*Hans Henseler**

Redactioneel

Het inzagerecht en de groeiende omvang van digitaal bewijs

In 2017 werden criminelen opgeschrikt toen de Nederlandse politie en het Openbaar Ministerie (OM) toegang hadden gekregen tot 3,6 miljoen versleutelde berichten die vermoedelijk voor een groot deel afkomstig zijn uit de georganiseerde misdaad.[1] Het bericht voorspelde dat de ontdekte gegevens kunnen leiden tot grote, beslissende doorbraken in strafzaken. Wie op de website van het OM zoekt op de term 'Ennetcom' ziet inderdaad dat sindsdien in verschillende zaken wordt aangegeven dat het bewijs vooral uit de ontsleutelde PGP-berichten[2] van aanbieder Ennetcom afkomstig is. Het bleef niet bij de PGP-berichten van Ennetcom. Onder de kop 'Aardschok voor georganiseerde misdaad' meldt het OM in juli van dit jaar: 'Meer dan 20 miljoen onderschepte chatberichten van criminelen zijn de afgelopen maanden voor een belangrijk deel LIVE meegelezen door politie en justitie in Nederland'.[3] Ditmaal waren de berichten afkomstig van het bedrijf EncroChat, een van de grootste aanbieders van versleutelde digitale communicatie.

Maar hoe wordt er nu in zo'n berg met informatie gezocht en gebeurt dat wel op een goede manier? Die vraag was inzet van de verdediging in de *Tandem II*-zaak. De algemene overwegingen met betrekking tot die zaak zijn te lezen in de uitspraak.[4] Het probleem is dat de 3,6 miljoen berichten in de Ennetcom-data bij elkaar een onhanteerbaar grote hoeveelheid vormen. Aan de hand van zoektermen in de door de rechter-commissaris goedgekeurde plannen van aanpak zijn de Ennetcom-data daarom eerst gefilterd. Het onderzoeksteam Tandem kon daardoor alleen verder onderzoek doen naar de gegevens die door middel van die zoektermen waren verkregen. Die gegevens (hierna ook wel: berichten) samen vormen de 'Tandem-dataset'. Deze filtering is uitgevoerd met behulp van de zoekmachine Hansken, die door het NFI is ontwikkeld.[5] Hansken is al eerder ingezet maar deze zaak is de eerste waarin de verdediging heeft gesteld dat zij onvoldoende mogelijkheden heeft gekregen om de integriteit en de betrouwbaarheid van de met behulp van Hansken verkregen resultaten te controleren.

In de uitspraak is te lezen dat de verdediging, in de aanloop naar de inhoudelijke behandeling, bij herhaling verzoeken heeft gedaan die betrekking hebben op Hansken als zoekmachine en op de manier waarop de onderzoeksresultaten zijn verkregen. De beslissingen van de rechtbank op die verzoeken hebben ertoe geleid dat de verdediging twee bezoeken heeft gebracht aan het NFI. Tijdens die bezoeken liet de verdediging zich bijstaan door een eigen deskundige. Tijdens het eerste bezoek kon de verdediging met behulp van Hansken in de Tandem-dataset zoeken. Daarbij had de verdediging toegang tot de berichten die waren gevoegd in het dossier (categorie 1) en berichten die volgens het OM niet-relevante berichten waren (categorie 2). Daarnaast was er een derde categorie berichten, waarvan de inzage aan de verdediging is onthouden. Aan de verdediging werd een cd-rom verstrekt met daarop de inhoud van de berichten van categorie 1 en 2. Uiteindelijk concludeert de rechtbank dat het verweer van de verdediging niet standhoudt. Niet op het punt van de controlemogelijkheid en ook niet op de andere punten die betrekking hadden op Hansken (en die hier verder niet worden uitgewerkt).

