafvorderlijk perspectief relevant zijn. In paragraaf 4 worden de door de wetgever in de memorie van toelichting bij het wetsvoorstel Modernisering Sv (MvT) genoemde relevante factoren ter invulling van het stelselmatigheidscriterium (de door de commissie-Koops voorgestelde factorencatalogus) toegepast op de inzet van webcrawlers, en wordt onderzocht in hoeverre die factoren de opsporingsautoriteiten in dat kader voldoende duidelijkheid bieden bij de vooraf uit te voeren stelselmatigheidstoets. In paragraaf 5 wordt dit artikel afgesloten met een overzicht van de belangrijkste bevindingen.

2. Het stelselmatig overnemen van persoonsgegevens uit publiek toegankelijke (internet)bronnen

In het eerder conceptwetsvoorstel tot vaststelling van Boek 2 van het nieuwe Wetboek van Strafvordering uit 2017 werd de strafvorderlijke bevoegdheid van het stelselmatig overnemen van persoonsgegevens uit publiek toegankelijke bronnen geregeld in artikel 2.8.2.4.1. De term 'open bron' in dit artikel is later – naar moet worden aangenomen mede naar aanleiding van aanbeveling 53 uit het rapport van de Commissie modernisering opsporingsonderzoek in het digitale tijdperk (hierna: commissie-Koops) – in het wetsvoorstel Modernisering Sv vervangen door de term 'publiek toegankelijke bron'.[11] Thans is het stelselmatig overnemen van persoonsgegevens geregeld in artikel 2.8.8 Sv (nieuw). Deze bepaling luidt als volgt:

> ' 1. In geval van verdenking van een misdrijf kan de officier van justitie bevelen dat een opsporingsambtenaar stelselmatig, al dan niet op geautomatiseerde wijze, persoonsgegevens uit publiek toegankelijke bronnen overneemt.
> 2. Het bevel wordt gegeven voor een periode van ten hoogste drie maanden. De geldigheidsduur kan telkens voor een periode van ten hoogste drie maanden worden verlengd.
> 3. Bij of krachtens algemene maatregel van bestuur worden regels gesteld over de geautomatiseerde wijze van overnemen van gegevens.'

De betekenis van diverse termen in deze bepaling is niet zonder meer duidelijk. Die termen worden hierna in cursief aangeduid. Wat moet worden verstaan onder '*publiek toegankelijke bron*'? Wanneer is sprake van het '*overnemen*' van persoonsgegevens, en waarin verschilt 'overnemen van persoonsgegevens' van '*onderzoek* van gegevens'? Wanneer wordt het overnemen van persoonsgegevens *stelselmatig*? In de subparagrafen 2.1-2.3 wordt verkend in hoeverre de wetgever de betekenis van de termen 'publiek toegankelijke bron', 'overnemen van gegevens' en 'stelselmatig' in de conceptmemorie van toelichting (MvT) verduidelijkt.

2.1 Publiek toegankelijke bron

De commissie-Koops heeft in haar rapport de term 'publiek toegankelijke bron' voorgesteld ter vervanging van de door de wetgever gebruikte term 'open bron'. Voor de betekenis van 'publieke toegankelijkheid' is aansluiting gezocht bij de strafbaarstelling van computervredebreuk (art. 138ab Wetboek van Strafrecht (Sr)), en meer in het bijzonder het bestanddeel 'binnendringen'. Publiek toegankelijke bronnen worden door de commissie gedefinieerd als 'alle bronnen die op een geautomatiseerd werk staan dat bereikbaar is vanaf het internet, die gepubliceerd of gedeeld zijn en *niet voorzien zijn van een beveiliging*' [cursivering door auteur].[12] De wetgever heeft aansluiting gezocht bij deze definitie.[13]
De wetgever benadrukt de prominente rol van 'de vraag of in enige mate een beveiliging wordt doorbroken', maar acht tevens relevant dat 'geen effectieve controle plaatsvindt bij het verstrekken van toegang', waarna (enigszins onbevredigend) wordt geconcludeerd: 'telkens zal moeten worden afgewogen in hoeverre een bron die beschikbaar is voor geautomatiseerde vergaring van gegevens ook daadwerkelijk publiek toegankelijk is'.[14] Naast de eis dat gegevens niet voorzien moeten zijn van een beveiliging, dient geen effectieve toegangscontrole plaats te vinden.[15] Door de definitie positief te omschrijven kan deze eenvoudiger worden verwoord: gegevens die niet zijn onderworpen aan effectieve toegangscontrole en waartoe toegang kan worden verkregen zonder het geautomatiseerd werk waarop deze zich bevinden binnen te dringen, zijn publiek toegankelijk. Hoewel deze definitie weinig aan de verbeelding over lijkt te laten kan bij nader inzien de vraag worden gesteld in hoeverre deze definitie houvast biedt voor de opsporingspraktijk. In de eerste plaats is de definitie negatief geformuleerd: van degene die geautomatiseerd persoonsgegevens wil overnemen, wordt gevraagd te beoordelen en onderbouwen dat en waarom een internetbron *niet* publiek toegankelijk is.
Van de opsporingsautoriteiten wordt immers verlangd dat zij onderbouwen dat en waarom gegevens *niet* zijn onderworpen aan effectieve toegangscontrole en dat *geen* doorbreking van enige beveiliging nodig is. In de tweede plaats lijkt de wetgever het perspectief van de opsporing en het perspectief van de gegevensbeheerder in de genoemde definitie te willen verenigen. Het per-

11 Commissie modernisering opsporingsonderzoek in het digitale tijdperk, *Regulering van opsporingsbevoegdheden in een digitale omgeving*, s.l. 2018 (hierna: commissie-Koops), p. 154-155.

12 Commissie-Koops, p. 153.
13 MvT, p. 498.
14 MvT, p. 498.
15 MvT, p. 498.

doi: 10.5553/PMSv/258950952022005003001

spectief van de opsporing is erop gericht dat geen gegevens worden verzameld voor zover daarvoor een beveiliging moet worden doorbroken. Het perspectief van de gegevensbeheerder is juist dat zolang effectieve toegangscontrole op persoonsgegevens in een internetbron plaatsvindt, deze persoonsgegevens niet geautomatiseerd mogen worden overgenomen. Voor een juist begrip van de term 'publiek toegankelijke bron' is inzicht vereist in de wijze waarop beide perspectieven elkaar complementeren. De definitie van de wetgever wordt hierna vanuit beide perspectieven belicht.

2.1.1 Het perspectief van de opsporingsautoriteiten: geen beveiliging doorbreken

Vanuit het perspectief van de opsporingsautoriteiten is met het oog op het geautomatiseerd overnemen van persoonsgegevens hoofdzakelijk van belang dat een internetbron geen beveiliging kent die moet worden doorbroken. Preciezer omschreven gaat het om de vraag of gegevens uit de internetbron zonder binnendringen in het geautomatiseerde werk – namelijk: de (web)server waar vanaf die internetbron wordt gehost – kunnen worden overgenomen. Het gaat bij de vraag of al dan niet een beveiliging wordt doorbroken aldus om de wijze van toegangsverschaffing, en niet om de online vindbaarheid van de gegevens. Ook gegevens die niet door zoekmachines (zoals Google) zijn geïndexeerd en door de internetgebruiker niet eenvoudig kunnen worden gevonden, zijn publiek toegankelijk. Waarschuwingen of (herhaaldelijke) oproepen tot registratie vormen nog geen beveiliging: de gebruiker kan ze negeren en daarmee wordt geen beveiliging doorbroken. Evenmin wordt bij registratie op een internetforum teneinde gegevens op dat forum zichtbaar te maken en te verzamelen een beveiliging doorbroken.[16] Dat wordt anders als opsporingsautoriteiten via een fakeaccount contact proberen te leggen met verdachte personen. Het aanmaken en vervolgens gebruiken van een fakeaccount, ongeacht of dat nu op het reguliere internet of op het darkweb gebeurt, valt onder het gebruik van een valse hoedanigheid als beschreven in artikel 138ab Sr.[17] Het gebruik van die valse hoedanigheid om persoonsgegevens te vergaren maakt dat de beveiliging van die gegevens wordt doorbroken.

Vanuit het perspectief van de opsporingsautoriteiten vormen afgeschermde internetbronnen een problematische categorie. Afgeschermde internetbronnen bevatten gegevens die zonder een (al dan niet tegen betaling verkregen) autorisatie niet toegankelijk zijn. Het geheel van afgeschermde internetbronnen wordt wel aangeduid als het *deep web*. Het gaat in de regel om gegevens die in een digitaal account of anderszins beveiligde omgeving zijn opgeslagen en slechts na invoer van een gebruikersnaam-wachtwoordcombinatie en/of meerfactorauthenticatie toegankelijk zijn. Voorbeelden van afgeschermde gegevens zijn:[18] gegevens die via een beveiligde omgeving worden ontsloten (zoals: een huisartsenportaal en registratiesystemen van de politie, maar ook Kluwer Navigator) en registergegevens waarvan slechts tegen betaling kennis kan worden genomen (zoals: KvK-gegevens en kadastrale gegevens). Vanuit het beveiligingscriterium beschouwd lijkt de term 'afgeschermde internetbron' te impliceren dat het per definitie gaat om beveiligde gegevens waarvan kennisneming door willekeurige derden niet mogelijk is zonder het doorbreken van die beveiliging, zodat die gegevens dus niet publiek toegankelijk zijn. Maar dat is te kort door de bocht, omdat geen rekening wordt gehouden met de mogelijkheid dat opsporingsautoriteiten een rechtmatig verkregen autorisatie gebruiken om de gegevens te raadplegen. Voor afgeschermde internetbronnen leidt dit tot de complicatie dat de publieke toegankelijkheid van de gegevens afhangt van het antwoord op de vraag of met gebruikmaking van een rechtmatig verkregen autorisatie toegang is verkregen tot de internetbron.

