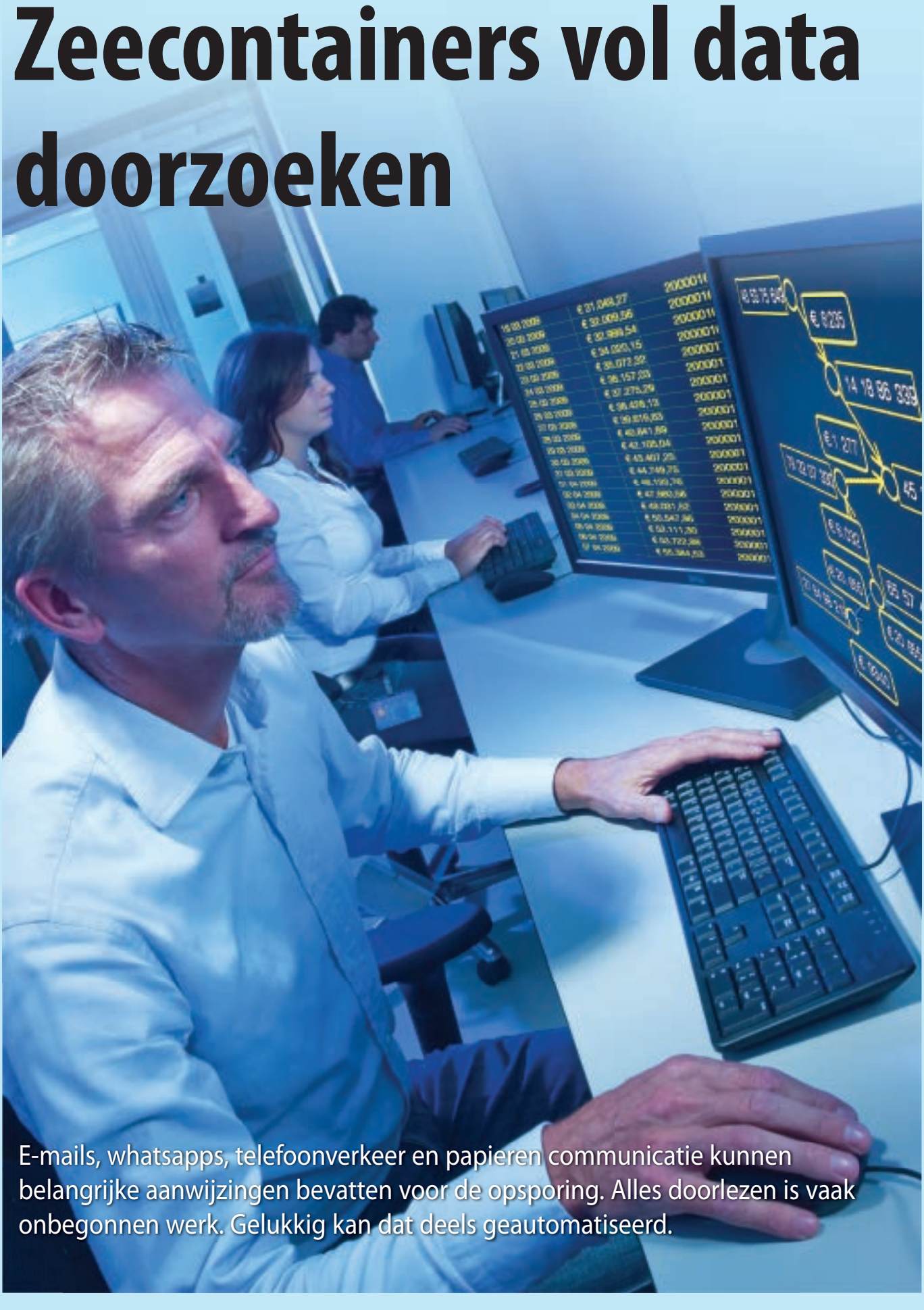
Zeecontainers vol data doorzoeken



E-mails, whatsapps, telefoonverkeer en papieren communicatie kunnen belangrijke aanwijzingen bevatten voor de opsporing. Alles doorlezen is vaak onbegonnen werk. Gelukkig kan dat deels geautomatiseerd.

Tekst: Joan Kurpershoek · **Foto's**: NFI · **Infographic**: MediagraphiX

De afdeling Digitale Opsporing van de eenheid Rotterdam heeft in korte tijd ingewikkelde netwerken van verdachten in kaart kunnen brengen, in samenwerking met het Kennis- en Expertisecentrum voor Intelligente Data-analyse. Dit centrum, kortweg 'Kecida', is onderdeel van het Nederlands Forensisch Instituut. "In één van die zaken hebben zij data uit mobiele telefoons geanalyseerd," zegt digitaal rechercheur Robert van Voorst, "maar dan specifiek op whatsapp-berichten. Wij wilden een analyse van het sociale netwerk tussen een aantal verdachten: wie had contact met wie, hoe lagen de verhoudingen, wie was de spin in het web? Kecida wist dat binnen een paar dagen inzichtelijk te maken. Hun conclusie kwam deels overeen met die van analisten bij de recherche, die al drie maanden onderzoek hadden gedaan. Behalve dat Kecida het dus veel sneller deed, bood het ook extra informatie over de contacten, bijvoorbeeld van wie het meeste initiatief uit ging. Ze hebben dat ook gevisualiseerd. In een paar plaatjes lieten ze zien hoe het netwerk van de persoon was opgebouwd. Dat is erg verhelderend, je haalt gelijk de kernfiguren eruit."