Bij mijn weten is dit de eerste keer dat in een strafzaak de controleerbaarheid van het filteren van informatie zo nadrukkelijk door de verdediging aan de orde is gesteld en dat de rechtbank daar vervolgens ook in mee gaat en ruimte maakt voor die controle door de verdediging. In het civiele recht en het bestuursrecht is men op dit gebied al een stuk verder. Zo heeft de NMa (nu ACM) al in 2007 een digitale werkwijze gepubliceerd.[6] Overigens is die werkwijze inmiddels meerdere malen herzien[7] en tot op de dag van vandaag onderwerp van discussie tussen advocaten en de ACM. Maar de NMa was destijds in Nederland en in Europa wel voorloper op dit terrein. In feite was zij een van de eersten die zich bezighielden met *E-Discovery*, een term die in 2007 nog vooral werd gebezigd in de Verenigde Staten. Het *E-Discovery*-vakgebied heeft zich de afgelopen tien jaar ook in Europa stormachtig ontwikkeld en er zijn vele *best practices* gepubliceerd. Inmiddels is er een complete industrie van

* Dr. ir. J. Henseler is lector *E-Discovery* en *Digital Forensics* bij Hogeschool Leiden, senior adviseur bij het Nederlands Forensisch Instituut en lid van het College NRGD. De auteur dankt prof. dr. Ton Broeders, dr. mr. M.J. Dubelaar, mr. Gert Haverkate en mr. ing. Nico Keijser voor hun waardevolle commentaar en aanvullingen op eerdere versies van dit redactioneel.

1. om.nl/actueel/nieuws/2017/03/09/versleutelde-berichten-schat-aan-criminele-informatie.
2. PGP is een afkorting voor Pretty Good Privacy. Dit is een type versleuteling (encryptie) die maakt dat berichten niet zonder meer gelezen kunnen (konden) worden of onderschept kunnen (konden) worden door derden.
3. om.nl/actueel/nieuws/2020/07/02/aardschok-voor-georganiseerde-misdaad.
4. Rb. Amsterdam 19 april 2018, ECLI:NL:RBAMS:2018:2504.
5. Meer informatie over Hansken is te vinden op de website hansken.org.
6. NMa digitale werkwijze 2007, *Stcrt.* 2007, 243.
7. Zie onder meer ACM Werkwijze voor geheimhoudingsprivilege advocaat 2014, *Stcrt.* 2014, 3991 en ACM Werkwijze voor onderzoek in digitale gegevens 2014, *Stcrt.* 2014, 3993.

bedrijven die *E-Discovery*-tools, -diensten en/of -opleidingen aanbieden.

Een bekende uitdaging in *E-Discovery*-projecten in de VS en andere landen met een *Common Law*-systeem is het filteren van informatie waarvoor het Legal Professional Privilege (LPP) geldt. Kort gezegd mag een bedrijf dat verplicht wordt tot het aanleveren van een set e-mails en documenten daar bepaalde vertrouwelijke informatie uit verwijderen. In Nederland spreken we in dit verband van geprivilegieerde gegevens en de ACM heeft een functionaris Verschoningsrecht die bepaalt of gegevens inderdaad geprivilegieerd zijn. Met de strengere AVG-wetgeving speelt dit probleem ook steeds meer met betrekking tot persoonlijke gegevens, bijvoorbeeld bij WOB-verzoeken. In opsporingsonderzoeken wordt meestal de term geheimhouderscommunicatie gebruikt, een term die al veel langer bekend is, bijvoorbeeld als het gaat om taps van telefoongesprekken van verdachte met zijn advocaat, die uit het dossier verwijderd behoren te worden. In recente publicaties in het *Advocatenblad* en het *Tijdschrift voor Bijzonder Strafrecht & Handhaving* wordt betoogd dat het filteren van vertrouwelijke e-mails met handmatig geselecteerde trefwoorden omslachtig is en dat de kosten door het gebruik van kunstmatige intelligentie omlaag kunnen.[8] De *E-Discovery*-specialisten die in het artikel aan het woord zijn, maken gebruik van commerciële *E-Discovery*-software. Het artikel stelt dat het OM en de recherche met kunstmatige intelligentie digitaal in tienduizenden e-mails de vertrouwelijke correspondentie met de advocaat kunnen identificeren. Het zou sneller, goedkoper en beter controleerbaar zijn dan mensenwerk, terwijl het verschoningsrecht meer wordt gerespecteerd. Het artikel noemt vooral de FIOD als voorbeeld en verwijst ook naar een recente uitspraak van de Hoge Raad waaruit blijkt dat de huidige werkwijze (die gebaseerd is op het filteren met zoekwoorden) is toegestaan.[9] Het OM en de FIOD zijn vooralsnog voorzichtig met het uitproberen van nieuwe technologie en lijken het voorlopig vooral bij het geaccepteerde filteren met zoekwoorden te houden.