De opsporingsambtenaar die een account aanmaakt bij het Kadaster en daarmee inlogt om gegevens op te vragen, gebruikt een rechtmatig verkregen autorisatie. Maar de opsporingsambtenaar die een kwetsbaarheid in de website van het Kadaster misbruikt en inloggegevens verwerft waarmee toegang wordt verkregen tot de database, doet dat niet. Kortom: uitsluitend bij rechtmatig gebruik van een autorisatie kan sprake zijn van een publiek toegankelijke bron. Dat lijkt een betrekkelijk eenvoudig beoordelingscriterium, maar bij nader inzien is echter sprake van een afbakeningsprobleem omdat geen scherp onderscheid valt te maken tussen het rechtmatig gebruik van een autorisatie en – zoals de commissie-Koops stelt – irreguliere manieren van toegangsverschaffing.[19] Hoe moet bijvoorbeeld worden gedacht over een gedeelte van een bedrijfsnetwerk dat door kwetsbaarheden toch benaderbaar blijkt te zijn voor werknemers die daartoe niet bevoegd zijn? Is zo'n digitale werkomgeving publiek toegankelijk? Het beveiligingscriterium werpt geen licht op dit afbakeningsprobleem, omdat deze manier van irreguliere toegangsverschaffing – waarbij door een (min of meer toevallige) kwetsbaarheid toegang wordt verkregen tot gegevens die afgeschermd hadden moeten zijn – niet op één lijn kan worden gesteld met binnendringen in een geautomatiseerd werk.

16 Vgl. B.J. Koops & J.J. Oerlemans, 'Formeel Strafrecht en ICT' (hoofdstuk 3), in: B.J. Koops & J.J. Oerlemans, *Strafrecht en ICT* (Monografieën Recht en Informatietechnologie), Deventer: Wolters Kluwer 2019, p. 193, met verwijzing naar het rapport van de commissie-Koops.

17 Koops & Oerlemans 2019, p. 193.

18 In deze opsomming ontbreken digitale accounts (zoals: cloudaccounts en e-mailaccounts), omdat de moderniseringswetgever een nieuwe strafvorderlijke bevoegdheid tot inloggen met rechtmatig verkregen gegevens beoogt te introduceren. In het huidige wetsvoorstel Modernisering Sv is thans nog geen strafvorderlijke bepaling voor deze bevoegdheid ingevoegd. Een niet uitputtend overzicht van de huidige wettelijke en verdragsrechtelijke mogelijkheden om de inhoud van een digitaal account veilig te stellen is te vinden in Rb. Den Haag 14 mei 2021, ECLI:NL:RBDHA:2021:6770, *NJFS* 2021/268.

19 Commissie-Koops, p. 155. De vraag of sprake is van rechtmatig gebruik van een autorisatie dan wel een irreguliere manier van toegangsverschaffing kan zich ook in een andere context voordoen, namelijk bij de interpretatie van het bestanddeel 'binnendringen' als beschreven in art. 138ab Sr. Vgl. over deze problematiek HR 30 november 2021, ECLI:NL:HR:2021:1691 en de daaraan voorafgaande conclusie van A-G Spronken van 31 augustus 2021, ECLI:NL:PHR:2021:777.

doi: 10.5553/PMSv/258950952022005003001

Een vergelijkbaar afbakeningsprobleem is aan de orde als webcrawlers persoonsgegevens *crawlen* die afgeschermd hadden moeten blijven. Het crawlen van gegevens die afgeschermd hadden moeten blijven is veelal een onvoorzien neveneffect van de wijze waarop de webcrawler is geconfigureerd: irreguliere toegangsverschaffing zonder dat wordt binnengedrongen in een geautomatiseerd werk.[20] Het beveiligingscriterium biedt ook hier geen uitkomst. Afbakeningsproblemen waartoe irreguliere manieren van toegangsverschaffing leiden, kunnen beter worden benaderd vanuit het perspectief van de gegevensbeheerder en het criterium van effectieve toegang.

2.1.2 Het perspectief van de gegevensbeheerder: effectieve toegangscontrole

Vanuit het perspectief van de beheerder van een internetbron is met het oog op het geautomatiseerd overnemen van persoonsgegevens hoofdzakelijk van belang dat gegevens waarvan de toegang effectief wordt beperkt, niet worden gedeeld of gepubliceerd met willekeurige derden. Effectieve toegangscontrole[21] impliceert dat de gegevensbeheerder toegangsvoorwaarden stelt waaraan moet worden voldaan om daadwerkelijk toegang tot de gegevens te verkrijgen. Waarschuwingen of (herhaaldelijke) oproepen tot registratie vormen nog geen effectieve toegangscontrole: de gebruiker kan ze negeren en alsnog toegang verkrijgen tot de gegevens. Ook een registratievoorwaarde op een website, zoals een internetforum, behelst niet zonder meer effectieve toegangscontrole. Daarvan is slechts sprake voor zover een registratie-eis de toegang tot gegevens op een website daadwerkelijk beperkt. De registratievoorwaarde op de website nu.nl is geen effectieve toegangscontrole, omdat de gebruiker ook zonder registratie nieuwsberichten kan blijven lezen. Dat bepaalde functionaliteiten, waaronder het lezen van meer dan vijf reacties onder berichten in de rubriek 'Nujij',[22] zonder registratie niet of minder beschikbaar zijn, belemmert evenmin de toegang tot de nieuwsberichten zelf.

Gegevens die zich op het darkweb bevinden, zijn evenmin onderworpen aan effectieve toegangscontrole. Voor het verkrijgen van toegang tot het Tor-netwerk, dat een onderdeel (darknet) van het darkweb is, is weliswaar speciale software (Tor-client) en een bijzondere netwerkconfiguratie vereist, maar dat zijn toegangsbeperkingen met betrekking tot het Tor-netwerk zelf en niet voor de (web)server in het Tor-netwerk en de gegevens die zich daarop bevinden. Dat wordt anders zodra de gegevens slechts kunnen worden geraadpleegd door een beperkte kring van gebruikers. Gebruikers van darkwebfora dienen niet zelden eerst een bepaald sociaal krediet op te bouwen, voordat toegang wordt verkregen tot meest illegale goederen (zoals de schadelijkste malware, de beste kwaliteit harddrugs, enz.). Dergelijke (transactie)gegevens zijn wel onderworpen aan effectieve toegangscontrole en niet publiek toegankelijk.

Zoals hiervoor al aangegeven bestaat er een grijs gebied van methoden van toegangsverschaffing waarbij geen beveiliging wordt doorbroken, maar wel zonder rechtmatig verkregen autorisatie kennis wordt genomen van (al dan niet gedeeltelijk) afgeschermde gegevens. Dat probleem is bijvoorbeeld aan de orde als webcrawlers persoonsgegevens crawlen of scrapen die afgeschermd hadden moeten blijven. Talloze bedrijven maken voor commerciële en/of *human resource*-doeleinden gebruik van webcrawlers of om op grote schaal persoonsgegevens (profielnaam, functie en cv) te verzamelen. Voor socialemediaplatforms, zoals LinkedIn, en gebruikers van openbare profielen, is niet volledig controleerbaar welke persoonsgegevens door webcrawlers worden verzameld, tenzij de gebruiker ervoor kiest het profiel volledig af te schermen. Volledig afschermen leidt er echter toe dat een profiel niet alleen voor willekeurige derden, maar ook voor andere LinkedIn-gebruikers niet meer zichtbaar is. Veruit het merendeel van de gebruikers tracht om die reden een balans te vinden tussen de vindbaarheid van persoonsgegevens, en het respecteren van de privacyinstellingen. Maar ook via de privacyinstellingen van een LinkedIn-account kan niet volledig controle worden uitgeoefend over welke persoonsgegevens zichtbaar zijn. Als persoonsgegevens onderdeel uitmaken van een LinkedIn-profiel, bestaat aldus het risico dat die persoonsgegevens door webcrawlers of geautomatiseerd worden overgenomen. Daarbij wordt weliswaar niet binnengedrongen in geautomatiseerde werken (servers van LinkedIn), maar er worden wel zonder dat LinkedIn en de betreffende gebruikers daarvoor expliciet toestemming hebben verleend op grote schaal persoonsgegevens overgenomen. Vanuit het perspectief van de gegevensbeheerder en het criterium van effectieve toegangscontrole is dat problematisch, omdat gegevens worden overgenomen terwijl inspanningen zijn verricht om die gegevens af te schermen. Derhalve kan de vraag worden gesteld of openbare LinkedIn-profielen bij het grootschalig gebruik van webcrawlers of vergelijkbare 'grijze' methoden van toegangsverschaffing nog steeds als publiek toegankelijke bronnen kunnen worden beschouwd. Een vergelijkbare problematiek betreft het onbedoeld beschikbaar komen van aanvankelijk afgeschermde gegevens door een datalek. Een voorbeeld is het datalek bij de GGD, waardoor persoonsgegevens van willekeurige Nederlanders via Telegram konden worden verhandeld.[23] Gelden die persoonsgegevens wanneer zij gecrawld worden als publiek toegankelijk? Het perspectief van de gegevensbeheerder en het criterium van effectieve toegangscontrole nopen ertoe de realiteit dat internetgebruikers geen volledige controle kunnen uit-

20 Zie over de configuratie van webcrawlers uitvoeriger par. 3.1-3.3.

21 De term 'effectieve toegangscontrole' is afkomstig uit het rapport van de commissie-Koops, zie commissie-Koops, p. 153. De wetgever spreekt over 'toegangscontrole met een minimale beveiliging', zie MvT, p. 499. De laatstgenoemde term verdient mijns inziens geen navolging, omdat daarbij uit het oog dreigt te worden verloren dat de definitie van 'publiek toegankelijke bron' zowel wordt beïnvloed door het opsporingsperspectief als het perspectief van de gegevensbeheerder.