Fraude Naast telefoondata doorzoekt, rangschikt en analyseert Kecida ook allerlei andere soorten bestanden, zegt teamleider Menno Israël. Zoals e-mails en oude, gedigitaliseerde dossiers. Of bankoverschrijvingen: welke personen en rekeningnummers staan in verband met opvallende geldbedragen, schulden of transacties? Zo wist Kecida bijvoorbeeld een 'carrouselfraude' in kaart te brengen, met miljoenen geldtransacties tussen buitenlandse en Nederlandse bedrijven. Daarbij vroegen bedrijven onterecht miljoenen euro's aan BTW terug.

"Wij zoeken naar opvallende zaken", zegt Israël. "Uitzonderingen eigenlijk, zoals afwijkend woordgebruik in documenten. Sommige woorden vallen op omdat ze niet passen bij het taalgebruik binnen de betreffende organisatie. Een simpel voorbeeld is 'pepernoten' als codewoord voor geld. Een ander voorbeeld: in e-mailwisseling werd gesproken over appels, sinaasappels en T-shirts. Het woord T-shirts valt op. Rechercheurs hadden appels en sinaasappels op de traditionele, handmatige manier ontdekt. Wij hebben nu een module ontwikkeld om T-shirt als code- of sluiertaal te detecteren, want zoekend met trefwoorden red je dat niet. Je weet immers niet waarnaar je moet zoeken. Het algoritme dat we gebruiken, is behoorlijk complex maar in feite een verdere uitwerking van: 'Dit woord komt normaal gesproken niet of heel weinig voor in combinatie met deze andere woorden.'"

Geheugen In omvangrijke zaken lijkt het voor rechercheurs onbegonnen werk om dat soort dingen zelf uit te zoeken. Helemaal als het gaat om miljoenen mailtjes, terwijl de verdachte organisatie ook nog eens regelmatig de 'sluiertaal' verandert. Israël: "Al werken er tien rechercheurs of analisten aan zo'n zaak, dan nog is het ingewikkeld. Niemand heeft overzicht van alle informatie. Dan ben je afhankelijk van het geheugen van die ene rechercheur die in één van die mailtjes iets heeft gelezen. Stel, er zijn honderd zaken die in verschillende regio's plaatsvonden, maar vermoedelijk wel met elkaar te maken hebben. De dossiers beslaan zo'n tienduizend pagina's en vijf rechercheurs lezen allemaal een deel daarvan. Een rechercheur komt één keer de naam Manke Teun tegen, omdat een team bij een observatie heeft gezien dat deze met een verdachte meeliep. Dan lijkt die Teun niet belangrijk. Tenzij blijkt dat hij in zaak drie, vijf en twaalf óók een keer voorkwam. Maar als die zaken destijds door andere recherche-teams zijn behandeld, bestaat het risico dat die informatie nooit naar boven komt."

Overzicht "Met intelligente data-analyse krijg je dat overzicht wel", vervolgt Israël. Mede doordat Kecida in staat is verschillende informatiebronnen aan elkaar te koppelen. "Door informatie uit navigatiegegevens van een auto - betrokken bij een ernstig misdrijf - te combineren met historische telefoongegevens van verdachten, hebben wij kunnen achterhalen wie er in die auto reed die drie maanden daarvoor was gestolen. De navigatiegegevens uit de TomTom, locaties en tijdstippen, en de telefoongegevens waren twee onafhankelijke databronnen, met elk veel registraties in een korte periode."

Zo'n onderzoeksmethode moet wel kloppen, benadrukt hij. "Die mag niet straks voor de rechter onderuitgehaald worden, doordat de advocaat zegt: 'Jullie wisten waarschijnlijk wat het telefoonnummer was en hebben het zo geordend dat dat nummer bovenaan kwam.'"

Het team van Kecida bestaat dan ook uit specialisten op het gebied van informatica, statistiek en kunstmatige intelligentie. En 'allrounders', bij wie alle onderzoeksresultaten samenkomen. "Je moet altijd vanuit het grotere geheel naar handelingen en scenario's toe. Dus moet je die gegevens goed kunnen combineren. Dat doen wij in een gesprek met z'n allen en de klant erbij, al noem ik de klant liever de samenwerkings- of projectpartner."