In de VS is het filteren met zoekwoorden sinds 2012 een gepasseerd station door de uitspraak in de inmiddels beroemde *Da Silva Moore*-rechtszaak.[10] In die uitspraak stelde Judge Peck dat het geoorloofd was om in plaats van trefwoorden gebruik te maken van *predictive coding*. Dit is een techniek die is gebaseerd op *machine learning* waarbij gegevens niet gefilterd worden met zoekwoorden maar waarbij een ervaren advocaat of een team van deskundigen een kleine set documenten stuk voor stuk classificeert als wel of niet relevant. Aan de hand van deze voorbeelden kan een computer een model vormen waarmee vervolgens alle overige (ongelezen) documenten in de set gerangschikt kunnen worden op relevantie. Deze aanpak is niet 100% betrouwbaar maar met behulp van steekproeven en onder bepaalde voorwaarden is het wel mogelijk om een betrouwbaarheid van 95% of 99% of zelfs hoger te behalen zonder alle documenten te hoeven reviewen. Overigens betekent dit zeker niet dat in de VS nu in alle gevallen gekozen wordt voor *predictive coding*. Judge Peck stelt nadrukkelijk dat er niet gestreefd moet worden naar perfectie maar dat de rechtsgang bovenal moet streven naar zoekresultaten die redelijk en proportioneel zijn (de inspanning moet in verhouding staan tot de omvang van de zaak en hetgeen in de zaak vereist is). Het valt te verwachten dat binnen niet al te lange tijd ook in Nederland nieuwe technologie zoals *predictive coding* ingezet zal worden. Misschien eerst in het civiele recht en het bestuursrecht maar het is een kwestie van tijd tot het ook in het strafrecht zal worden ingezet.

Terug nu naar de uitspraak in de *Tandem*-zaak. Ten aanzien van het verweer van de verdediging dat zij onvoldoende mogelijkheden heeft gehad voor 'contra-expertise' haalt de rechtbank het volgende toetsingskader aan: 'Het recht op een eerlijk proces, zoals vastgelegd in artikel 6 EVRM, veronderstelt onder meer dat een verdachte kennis kan nemen van het volledige procesdossier en reële en effectieve mogelijkheden dient te hebben om tegen het hem gemaakte verwijt in te brengen wat hij in het belang van zijn verdediging acht.' Zoals eerder beschreven, heeft dit ertoe geleid dat de rechtbank de verdediging in staat heeft gesteld om met behulp van Hansken in de Tandem-dataset te zoeken in berichten die zijn toegevoegd aan het dossier en in berichten die het OM niet relevant vond. Maar daarnaast zijn er ook berichten in de Tandem-dataset die niet relevant zijn maar waarvan inzage aan de verdediging is onthouden omdat er zwaarwegende opsporingsbelangen in andere opsporingsonderzoeken aan inzage door de verdediging in de weg staan (categorie 3 berichten). De verdediging vond dat zij in het kader van een eerlijk proces recht had om ook deze gegevens in te zien. De rechter oordeelde dat het besluit om berichten aan de verdediging te onthouden geoorloofd was omdat de rechter-commissaris deze beslissing in nauw overleg met het onderzoeksteam heeft genomen.