22 Zie: 'Waarom we je vragen om in te loggen', NU.nl (geraadpleegd op 22 februari 2022).

23 Zie 'Datalek bij GGD: gegevens van miljoenen Nederlanders in criminele handen', *de Volkskrant* 26 januari 2021.

oefenen over persoonsgegevens die zij in afgeschermde internetomgevingen delen of publiceren onder ogen te zien en – indien van toepassing – te betrekken bij de beoordeling of gegevens al dan niet publiek toegankelijk zijn.

2.2 Stelselmatig overnemen van persoonsgegevens

De moderniseringswetgever heeft in de memorie van toelichting op titel 7.3 van Boek 2 de betekenis van het 'overnemen' van persoonsgegevens verduidelijkt.[24] Overnemen betreft het kopiëren van gegevens die al zijn opgeslagen op een geautomatiseerd werk dat bereikbaar is vanaf het internet en die gepubliceerd of gedeeld zijn. Van belang is dat het enkel kennisnemen van persoonsgegevens uit publiek toegankelijke bronnen er niet onder valt. De ratio hiervan is dat bij enkel kennisnemen door een persoon niet achteraf (zoals bij overnemen wel het geval is) een volledige reproductie van de waargenomen informatie mogelijk is en er navenant minder risico's zijn voor de verspreiding van de gegevens. Het overnemen van gegevens verschilt van het vastleggen van gegevens, omdat het bij het vastleggen van gegevens gaat om het registreren van signalen die niet (per definitie) al elders zijn opgeslagen, terwijl dat bij het overnemen van gegevens wel altijd het geval is.[25] Onder 'onderzoek van gegevens' vallen alle handelingen die moeten worden verricht om gegevens over te nemen of daarvan kennis te nemen. Het onderzoek van gegevens omvat daardoor zowel het kennisnemen van gegevens als het overnemen van gegevens.[26] Tot zover de voor het geautomatiseerd overnemen van persoonsgegevens belangrijkste wettelijke definities.

Om de lezer een globale indruk te geven van de relevantie van de wettelijke definities van 'overnemen van gegevens' en 'onderzoek van gegevens' voor het gebruik van webcrawlers in de opsporing, wordt daarvan nu een korte beschrijving gegeven. Als wordt gesproken over 'overnemen van gegevens' wordt daarmee enkel gedoeld op het crawlingproces als zodanig. Kort samengevat betreft het crawlingproces een zich oneindig herhalend proces waarin voortdurend publiek toegankelijke gegevens worden overgenomen. Als wordt gesproken over 'onderzoek van gegevens' wordt daarmee gedoeld op het geheel van het crawlingproces en de (al dan niet geautomatiseerde) verwerking van de in het crawlingproces verzamelde gegevens. In paragraaf 3 worden de wettelijke definities vervolgens gedetailleerder in verband gebracht met het webcrawlingproces en de werking van de verschillende onderdelen van een webcrawler.

2.3 Het stelselmatigheidscriterium in relatie tot het geautomatiseerd overnemen van persoonsgegevens

Het sleutelbegrip in artikel 2.8.8 Sv (nieuw) is het stelselmatigheidscriterium: de uitoefening van een bevoegdheid is stelselmatig als daarbij op voorhand redelijkerwijs voorzienbaar een min of meer volledig beeld van bepaalde aspecten van iemands privéleven kan ontstaan.[27] Diverse auteurs hebben betoogd dat de wetgever in de MvT weinig aanknopingspunten aanreikt voor de beoordeling wanneer bij het stelselmatig overnemen van persoonsgegevens uit publiek toegankelijke bronnen sprake is van stelselmatig onderzoek.[28] De commissie-Koops deed destijds de aanbeveling dat de wetgever in de MvT expliciet zou moeten opnemen dat de opsporingsautoriteit vooraf een inschatting moet maken en vastleggen over de te verwachten inbreuk op de persoonlijke levenssfeer van de verdachte, en dat de wetgever tevens handvatten zou moeten bieden over de wijze waarop die inschatting gemaakt kan worden (aanbeveling 57).[29] Tevens deed de commissie-Koops met betrekking tot de inzet van webcrawlers de specifieke aanbeveling dat in de MvT zou moeten worden ingegaan op de voorzienbaarheid van de privacyinbreuk, en de proportionaliteit en subsidiariteit, in relatie tot de wijze van inrichting van de webcrawler (onder meer: de duur van de inzet, het zoekbereik en andere aangebrachte beperkingen) (aanbeveling 58).[30]

Ten slotte heeft de commissie-Koops de aanbeveling gedaan dat in de MvT meer expliciet handvatten zouden moeten worden geboden voor wat betreft de factoren op basis waarvan het stelselmatigheidscriterium, dat voorafgaand aan het stelselmatig overnemen van persoonsgegevens zou moeten worden toegepast, moet worden ingevuld (aanbeveling 59).[31] In dit verband heeft de commissie-Koops vier categorieën van relevante factoren genoemd: de omvang en het type van de (over te nemen) gegevens, de aard van de bron, de wijze van zoeken en het gebruik van gegevens en de mogelijke impact op de persoon.[32] De wetgever heeft de aanbevelingen van de commissie-Koops in de MvT bij artikel 2.8.8. Sv (nieuw) grotendeels gevolgd en de voorgestelde vier categorieën opgenomen.[33] Ook heeft de wetgever toegelicht op welke manier die factoren in de door de opsporingsautoriteiten op voorhand te maken afweging zouden moeten worden betrokken.

De operationalisering van het stelselmatigheidscriterium binnen de context van de inzet van webcrawlers stelt de opsporing voor bijzondere uitdagingen. Benadrukt moet worden dat deze bijzondere uitdagingen specifiek aan de orde zijn binnen de context van de inzet van webcrawlers.

24 MvT, p. 245, p. 497.
25 MvT, p. 245.
26 MvT, p. 245.
27 Commissie-Koops, p. 38.
28 Zie meest recent R.S. Veen, 'Digitale opsporing. Het EHRM en het stelselmatige overnemen van persoonsgegevens uit publiek toegankelijke bronnen', *DD* 2019/29; W.Ph. Stol & L. Strikwerda, 'Online vergaren van informatie voor opsporingsonderzoek. Een beknopte evaluatie van voorgestelde wetgeving', *Tijdschrift voor Veiligheid* 2018, afl. 1-2, p. 8-22; J.J. Oerlemans, *Investigating cybercrime* (diss. Leiden), Amsterdam: University Press 2017, p. 161-163; A.R. Lodder & M.B. Schuilenburg, 'Politie-webcrawlers en Predictive policing', *Computerrecht* 2016/81.
29 Commissie-Koops, p. 160.
30 Commissie-Koops, p. 162.
31 Commissie-Koops, p. 165.
32 Commissie-Koops, p. 163-164.
33 MvT, p. 500-502.

doi: 10.5553/PMSv/258950952022005003001

De reden is dat de inzet van een webcrawler *in theorie* op verschillende momenten binnen het 'onderzoek van gegevens' (het webcrawlingproces plus de verwerking van in het crawlingproces verzamelde gegevens) aan het stelselmatigheidscriterium kan worden getoetst, namelijk bij de configuratie van een webcrawler, bij de beoordeling van de relevantie van de gecrawlde gegevens en bij het onderzoeken van de gecrawlde gegevens. De vraag kan echter worden gesteld of toetsing aan het stelselmatigheidscriterium in deze stadia van het onderzoek ook steeds *praktisch* mogelijk en haalbaar is voor de opsporingspraktijk. Om de vraag te kunnen beantwoorden in hoeverre de door de wetgever genoemde factoren en de toelichting daarop de opsporingsautoriteiten bij het uitvoeren van de stelselmatigheidstoets in het kader van de inzet van webcrawlers voldoende duidelijkheid bieden, is een goed begrip vereist van wat webcrawlers precies zijn en op welke wijze het webcrawlingproces verloopt. In paragraaf 3 wordt daarom aandacht besteed aan de strafvorderlijk relevante onderdelen van webcrawlers en aan het webcrawlingproces. In paragraaf 4 wordt vervolgens getracht de vraag te beantwoorden in hoeverre de door de wetgever genoemde factoren en de toelichting daarop de opsporingsautoriteiten bij het uitvoeren van de stelselmatigheidstoets in het kader van de inzet van webcrawlers voldoende duidelijkheid bieden.

3. De strafvorderlijk relevante technische aspecten van webcrawlers

Webcrawlers zijn softwareprogramma's die op geautomatiseerde wijze op internet naar digitale gegevens zoeken, zoals NAW-gegevens, telefoonnummers, e-mailadressen of accountnamen, en de gevonden zoekresultaten voor de gebruiker genereren en ordenen. Webcrawlers zijn een geïntegreerd onderdeel van online zoekmachines. Een bekende webcrawler is 'Googlebot', dat is de generieke naam van de in de Google zoekmachine gebruikte webcrawler.[34] Webcrawlers kunnen voor uiteenlopende legitieme doelen gegevens verzamelen en ordenen, bijvoorbeeld voor het vergelijken van online adviesprijzen van producten, het verzamelen van gegevens over de bezoekers van een bepaalde webpagina en het zoeken naar en aanvoeren van relevante nieuwsberichten.

Webcrawlers kunnen ook ten behoeve van de opsporing worden ingezet om publiek toegankelijke informatie te verzamelen. Opsporingsautoriteiten zijn in dit verband hoofdzakelijk geïnteresseerd in metadata: een paraplubegrip waarmee gegevens die de eigenschappen van bepaalde gegevens beschrijven worden aangeduid. Metadata omvatten zeer uiteenlopende categorieën van gegevens, zoals metadata van Microsoft Office-documenten (hierna: MS Office-documenten). In deze categorie vallen bijvoorbeeld de naam van de auteur, de datum en het tijdstip waarop dat bestand voor het eerst op de computer is opgeslagen ('File Created Date') en het aantal pagina's. De gebruiker kan deze bestandsmetadata inzien door de bestandseigenschappen van het MS Office-document te raadplegen. Bij metadata van een website moet aan andersoortige gegevens worden gedacht, zoals de titels van de webpagina's van een website en de metaomschrijvingen van de hoofdpagina. Een webpaginatitel bestaat uit de site- en domeinnaam (www.sitenaam.domeinnaam). De metaomschrijving van de hoofdpagina bestaat uit een korte omschrijving van de website. De webpaginatitel en bijbehorende metaomschrijving worden onder meer weergegeven in de zoekresultaten die na een ingegeven zoekopdracht in een online zoekmachine, zoals de Google zoekmachine, voor de gebruiker zichtbaar worden.

Webcrawlers bestaan uit onderdelen met verschillende functionaliteiten, die vanuit strafvorderlijk perspectief bezien niet allemaal relevant zijn. In deze paragraaf worden deze onderdelen naar chronologie van het webcrawlingproces besproken. Voor de softwarematige onderdelen van een webcrawler bestaan geen eenduidige benamingen. Bovendien kan de precieze softwarearchitectuur per type webcrawler, voor zover al inzichtelijk, enigszins verschillen.[35] Ik sluit om die reden aan bij de algemene softwarearchitectuur zoals gepresenteerd in het informaticaboek van Olsten en Najork.[36]

Zij onderscheiden de volgende zeven webcrawleronderdelen: de 'frontier data structure', de 'HTTP-fetcher', de 'link extractor', de 'URL distributor', de 'custom URL filter', de 'duplicate URL eliminator' en de 'URL prioritizer'. Ik meen dat het laatstgenoemde onderdeel vanuit strafvorderlijk perspectief het meest relevante onderdeel van een webcrawler is, zodat die functionaliteit hierna uitvoeriger wordt besproken. Voor de helderheid: de activiteit 'webcrawling' is technisch gezien een zich oneindig herhalend zoek- en verwerkingsproces. Het crawlingproces is, nadat het is doorlopen, niet afgerond (tenzij een webcrawler zodanig is ingesteld dat een zoekopdracht slechts eenmalig wordt uitgevoerd). Het 'webcrawlingproces' betreft dus in principe een doorlopende toepassing van de onderdelen die in deze paragraaf worden beschreven.

34 Meer informatie is hier te vinden: 'What Is Googlebot', Google Search Central, Google Developers (geraadpleegd op 22 januari 2022).

35 Daarbij kan meespelen dat softwareontwikkelaars, zoals de ontwikkelaars van de Googlebot, een belang hebben om de softwarearchitectuur van eigen webcrawlers af te schermen van concurrenten.

36 C. Olsten & M. Najork, 'Web Crawling', *Foundations and Trends in Information Retrieval* 2010, nr. 3, p. 175-246, i.h.b. p. 184-185. Om auteursrechtelijke redenen wordt de grafische weergave van de basissoftwarearchitectuur van webcrawlers uit dit boek niet getoond, maar deze kan via Google Scholar hier wel worden geraadpleegd.

doi: 10.5553/PMSv/258950952022005003001

3.1 Frontier data structure, link extractor, link distributor en duplicate URL eliminator

De *frontier data structure*[37] betreft een database met nog niet in het webcrawlingproces betrokken hyperlinks/URL's[38] die worden ingeladen in het webcrawlingproces. Deze database bevat een verzameling van hyperlinks waarop de webcrawler wordt toegepast. Deze database vormt het fundament van het webcrawlingproces. Afhankelijk van de configuratie van de *frontier data structure* is het mogelijk om hyperlinks volgens bepaalde prioriteitscriteria, zoals doorzoekbaarheid, te rangschikken.[39] De prioritering van hyperlinks bepaalt de volgorde waarin hyperlinks/URL's worden gecrawld en op die manier de aard van de zoekopdracht. Weer anders gezegd bepaalt de prioritering de volgorde van de wachtrij (queue) van hyperlinks/URL's die worden betrokken in het webcrawlingproces.

Een strafvorderlijk zeer relevant technisch aspect is dat de *frontier data structure* zelf geen hyperlinks/URL's kan selecteren. Degene die de webcrawler configureert (lees: de opsporingsautoriteit) bepaalt welke hyperlinks/URL's worden ingeladen.

De absolute reikwijdte van het sleepnet van een 'webcrawler' wordt derhalve bepaald door de opsporingsautoriteit die de webcrawler inzet. Dit aspect is van groot belang voor de beoordeling van de stelselmatigheid op het moment van het configureren van een webcrawler. De reikwijdte van het sleepnet kan op verschillende manieren worden beperkt, bijvoorbeeld tot een bepaald internetdomein (.nl of .eu) of meer specifiek tot bepaalde (sub)mappen van een webserver (http://www.voorbeeld.nl/map1/submap1, enz.). Naarmate minder of geen beperkingen worden gesteld aan de absolute reikwijdte van het sleepnet, zullen de aard en ernst van de inbreuk op de persoonlijke levenssfeer in beginsel groter zijn. Het grondig en doordacht configureren van de *frontier data structure* biedt opsporingsautoriteiten een belangrijke mogelijkheid om de aard en ernst van de inbreuk op de persoonlijke levenssfeer, bezien in het licht van de aard en ernst van de verdenking, op voorhand zo veel mogelijk te beperken.

Een ander onderdeel, de *link extractor*, faciliteert dat de HTML-code van de webpagina en de hyperlink(s) op de webpagina veilig worden gesteld. Kort gezegd betreft HTML-code het deel van een website dat verantwoordelijk is voor de inhoud ervan. De *link extractor* filtert snelkoppelingen (Engels: hyperlinks) uit deze HTML-code en heeft daardoor een vitale, operationele functie in een webcrawler.[40]

Weer een ander onderdeel, de *URL distributor*, faciliteert dat de door de *link extractor* veiliggestelde hyperlinks in het webcrawlingproces worden betrokken. Dit onderdeel is ervoor verantwoordelijk dat de *frontier data structure* gedurende het webcrawlingproces wordt aangevuld met hyperlinks/URL's die door de *link extractor* zijn verkregen.

Ten slotte twee onderdelen met louter technische relevantie, namelijk de *frontier manager* en *duplicate URL eliminator*. De *frontier manager* (ook wel: 'crawl frontier API' genoemd) betreft een onderdeel met louter technische relevantie, en is verantwoordelijk voor de communicatie tussen de webcrawler en de database met hyperlinks/URL's. Dit onderdeel zorgt er (kort gezegd) voor dat het webcrawlingproces conform de configuratie van de webcrawler wordt uitgevoerd. De *duplicate URL eliminator* controleert (kort gezegd) of dezelfde hyperlink/URL niet twee keer in de *frontier data structure* wordt opgenomen. Doublures doen immers afbreuk aan de efficientie van het crawlingproces, en leiden ertoe dat meer bandbreedte wordt verbruikt dan strikt noodzakelijk is.

3.2 Http-fetcher

De *http-fetcher* faciliteert (kort gezegd) dat de webcrawler verbinding maakt met de hyperlinks uit de *Frontier data structure*. Met dit onderdeel wordt getracht de inhoud van de achterliggende webpagina's of websites te downloaden.[41] Het onderdeel wordt na het inladen van de *frontier data structure* geactiveerd, en is onmisbaar voor het starten van het webcrawlingproces.

De capaciteit voor het downloaden van webpagina's en het aantal crawlingprocessen dat kan worden doorlopen wordt beperkt door de bandbreedte van de internetverbinding van de gebruiker van de webcrawler.[42] De gedownloade webpagina's of websites ('harvested web pages') zijn meestal niet zichtbaar in de webcrawlerapplicatie zoals die voor de gebruiker wordt weergegeven,[43] maar webcrawlers kunnen wel ingericht zijn voor 'web archiving': het archiveren van complete websites of delen daarvan. Een bekend webarchief is de 'Wayback

37 Op Wikipedia wordt de 'frontier data structure' aangeduid als de 'crawl frontier'. Zie het lemma 'crawl frontier' op Wikipedia. Voor meer technische informatie over dit onderdeel kan het basisdocument 'crawlfrontier documentation release 0.2.0 (versie 15 april 2015)' van de IT-ontwikkelaar 'ScrapingHub' worden geraadpleegd. Bij de crawl frontier hoort een aparte basissoftwarearchitectuur, maar deze wordt in dit artikel niet besproken. Zie daarover bijvoorbeeld https://en.wikipedia.org/wiki/Crawl_frontier.

38 Een hyperlink of URL is een verwijzing van de ene webpagina of website naar een andere webpagina of website. Beide termen zijn synoniemen, maar worden in het vervolg omwille van herkenbaarheid naast elkaar gebruikt.

39 Het aanbrengen van een rangschikking kan ook door een ander onderdeel van de webcrawler plaatsvinden, namelijk de *URL prioritizer*, die hierna in par. 3.3 wordt besproken. Dit onderdeel brengt op basis van door de gebruiker (lees: opsporingsautoriteit) opgegeven voorkeuren een rangorde aan in de hyperlinks die in de *frontier data structure* zijn opgenomen.

40 Hyperlink extractietools worden als zelfstandige functionaliteit (los van een webcrawler infrastructuur) online vrijelijk aan internetgebruikers beschikbaar gesteld. Zie bijvoorbeeld https://urlextractor.net.