In Engeland is er de laatste jaren veel aandacht voor het reviewen van digitale documenten. Dit onderwerp is daar onderdeel van het bredere thema *disclosure*. *Disclosure* kan kort worden gedefinieerd als '(…) *the process in a criminal case by which someone charged with a crime is provided with copies of, or access to, material from the investigation that is capable of undermining the prosecution case against them and/or assisting their defence. Without this process taking place a trial would not be fair.*'[11] Na een aantal grote schandalen over falende *disclosure* in de pers,[12] heeft de UK Crown Prosecution Service (CPS) de richtlijnen voor het reviewen van digitaal

8. Zie H.B.J. Sluijsmans & V.J.C. de Bruijn, 'Mogelijkheden voor het beter waarborgen van het verschoningsrecht door beheerst gebruik van machine learning', *Tijdschrift voor Bijzonder Strafrecht & Handhaving* 2020, afl. 3.
9. HR 16 juni 2020, ECLI:NL:HR:2020:1048.
10. Voor een samenvatting zie lexisnexis.com/community/casebrief/p/casebrief-da-silva-moore-v-publicis-groupe. De volledige uitspraak is te vinden op cite.case.law/frd/287/182/.
11. assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/756436/Attorney_General_s_Disclosure_Review.pdf.
12. telegraph.co.uk/news/2018/07/12/policeand-prosecution-lawyers-fail-correctly-disclose-evidence/.

materiaal in december 2018 herzien.[13] Het voert te ver om hier uitvoerig op deze handleiding in te gaan maar de geïnteresseerde lezer kan een samenvatting (ten aanzien van de digitale aspecten) lezen die ik in 2019 heb gepubliceerd op *LinkedIn.*[14] De principes die zijn geformuleerd, lijken in grote lijnen overeen te komen met het besluit van het OM in Nederland in de *Tandem*-zaak. Het is interessant om te lezen wat er gezegd wordt over het onthouden van inzage aan de verdediging. In de aangepaste richtlijn wordt niet gesproken over zwaarwegende opsporingsbelangen maar over het onbedoeld inzage geven aan de verdediging in gegevens die vertrouwelijk of gevoelig zijn voor de eigenaar van die data, zoals bijvoorbeeld gegevens op de telefoon van het slachtoffer.[15]

De datasets die aan de hand van zoekwoorden worden samengesteld uit datasets zoals de Ennetcom-dataset en de EncroChat-dataset zijn ondanks het filteren nog steeds zeer omvangrijk. Ook het digitaal bewijs dat tegenwoordig uit een enkele moderne smartphone wordt geëxtraheerd, kan flink oplopen. Het analyseren van meerdere van zulke datasets in een onderzoek naar een strafbaar feit mag gerust een uitdaging genoemd worden. Niet alleen omdat dit bijzondere analytische vaardigheden vergt van rechercheurs maar vooral ook omdat de apparatuur en software die hun daarbij ter beschikking staat niet krachtig genoeg is. Interessant gegeven in dit verband is een voorbeeld van een onderzoek waarbij de extractie van de gegevens uit één iPhone een pdf-document opleverde van maar liefst 77.486 pagina's. Dit feit is te lezen in de achtergrond die geschetst wordt ter onderbouwing van een nieuw te bouwen Remote Search & Review platform waarmee rechercheurs ondersteund worden in het reviewen van digitaal bewijs.[16] Iedereen die wel eens geprobeerd heeft om een pdf-bestand met zoveel pagina's te openen, weet dat een standaard kantoorcomputer het daar lastig mee heeft. Zelfs een zwaar workstation van een digitaal forensische expert heeft daar problemen mee. Daar komt nog bij het probleem dat met alleen de ingebouwde Ctrl+f-zoekfunctie het lastig analyseren is in zoveel pagina's ongestructureerde data. Ook de gratis versies van de forensische tools (zogenaamde Reader-versies) waarmee wel slim gefilterd kan worden, vergen stevige hardware en het beschikbaar stellen van zo'n *case*-bestand via verwisselbare opslagmedia is omslachtig, tijdrovend en verhoogt het risico op dataverlies. Als vervolgens ook nog gevraagd wordt om verbanden te vinden tussen verschillende telefoons en datasets en om samen te kunnen werken met collega's dan mag duidelijk zijn dat dit op deze manier onbegonnen werk is.