41 Webprogrammeurs kunnen door het gebruik van een 'robots.txt'-bestand het dataverkeer dat webcrawlers genereren beperken. In een 'robots.txt'-bestand kunnen links naar webpagina's en/of bestanden op de betreffende website worden opgenomen die door webcrawlers niet mogen worden overgenomen. Meer informatie over het blokkeren van webcrawlers biedt deze uitleg door Google: https://developers.google.com/search/docs/advanced/robots/intro.

42 Vgl. J. Edwards, K. McCurley & J. Tomlin, 'An Adaptive Model for Optimizing Performance of an Incremental Web Crawler', in: *Proceedings of the 10th international conference on World Wide Web* 2001, p. 106-113, i.h.b. p. 106: 'Given that the bandwith for downloading crawls is neither infinite nor free it is becoming essential to crawl the web in an not only scalable, but *efficient* way if some reasonable measure of quality or freshness is to be maintained.'

43 Olston & Najork 2010, p. 184.

doi: 10.5553/PMSv/258950952022005003001

Machine',[44] dat gearchiveerde versies van een enorme hoeveelheid webpagina's en websites bevat. In december 2020 bevatte dit webarchief in totaal 70 petabyte[45] aan internetgegevens, waaronder ongeveer 525 miljard webpagina's.

Voor de beoordeling van de stelselmatigheid van de inzet van een webcrawler op het moment van het configureren daarvan heeft de *http-fetcher* geen betekenis. De *http-fetcher* heeft namelijk geen sturende rol in het webcrawlingproces. Maar voor zover een webcrawler het archiveren van websites faciliteert, wordt daarmee wel de controle van de gecrawlde gegevens vergemakkelijkt. De configuratie van de *http-fetcher* is, voor zover deze archivering faciliteert, mogelijk dus relevant voor het beoordelen van de stelselmatigheid op het moment van het onderzoeken van de gecrawlde gegevens. Stel dat achteraf blijkt dat (ondanks een zorgvuldige selectie van de hyperlinks/URL's die werden ingeladen in het webcrawlingproces) door inzet van een webcrawler de Instagramprofielen van alle contactpersonen van de verdachte zijn overgenomen, terwijl de verdenking enkel zag op online belaging van diens partner, dan lijken de aard en de ernst van de privacyinbreuk, bezien in het licht van de aard en ernst van de verdenking, niet gerechtvaardigd te kunnen worden.

3.3 URL-prioritizer en custom URL filter

Hoewel de technische werking van de *URL-prioritizer* en *custom URL filter* lastig van elkaar kunnen worden onderscheiden, is de functie van beide onderdelen wezenlijk verschillend. De *custom URL filter* is ervoor verantwoordelijk dat op basis van technische filtercriteria bepaalde hyperlinks worden uitgesloten van het webcrawlingproces. Van belang is dat de *custom URL filter* webpagina's louter op basis van technische criteria (en dus niet inhoudelijke selectiecriteria) uitsluit, bijvoorbeeld omdat op een webpagina bestandstypen voorkomen die niet bruikbaar zijn voor de gebruiker.

In zoverre verschilt de functie van de *custom URL Filter* van de hierna te bespreken *URL prioritizer*, die op basis van door de gebruiker opgegeven voorkeuren bepaalde hyperlinks voorrang geeft in het webcrawlingproces. Daarmee kan het crawlingproces namelijk wel inhoudelijk worden gestuurd.

De *URL prioritizer* brengt op basis van door de gebruiker opgegeven voorkeuren een bepaalde rangorde aan in de hyperlinks/URL's in de *frontier data structure*.[46] Zoals eerder besproken kan, wanneer de *frontier data structure* (mede) uit een extern ingeladen database van hyperlinks bestaat, voor (dat deel van) de database reeds een rangorde zijn aangebracht, zodat de *URL prioritizer* die is ingebouwd in de webcrawler dan voor (dat deel van) de database geen functie vervult.

De *URL prioritizer* kan in strafvorderlijke zin als het meest relevante onderdeel van een webcrawler worden beschouwd, omdat dit onderdeel (mede) bepalend is[47] voor de selectie van de hyperlinks die met voorrang in het webcrawlingproces worden betrokken en door de webcrawler worden verwerkt.

Naarmate de gegevensoverdrachtsnelheid (bandbreedte) van de internetverbinding die voor het webcrawlingproces beschikbaar is beperkter is, zullen de instellingen van de *URL prioritizer* relevanter zijn voor het functioneren van de webcrawler. Het belang van de rangorde van de hyperlinks/URL's zal immers toenemen naarmate de webcrawler door een trage(re) internetverbinding minder snel kan opereren, en de kans bestaat dat hyperlinks/URL's met een lagere prioriteit niet in het webcrawlingproces kunnen worden meegenomen.

De voorkeuren op basis waarvan de *URL prioritizer* opereert, worden primair bepaald door het doel waarvoor de gebruiker de webcrawler hanteert. Voor zover het gaat om door de politie ontwikkelde webcrawlers zullen de gehanteerde voorkeursinstellingen in de *URL prioritizer* waarschijnlijk worden bepaald door de opsporingsautoriteiten. Deze voorkeursinstellingen zullen bepalend zijn voor de relatieve reikwijdte van het 'sleepnet' van de webcrawler. Het gaat om relatieve reikwijdte, omdat de aard en de omvang van gegevens die worden overgenomen mede afhankelijk zijn van de inrichting en indexering van de webpagina's waarvan gegevens worden overgenomen en omdat de absolute reikwijdte van het sleepnet wordt bepaald door de hyperlinks/URL's die in de *frontier data structure* van de webcrawler worden ingeladen.

De voorkeurinstellingen van de *URL prioritizer* beïnvloeden, samen met de hyperlinks/URL's die worden ingeladen in de *frontier data structure*, de invulling van het stelselmatigheidscriterium op het moment van het configureren van een webcrawler. Naarmate de voorkeursinstellingen van de *URL prioritizer* ertoe leiden dat de webcrawler meer privacygevoelige informatie van de verdachte verwerkt, zal, afhankelijk van de overige factoren die in het concrete geval van invloed zijn op de invulling van het stelselmatigheidscriterium, eerder sprake zijn van het stelselmatig geautomatiseerd overnemen van persoonsgegevens. Omgekeerd zal naarmate de voorkeursinstellingen zodanig zijn dat (zeer) beperkt privacygevoelige gegevens door de webcrawler worden verwerkt, afhankelijk van de overige factoren die in het concrete geval van invloed zijn op de invulling van het stelselmatigheidscriterium, minder snel sprake kunnen zijn van het stelselmatig geautomatiseerd overnemen van persoonsgegevens. Wel moet onderkend worden dat

44 Zie https://en.wikipedia.org/wiki/Wayback_Machine. Deze database is opgericht door de non-profitorganisatie 'The Internet Archive'.

45 1 Petabyte is 1000 terabyte, 1 terabyte is 1000 gigabyte.

46 Olsten & Najork 2010, p. 184: '(…) the URL prioritizer selects a position for the URL in the Frontier, based on factors such as estimated page importance or rate of change.'

47 De rangorde van hyperlinks in de *frontier data structure* wordt ook bepaald door de *custom URL filter*, maar zoals hiervoor reeds besproken gaat het bij die functionaliteit om een selectie op basis van technische, operationele voorkeuren, die primair door de softwareprogrammeur op basis van bepaalde randvoorwaarden worden bepaald, en niet (primair) om voorkeuren die door de gebruiker van de webcrawler zijn opgegeven. Bovendien is voorstelbaar dat de gebruiker niet of slechts gedeeltelijk bekend is met de technische voorkeuren op basis waarvan de *custom URL filter* opereert.

de hiervoor genoemde uitgangspunten de opsporingsautoriteiten in concrete gevallen lang niet altijd voldoende houvast zullen kunnen bieden bij het beoordelen van de stelselmatigheid op het moment van configureren van een webcrawler. Bovendien moet er rekening mee worden gehouden dat de uitkomst van het webcrawlingproces niet uitsluitend door de voorkeurinstellingen van een webcrawler wordt bepaald, maar tevens door de inrichting en indexering van een website of webpagina. De uitwerking van de voorkeurinstellingen van een webcrawler op een bepaalde website of webpagina valt op voorhand redelijkerwijs niet volledig te voorzien, en dat kan leiden tot bijvangst: zoekresultaten die vooraf redelijkerwijs niet voorzienbaar waren.

3.4 Tussenconclusie: drie aangrijpingspunten voor het beoordelen van de stelselmatigheid

Uit de hiervoor gegeven analyse van de verschillende onderdelen van een webcrawler kwam naar voren dat de *frontier data structure*, *http-fetcher* en *URL Prioritzer* met het oog op het beoordelen van de stelselmatigheid van het geautomatiseerd overnemen van persoonsgegevens met behulp van webcrawlers als kernonderdelen kunnen worden beschouwd. Het meest wezenlijke onderdeel is de *URL prioritizer*, omdat dit onderdeel op basis van door de gebruiker opgegeven voorkeuren een bepaalde rangorde aanbrengt in de hyperlinks/URL's in de *frontier data structure*.

De voorkeurinstellingen van de *URL prioritizer* beïnvloeden, samen met de hyperlinks/URL's die worden ingeladen in de *frontier data structure*, de invulling van het stelselmatigheidscriterium op het moment van het configureren van een webcrawler. De configuratie van de *http-fetcher* is, voor zover deze is ontworpen om archivering te faciliteren, mogelijk relevant voor het beoordelen van de stelselmatigheid op het moment van het onderzoeken van de gecrawlde gegevens. De *frontier data structure*, *http-fetcher* en *URL Prioritzer* vormen in zoverre aangrijpingspunten voor het beoordelen van de stelselmatigheid van het geautomatiseerd overnemen van gegevens met behulp van webcrawlers.

4. De praktische bruikbaarheid van het stelselmatigheidscriterium in relatie tot het geautomatiseerd overnemen van persoonsgegevens met webcrawlers voor de opsporingspraktijk

In paragraaf 3 is een overzicht geboden van de (werking van) de onderdelen van een webcrawler, en zijn de in strafvorderlijke zin relevante onderdelen daarvan aan bod gekomen. Het is nu tijd om te bezien in hoeverre het stelselmatigheidscriterium in de context van de inzet van webcrawlers voor de opsporingspraktijk praktisch bruikbaar is. Als vertrekpunt daarvoor dienen de door de commissie-Koops in relatie tot het overnemen van gegevens uit publiek toegankelijke bronnen genoemde[48] (en door de moderniseringswetgever overgenomen[49]) vier categorieën van relevante factoren: de omvang en het type van de (over te nemen) gegevens, de aard van de bron, de wijze van zoeken en het gebruik van gegevens en de mogelijke impact op de persoon.

Per categorie van factoren heeft de wetgever tevens de door de commissie-Koops gegeven toelichting, waarin de praktische betekenis van de factoren wordt uiteengezet, grotendeels overgenomen. De wetgever heeft in de MvT geen rangorde of hiërarchie tussen de categorieën van factoren en de daarbinnen onderscheiden individuele factoren aangebracht, zodat de indruk ontstaat dat het relatieve gewicht van de relevante factoren in abstracto bezien gelijk is.

In lijn met de onderzoeksvraag zal nu worden geanalyseerd in hoeverre de door de commissie-Koops onderscheiden categorieën van factoren de opsporingspraktijk handvatten bieden bij het uitvoeren van de stelselmatigheidstoets op de volgende momenten binnen het crawlingproces en de (al dan niet geautomatiseerde) verwerking van de in het crawlingproces verzamelde gegevens:

- het configureren van een webcrawler;
- het beoordelen van de relevantie van de gecrawlde gegevens en;
- het onderzoeken van de gecrawlde gegevens.

4.1 Categorie 1: de omvang en het type van de over te nemen gegevens

Tot de categorie 'omvang en het type van de over te nemen gegevens' behoren de volgende drie factoren:

- de hoeveelheid van de gegevens;
- de aard van de gegevens; en
- de diversiteit van de gegevens.

De commissie-Koops verduidelijkt in haar rapport dat deze factoren *mede* samenhangen met de mate waarin vooraf bekend is met welke frequentie en op welke wijze de verdachte zich op het internet manifesteert. De wetgever heeft deze toelichting in de MvT overgenomen, waarbij opvalt dat het woord '*mede*' is geschrapt.[50] Even verderop maakt de wetgever het voorbehoud dat de hoeveelheid, de plaats en de aard van de gegevens die in beeld zullen komen vooraf niet altijd goed kunnen worden ingeschat.[51]

In paragraaf 3 kwam reeds aan de orde dat de uitwerking van de voorkeursinstellingen van een webcrawler op de inrichting en indexering van de webpagina die wordt gecrawld, op voorhand redelijkerwijs niet volledig valt te voorzien (bijvangst). De wetgever merkt op dat in twij-

48 Commissie-Koops, p. 163-164.
49 MvT, p. 500-502.
50 MvT, p. 501.
51 MvT, p. 502.

doi: 10.5553/PMSv/258950952022005003001

felgevallen – gevallen waarin op voorhand duidelijk is dat een gerede kans bestaat op bijvangst (zoals het crawlen van gegevens die afgeschermd hadden moeten blijven) – aan de toepassingsvoorwaarden van artikel 2.8.8. (nieuw) Sv zou moeten worden voldaan en een bevel van de officier van justitie zou moeten worden aangevraagd. In de visie van de wetgever komt het erop aan te beoordelen of sprake is van een 'twijfelgeval', waarbij de factor 'aard van de gegevens' vermoedelijk veelal doorslaggevend zal zijn. Naarmate de over te nemen gegevens gevoeliger zijn, zal een kans op bijvangst eerder als onaanvaardbaar worden beoordeeld en valt te verwachten dat zekerheidshalve een bevel van de officier van justitie zal worden aangevraagd. In het verlengde daarvan ligt het in de rede om de 'aard van de gegevens' op te vatten als de mate van privacygevoeligheid van de gegevens. De factor 'aard van de gegevens' kan in zoverre dienen als een indicator voor stelselmatigheid op het moment van het configureren van een webcrawler.

Een andere relevante kwestie in verband met de factor 'aard van de gegevens' is de door de wetgever geopperde mogelijkheid dat een opsporingsambtenaar ervoor kiest om eerst een oppervlakkige zoekslag te maken, om een eerste beeld te krijgen van wat een uitgebreide zoekslag op zou kunnen leveren.[52] De ratio hiervan lijkt te zijn dat een oppervlakkige zoekslag slechts tot een geringe inbreuk op de persoonlijke levenssfeer van de verdachte leidt. Maar wat houdt een 'oppervlakkige zoekslag' precies in? En biedt de term 'oppervlakkige zoekslag' opsporingsautoriteiten op voorhand voldoende houvast? De term 'oppervlakkige zoekslag' (los van de vraag wat dat precies betekent) heeft in de context van de inzet van webcrawlers slechts betekenis voor zover webcrawlers zodanig kunnen worden geconfigureerd dat zij slechts inschatten of een bepaalde internetbron mogelijk voor het opsporingsonderzoek relevante informatie bevat. Betwijfeld moet worden of het praktisch werkbaar is om een webcrawler op zodanige wijze te configureren. Voorstelbaar is dat de selectie van relevante hyperlinks/URL's die in de *frontier data structure* worden ingeladen al zeer tijdrovend kan zijn, en dat zal nog sterker gelden voor het selecteren van gegevens die worden overgenomen. Bovendien kan de vraag worden gesteld in hoeverre het programmeren van een oppervlakkige zoekslag technisch mogelijk is: in hoeverre is het mogelijk om onbedoelde bijvangst op voorhand (nagenoeg) geheel uit te sluiten? De beantwoording van deze vraag zou een vergelijkend technisch onderzoek vergen naar verschillende webcrawlerconfiguraties die de politie in de opsporingscontext hanteert, waarbij wordt getoetst in hoeverre die configuraties in staat zijn om bijvangst te vermijden. Bij deze stand van zaken zou vanuit rechtsbeschermingsperspectief kunnen worden geconcludeerd dat – nu vooralsnog niet duidelijk is of webcrawlers een oppervlakkige zoekslag kunnen verrichten – het geautomatiseerd overnemen van persoonsgegevens met webcrawlers in gevallen als stelselmatig moet worden aangemerkt en derhalve altijd een bevel van de officier van justitie is vereist. Deze benadering verdient naar mijn mening echter geen navolging, omdat daardoor een standaardwerkwijze tot ontwikkeling zou kunnen komen waarbij verlening van een bevel het uitgangspunt is. Daardoor zouden de stelselmatigheidstoets en daarmee de privacybelangen van de verdachte en de gegevensbeheerder naar de achtergrond verdwijnen. Het behoeft geen betoog dat dit vanuit rechtsstatelijk oogpunt niet wenselijk is en uiteindelijk ook niet in het belang kan zijn van de opsporingspraktijk.

Ondertussen lijkt de 'aard van de gegevens' (lees: de privacygevoeligheid van de gegevens) ook relevant voor het beoordelen van stelselmatigheid bij het onderzoeken van de gecrawlde gegevens. Evenzeer geldt dat voor de factor 'hoeveelheid van de gegevens'. Naarmate bij de verwerking van gecrawlde gegevens kennis wordt genomen van een grotere hoeveelheid gevoelige gegevens, nemen immers de aard en de ernst van de privacyinbreuk toe en zal eerder sprake zijn van stelselmatigheid. Dat zou wellicht anders kunnen zijn indien de verwerking van de gecrawlde gegevens volledig geautomatiseerd, zonder tussenkomst van een opsporingsambtenaar, zou kunnen plaatsvinden. Door een volledig geautomatiseerde verwerking van gecrawlde gegevens zou de inbreuk op de persoonlijke levenssfeer maximaal worden beperkt. Hoewel een volledig geautomatiseerde verwerking van gegevens technisch mogelijk is (denk bijvoorbeeld aan het inladen van verzamelde digitale sporen in de forensische zoekmachine Hansken),[53] bestaan er geen aanwijzingen dat geautomatiseerde gegevensverwerking in het kader van het geautomatiseerd overnemen van persoonsgegevens met webcrawlers ook al mogelijk is.

Dan de factor 'diversiteit van de gegevens'. Noch het rapport van de commissie-Koops, noch de MvT bevat aanwijzingen voor de wijze waarop deze factor zou moeten worden ingevuld. Een gedachte zou kunnen zijn dat naarmate de beoordeling van de privacygevoeligheid van de gegevens die worden overgenomen in verband met de diverse samenstelling van die gegevens lastiger valt te maken, sneller gekozen zou moeten worden voor het aanwenden van de strafvorderlijke bevoegdheid tot het stelselmatig overnemen van persoonsgegevens uit publiek toegankelijke bronnen.

Het gegevensbestand van het Kadaster, dat naast registratiegegevens van onroerend goed onder meer naam, adres, geboortedatum en geboortedatumplaats en nummers van identiteitsdocumenten zoals paspoort of rijbewijs van gerechtigden bevat, zou in dit verband al snel kwalificeren als een 'diverse verzameling van gegevens', zodat een bevel tot het stelselmatig overnemen van persoonsgegevens zou moeten worden aangevraagd. Dat zou anders liggen bij specifieke gegevensbestanden waarin slechts een of enkele persoonsgegevens zijn

52 MvT, p. 500.

53 Meer informatie over de technische werking van Hansken is te vinden in het online te raadplegen Informatieblad 'Forensische waarborgen in Hansken' (versie: 4 november 2021). Vindplaats: www.forensischinstituut.nl/publicaties/publicaties/2021/11/4/de-forensische-waarborgen-hansken.

doi: 10.5553/PMSv/258950952022005003001

vastgelegd, zoals het telefoonboek. Als de hiervoor genoemde interpretatie van de factor 'de diversiteit van de gegevens' juist is, verdient deze gedachtegang mijns inziens navolging. De factor 'diversiteit van de gegevens' kan in zoverre dienen als indicator voor stelselmatigheid op het moment van het configureren van een webcrawler.