Niet alleen de rechercheurs ervaren deze problemen. Na de twee bezoeken van de verdediging in de *Tandem II*-zaak aan het NFI hebben zij de beschikking gekregen over een cd-rom met de berichten in categorie 1 en 2. Bij herhaling heeft de verdediging aangevoerd dat de cd-rom wat haar betreft niet goed toegankelijk is. Het ontbreken van goede tools om in digitaal bewijs te zoeken heeft er in Engeland mede toe geleid dat rechercheurs belangrijk bewijs over het hoofd zagen waardoor rechtszaken stukliepen. Een sprekend voorbeeld is een vermeende verkrachtingszaak. Het verzoek van een advocaat van de verdachte om berichten van de telefoon van het slachtoffer in te mogen zien, werd in eerste instantie geweigerd. Op de eerste dag van de rechtszaak gaf de rechter haar echter wel toestemming. Na het doorlezen van 40.000 berichten die nacht en de volgende ochtend bleek dat het slachtoffer wel degelijk had ingestemd met de seksuele handelingen.[17] Een officieel onderzoek naar het falen van het verlenen van inzage in ongebruikt bewijsmateriaal[18] wijst ook de gebrekkige ICT-systemen bij de Politie en de *Crown Prosecution Service* als een van de oorzaken aan en beveelt aan om al het ongebruikte digitale bewijs in het vervolg op één centrale locatie in het CPS-systeem op te slaan.

De zoekmachine Hansken biedt in meerdere opzichten een oplossing voor de problemen waar onze collega's in Engeland mee worstelen. Nederland lijkt hiermee voorop te lopen in vergelijking met andere landen maar kan zeker nog niet achteroverleunen. De contra-expert die is ingeschakeld door de verdediging in de *Tandem II*-zaak, heeft gelijk als hij stelt dat het filteren met zoekwoorden zeer beperkt is. De rechter concludeert echter terecht dat de expert nergens concreet betwist dat de resultaten onbetrouwbaar zijn. De ontwikkelingen in commerciële *E-Discovery*-tools leren ons dat er nog veel valt te verbeteren en dat we met nieuwe technologie de huidige tools nog slimmer kunnen maken. Maar het belangrijkste waar het nu om gaat, is dat er eindelijk zicht is op een aanpak waarmee rechercheurs beter kunnen zoeken, beter kunnen samenwerken met elkaar en met het OM en waarbij bovendien de verdediging in de gelegenheid gesteld kan worden om met behulp van dezelfde technologie, weliswaar met restricties, ook onderzoek te doen. Niet alleen in het bewijs dat het OM relevant vindt maar ook in het overige bewijs dat in beslag genomen is, waarbij de rechter-commissaris uitzonderingen mag maken.

13. Zie p. 90-95 www.cps.gov.uk/legal-guidance/disclosure-manual.
14. linkedin.com/pulse/from-crime-court-review-principles-uk-disclosure-hans-henseler/.
15. theguardian.com/society/2020/feb/16/impact-on-victims-of-police-phone-seizures-to-be-reviewed.
16. london.gov.uk/what-we-do/mayors-office-policing-and-crime-mopac/governance-and-decision-making/mopac-decisions-0/forensics-remote-search-and-review-pathfinder.
17. independent.co.uk/news/uk/crime/rape-trial-collapse-sex-text-messages-police-funding-cuts-liam-allan-disclosure-phone-innocent-miscarriages-justice-a8113011.html.
18. justiceinspectorates.gov.uk/cjji/inspections/making-it-fair-the-disclosure-of-unused-material-in-volume-crown-court-cases/.