4.2 Categorie 2: de aard van de bron

Tot de categorie 'aard van de bron' behoren de volgende factoren:

- de aard van de locatie waar de gegevens te vinden zijn;
- de menselijke bron van de gegevens;
- het doel waarmee de gegevens in eerste instantie zijn vergaard of gepubliceerd;
- de plaats van de gegevens; en
- de feitelijke bekendheid van de gegevens.

Mijn indruk is dat enkel de factor 'het doel waarmee de gegevens in eerste instantie zijn vergaard of gepubliceerd' relevant is voor het op voorhand bepalen van de mate van inbreuk op de persoonlijke levenssfeer van de verdachte, en dat de overige factoren die binnen deze categorie worden onderscheiden slechts gezichtspunten betreffen die bij de invulling van deze factor kunnen worden betrokken. Ik licht dat hierna toe.

De factor 'het doel waarmee de gegevens in eerste instantie zijn vergaard of gepubliceerd' brengt naar mijn mening op heldere wijze tot uitdrukking dat in het kader van de invulling van het stelselmatigheidscriterium bij geautomatiseerd overnemen van persoonsgegevens met behulp van een webcrawler mede van belang is in hoeverre de verdachte redelijkerwijs mag verwachten dat bepaalde persoonsgegevens uit openbare internetbronnen kenbaar zijn. Als objectieve aanwijzingen ontbreken dat inloggegevens van een e-mailaccount eerder bij een datalek onbedoeld openbaar zijn geworden, geldt die redelijke verwachting uiteraard niet. Maar als die objectieve aanwijzingen wel degelijk bestonden, bijvoorbeeld omdat de inloggegevens op de website 'Have I been Pwned?' (zie: https://haveibeenpwned.com) worden vermeld, zou die redelijke verwachting wellicht kunnen worden aangenomen. Bij een verdachte die redelijkerwijs mocht verwachten dat zijn persoonsgegevens deugdelijk waren afgeschermd, levert het na inzet van een webcrawler aantreffen van die persoonsgegevens een ingrijpender inbreuk op de persoonlijke levenssfeer op dan bij een verdachte die zijn persoonsgegevens niet behoorlijk heeft afgeschermd. De factor 'het doel waarmee de gegevens in eerste instantie zijn vergaard of gepubliceerd' kan in zoverre dienen als contra-indicatie voor stelselmatigheid op het moment van het onderzoeken van de gecrawlde gegevens.

Het gewicht van deze factor dient mijns inziens echter te worden gerelativeerd voor zover het afgeschermde gegevens betreft die op geen enkele manier in verband staan met de persoonlijke levenssfeer van de verdachte. De relevantie van deze factor neemt af naarmate de gegevens minder privacygevoelig zijn.

Benadrukt moet worden dat het niet waarschijnlijk is dat met inzet van een webcrawler toegang kan worden verkregen tot daadwerkelijk afgeschermde persoonsgegevens. Indien persoonsgegevens bijvoorbeeld in een beveiligd digitaal account zijn opgeslagen, dan dienen (afhankelijk van de feitelijke situatie) andere strafvorderlijke bevoegdheden te worden aangewend, zoals de netwerkzoeking (art. 125j Sv), de wettelijke hackbevoegdheid (art. 126nba Sv) of een rechtshulpverzoek gericht aan een buitenlandse autoriteit teneinde bij een clouddienstaanbieder (zoals Google of Facebook) opgeslagen persoonsgegevens te verkrijgen.[54] De factor is dus niet van toepassing op afgeschermde persoonsgegevens, nu het niet waarschijnlijk is dat dergelijke gegevens met inzet van een webcrawler worden verkregen.

Wel kan, zoals hiervoor reeds besproken, de situatie zich voordoen dat persoonsgegevens in een beveiligd digitaal account of andere beveiligde internetomgeving worden verkregen door het (tijdelijk) falen van de beveiliging, waardoor dergelijke gegevens plotseling openbaar toegankelijk worden. Hoewel dergelijke datalekken veelal binnen enkele uren worden hersteld, is niet ondenkbaar dat dergelijke persoonsgegevens in het kader van de opsporing, al dan niet bewust, met inzet van een webcrawler worden verkregen. Voor zover in dergelijke gevallen sprake is van het na een datalek bewust inzetten van een webcrawler, is juridisch gezien inderdaad sprake van een ingrijpender inbreuk op de persoonlijke levenssfeer dan in het geval dergelijke persoonsgegevens min of meer toevallig worden verkregen. Een ongewenst gevolg van die opvatting zou mogelijk kunnen zijn dat de daadwerkelijke reden van de inzet van de webcrawler niet in het proces-verbaal wordt vermeld.

In gevallen waarin persoonsgegevens worden verkregen waarvan concrete aanwijzingen bestaan dat zij afkomstig zijn uit een beoogd beveiligde internetomgeving, verdient het vanuit rechtsbeschermingsperspectief de voorkeur dat in strafvorderlijke zin wordt gehandeld alsof toegang wordt verkregen tot afgeschermde persoonsgegevens, en wordt teruggegrepen op de in de vorige alinea genoemde strafvorderlijke bevoegdheden.

De overige factoren, namelijk 'de aard van de locatie waarop de gegevens te vinden zijn', 'de plaats van de gegevens', 'de feitelijke bekendheid van de gegevens' en 'de menselijke bron van de gegevens' zijn mijns inziens betreft veeleer te beschouwen als gezichtspunten, die mede bepalen in hoeverre voor de verdachte redelijkerwijs voorzienbaar is dat zijn of haar persoonsgegevens door inzet van een webcrawler kunnen worden geraadpleegd. Daarbij is van belang dat deze gezichtspunten weinig onderscheidend vermogen bezitten, omdat ze steeds verwijzen naar het doel waarmee de persoonsgegevens op internet zijn geplaatst. Als de verdachte zijn of haar persoonsgegevens wil afschermen, worden die persoonsgegevens opgeslagen in een beveiligde inter-

54 Zie de eerdere jurisprudentieverwijzing in voetnoot 19.

doi: 10.5553/PMSv/258950952022005003001

netomgeving, en genieten die gegevens doorgaans weinig bekendheid.

4.3 Categorie 3: de wijze van zoeken

Tot de categorie 'wijze van zoeken' behoren de volgende vier factoren:

- de 'geavanceerdheid' van het gebruikte technisch hulpmiddel;
- het doel van de zoekactie;
- de specificiteit van de zoekvraag; en
- de samenhang tussen de zoekvraag en het strafbare feit.

Deze categorie van factoren vormt in combinatie met de factor 'aard van de gegevens' (onder categorie 1) de kern van de bij het configureren van een webcrawler te verrichten stelselmatigheidstoets. Niettemin geldt ook voor deze categorie dat de individuele factoren in het kader van de inzet van webcrawlers niet allemaal evenveel gewicht in de schaal leggen. De factor 'de geavanceerdheid van het gebruikte technisch hulpmiddel' is mijns inziens bijvoorbeeld niet zo relevant, omdat de verschillende configuraties van webcrawlers wat betreft complexiteit geen grote variëteit laten zien. In dit verband wordt in herinnering gebracht dat het mogelijk is een uniforme beschrijving van de onderdelen van een webcrawler te geven (zie par. 3).

Mijn indruk is dat met name de factor 'de specificiteit van de zoekvraag' praktisch bruikbaar is voor de opsporingspraktijk, en dat de overige factoren die binnen deze categorie worden onderscheiden slechts gezichtspunten betreffen die bij de invulling van 'de specificiteit van de zoekvraag' kunnen worden betrokken. Ik licht dat hierna toe.

De factor 'specificiteit van de zoekvraag' dient binnen de context van webcrawlers niet in verband te worden gebracht met de invoer van zoektermen. In zoverre zou in plaats van 'zoekvraag' eigenlijk beter kunnen worden gesproken van een 'sleepnet'. Webcrawlers werken niet op basis van ingevoerde zoektermen, maar op basis van ingevoerde hyperlinks/URL's (de *frontier data structure*, zoals besproken in par. 3.1). Deze URL's/hyperlinks, bepalen 'de specificiteit van het sleepnet'.

Men zou immers kunnen redeneren dat hoe directer de relatie tussen de URL's/hyperlinks en het opsporingsdoel en de aard van de verdenking, hoe specifieker of gerichter het sleepnet is.

Vanuit rechtsbeschermingsperspectief is met name van belang dat de mate van inbreuk op de persoonlijke levenssfeer waartoe de inzet van een webcrawler leidt, niet ingrijpender is dan het opsporingsdoel strikt genomen rechtvaardigt. Als de interesse van de opsporingsautoriteiten bijvoorbeeld specifiek uitgaat naar het aantal in kinderpornografie geïnteresseerde leden in nieuwsgroepen en chatfora in Groningen, dan is vanuit de in artikel 2.1.3 Sv (nieuw) verankerde proportionaliteits- en subsidiariteitseis niet aanvaardbaar dat de webcrawler alle nieuwsgroepen en chatfora in Nederland onderzoekt. De factor 'specificiteit van de zoekvraag' kan in zoverre dienen als contra-indicatie voor stelselmatigheid op het moment van het configureren van een webcrawler.

4.4 Categorie 4: de opslag en het gebruik van de gegevens en de mogelijke gevolgen voor de persoon

Tot de categorie 'opslag en het gebruik van de gegevens en de mogelijke gevolgen voor de persoon' behoren de volgende drie factoren:

- de mate waarin gegevens worden overgenomen en de selectiviteit die daarbij wordt gehanteerd;
- de in het onderzoek al bekende informatie; en
- de combinatie van gegevens uit verschillende bronnen.

Deze categorie van factoren vergt van de opsporingsautoriteiten dat zij voorafgaand aan de inzet van een webcrawler een inschatting maken van de (voorlopige) resultaten van het opsporingsonderzoek en de wijze waarop het stelselmatig overnemen van persoonsgegevens zich daartoe verhoudt. Het maken van deze inschatting zal met name in het beginstadium van het opsporingsonderzoek ingewikkeld kunnen zijn, omdat in dat geval op voorhand slechts beperkt inzichtelijk is op welke manier de verkregen persoonsgegevens zullen bijdragen aan het verloop van het opsporingsonderzoek, zeker indien die persoonsgegevens al dan niet op geautomatiseerde wijze met gegevens uit andere bronnen worden gecombineerd. Hierna worden enkele kanttekeningen geplaatst bij de praktische bruikbaarheid van deze factoren.

De factor 'de mate waarin gegevens worden overgenomen en de selectiviteit die daarbij wordt gehanteerd' is slechts beperkt bruikbaar, omdat op voorhand nu eenmaal geen absolute zekerheid kan worden verkregen over de hoeveelheid en aard van de persoonsgegevens die bij het hanteren van bepaalde voorkeurinstellingen zullen worden verkregen. Anders dan bij zoekmachines bestaat bij webcrawlers een minder rechtstreeks verband tussen de zoekopdracht (de ingevoerde URL's/hyperlinks) en de resultaten die worden gegenereerd en overgenomen.

Een andere complicatie is dat webcrawlers in principe geen functionaliteit kennen waarmee binnen de inhoud van de ingevoerde URL's/hyperlinks een nadere selectie kan worden gemaakt van de persoonsgegevens die worden overgenomen. De inhoud van de ingevoerde URL's/hyperlinks wordt al naar gelang de rangorde daarvan 'gecrawld', en daarbij is een scherpe selectie van de onderdelen van de webpagina's die worden overgenomen doorgaans niet mogelijk.

Een Osint-specialist/opsporingsambtenaar maakt vervolgens een selectie van de relevante persoonsgegevens die in het dossier moeten worden opgenomen. Problematisch is wel dat de opsporingsambtenaar in dat kader ook kennisneemt van de niet-relevante persoonsgegevens, en dat naarmate het aandeel persoonsgegevens daarin groter is daardoor een ingrijpender inbreuk wordt gemaakt op de persoonlijke levenssfeer van de verdachte. De hiervoor beschreven werkwijze maakt duidelijk

doi: 10.5553/PMSv/258950952022005003001

dat het in de MvT genoemde gezichtspunt ('worden gegevens alleen beperkt en gericht overgenomen in politiesystemen, of is er juist sprake van een brede en weinig selectieve overneming van wat is aangetroffen?'), niet aansluit bij de werkelijkheid waarin na het crawlen van de inhoud van webpagina's een nadere selectie plaatsvindt door een opsporingsambtenaar. Bij de ontwikkeling van webcrawlers die in het kader van de opsporing worden ingezet zou het ontwikkelen van functionaliteiten waarmee een scherpe(re) selectie van de onderdelen van de webpagina's die geautomatiseerd worden overgenomen mogelijk wordt meer prioriteit moeten krijgen. Naarmate webcrawlers de inhoud van webpagina's op selectievere wijze kunnen crawlen, zal de normerende rol van deze factor in belang afnemen. Dit leidt tot de interessante paradox dat hoewel de wetgever de 'geavanceerdheid van het gebruikte technisch hulpmiddel' als aanwijzing ziet voor een grotere inbreuk op de persoonlijke levenssfeer, het geavanceerder/selectiever worden van webcrawlers er juist toe leidt dat de inbreuk op de persoonlijke levenssfeer geringer wordt.

De factor 'de in het onderzoek al bekende informatie' vertoont in belangrijke mate overlap met de factor 'de specificiteit van de zoekvraag'. Algemeen gesteld geldt immers dat de opgegeven voorkeurinstellingen (de verzameling URL's/hyperlinks) specifieker zullen zijn naarmate het opsporingsonderzoek zich in een verder gevorderd stadium bevindt. De wetgever noemt in de MvT als gezichtspunt bij deze factor 'hoeveel de zoekactie zal toevoegen aan het bestaande beeld'. Daarvoor geldt, zoals hiervoor besproken, weer het bezwaar dat op voorhand niet met zekerheid kan worden voorspeld hoeveel en welke persoonsgegevens bij het hanteren van bepaalde voorkeurinstellingen zullen worden verkregen. Kortom: deze factor heeft mijns inziens weinig meerwaarde. Ten slotte wordt over de factor 'de combinatie van gegevens uit verschillende bronnen' kortheidshalve opgemerkt dat de verwerking van persoonsgegevens in geautomatiseerde zoeksystemen niet binnen de reikwijdte van de strafvorderlijke bevoegdheid tot het stelselmatig overnemen van persoonsgegevens valt. Daarom valt naar mijn mening niet goed in te zien op welke manier deze factor van invloed is op de op voorhand te verrichten stelselmatigheidstoets in het kader van de inzet van webcrawlers bij het stelselmatig overnemen van persoonsgegevens.

5. Afsluiting

In dit artikel heb ik getracht de vraag te beantwoorden in hoeverre opsporingsautoriteiten op basis van de relevante gezichtspunten/factoren voor stelselmatigheid en bezien in het licht van de huidige technische mogelijkheden van een webcrawler de mate van inbreuk op de persoonlijke levenssfeer kunnen inschatten en de aard en ernst van die inbreuk zo veel mogelijk kunnen beperken.

De operationalisering van het stelselmatigheidscriterium binnen de specifieke context van de inzet van webcrawlers stelt de opsporing voor bijzondere uitdagingen, omdat de stelselmatigheidstoets zich op drie momenten gedurende de inzet van een webcrawler opdringt, namelijk: bij de configuratie van een webcrawler, bij de beoordeling van de relevantie van de gecrawlde gegevens en bij het onderzoeken van de gecrawlde gegevens.

Geanalyseerd is in hoeverre de factoren die de commissie-Koops specifiek voor de strafvorderlijke normering van het geautomatiseerd overnemen van persoonsgegevens in haar rapport noemt voor opsporingsautoriteiten bruikbaar zijn om de stelselmatigheid te beoordelen. In de hierna opgenomen tabel wordt een overzicht gegeven van factoren die mijns inziens voor de opsporingspraktijk binnen de context van webcrawlers praktisch bruikbaar zijn. Per factor wordt de doorwerking ervan op de stelselmatigheid en het niveau van doorwerking weergegeven. Uit de tabel kunnen mijns inziens twee relevante bevindingen over de praktische bruikbaarheid van het stelselmatigheidscriterium voor opsporingsautoriteiten worden gedestilleerd. De eerste bevinding luidt dat de criteria op basis waarvan het stelselmatigheidscriterium kan worden ingevuld op (ten minste) twee niveaus doorwerken, namelijk op het moment van configuratie van een webcrawler en op het moment van onderzoek van gecrawlde gegevens. De tweede bevinding luidt dat het aantal voor opsporingsautoriteiten praktisch bruikbare factoren aanzienlijk kan worden teruggebracht, namelijk tot vier factoren (aard van de gegevens, diversiteit van de gegevens, het doel waarmee de gegevens in eerste instantie zijn vergaard of gepubliceerd en specificiteit van de zoekvraag (lees: sleepnet). Tot besluit een geruststelling: naarmate de configuratie van webcrawlers die in het kader van de opsporing worden ingezet een steeds specifieker sleepnet toelaat, zal de inbreuk op de persoonlijke levenssfeer van de verdachte geringer worden, en de complexiteit van de toepassing van het stelselmatigheidscriterium binnen de specifieke context van webcrawlers steeds verder afnemen.

doi: 10.5553/PMSv/258950952022005003001

Tabel 1 Overzicht van voor de opsporingspraktijk praktisch bruikbare factoren bij de inzet van webcrawlers

Factor	Invloed op stelselmatigheid	Niveau
Aard van de gegevens (lees: privacygevoeligheid van de gegevens)	Naarmate de privacygevoeligheid van de gegevens toeneemt, wordt de kans op onbedoelde bijvangst eerder onaanvaardbaar beoordeeld en is eerder sprake van stelselmatigheid Naarmate bij het onderzoek van gecrawlde gegevens kennis wordt genomen van gevoeligere gegevens, nemen de aard en ernst van de privacyinbreuk toe en is eerder sprake van stelselmatigheid	Configuratie Onderzoek van gecrawlde gegevens
Diversiteit van de gegevens	Naarmate de beoordeling van de privacygevoeligheid van de gegevens die worden overgenomen in verband met de diverse samenstelling van die gegevens lastiger valt te maken, is eerder sprake van stelselmatigheid	Configuratie
Het doel waarmee de gegevens in eerste instantie zijn vergaard of gepubliceerd	Naarmate de verdachte redelijkerwijs mag verwachten dat bepaalde persoonsgegevens uit openbare internetbronnen kenbaar zijn, is minder snel sprake van stelselmatigheid	Onderzoek van gecrawlde gegevens
Specificiteit van de zoekvraag (lees: sleepnet)	Naarmate de relatie tussen de URL's/hyperlinks en het opsporingsdoel en de aard van de verdenking directer is, is minder snel sprake van stelselmatigheid	Configuratie

doi: 10.5553/PMSv/258950952022005003001