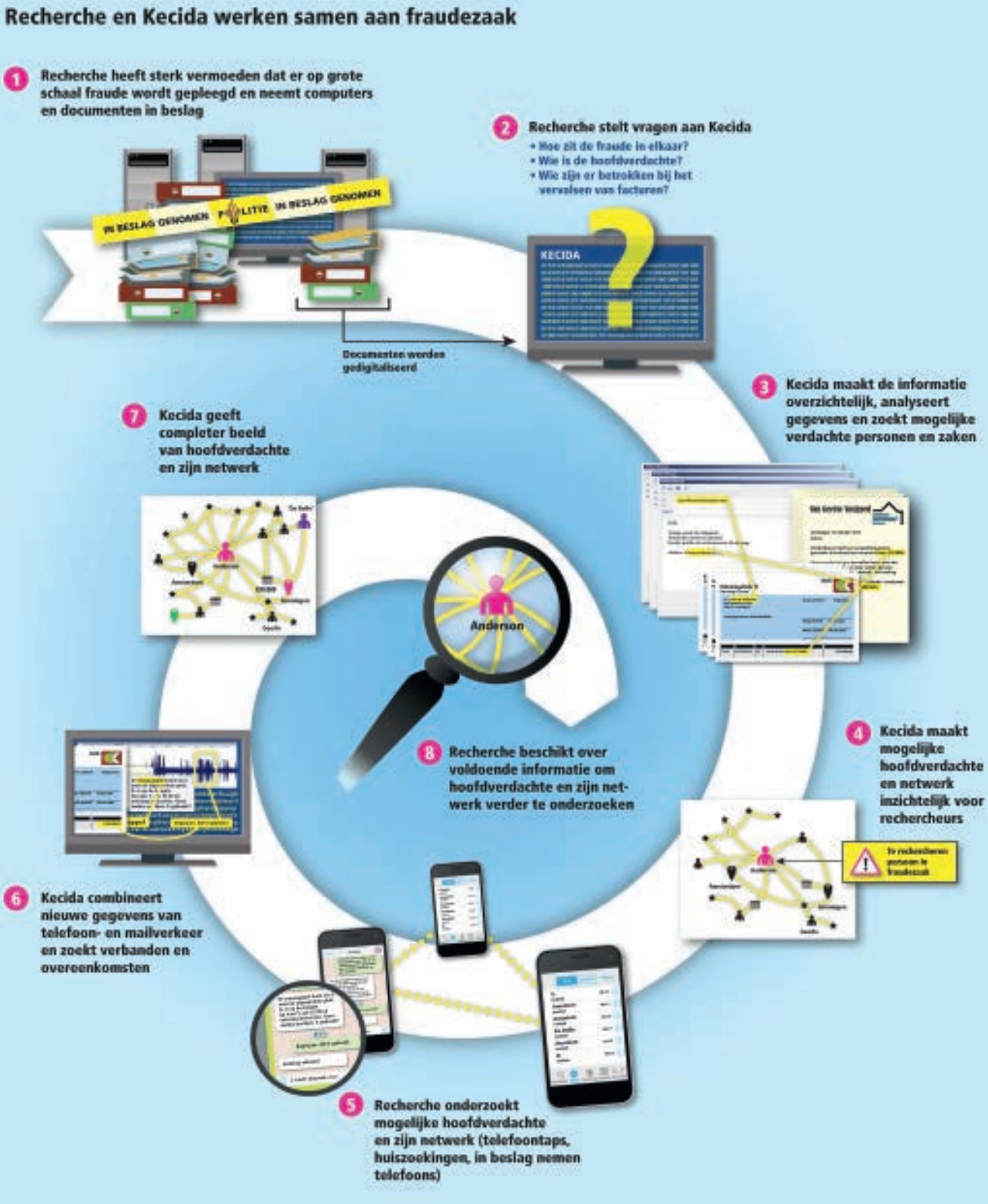
Dahliastraat Kecida maakt het makkelijker voor de rechercheur, zegt Israël. "Maar de rechercheur blijft degene die het resultaat uiteindelijk op waarde moet schatten." Om te beginnen moet de rechercheur bij het NFI de data aanleveren, die zijn veiliggesteld van bijvoorbeeld telefoons en computers. Vaak moeten die data, als ze van verschillende bronnen komen, nog omgezet worden naar hetzelfde format of worden gecorrigeerd. "Telecomproviders hebben allemaal andere formats en uitdraaien. Vaak zitten er fouten in documenten die met optical character recognition (OCR) zijn gescand, en soms kloppen adressen gewoon niet. Je houdt niet voor mogelijk op hoeveel manieren Dahliastraat geschreven kan worden! Dat soort data-intensieve zaken kosten altijd de meeste tijd, zo'n tachtig procent van de hele onderzoeksduur. Maar dat neemt af, want de kennis die wij opdoen bij het ene onderzoek, gebruiken we weer bij het volgende." Behalve voor de politie doet Kecida dit soort onderzoeken ook voor andere opdrachtgevers, zoals de KMar, de FIOD en de Financial Intelligence Unit.

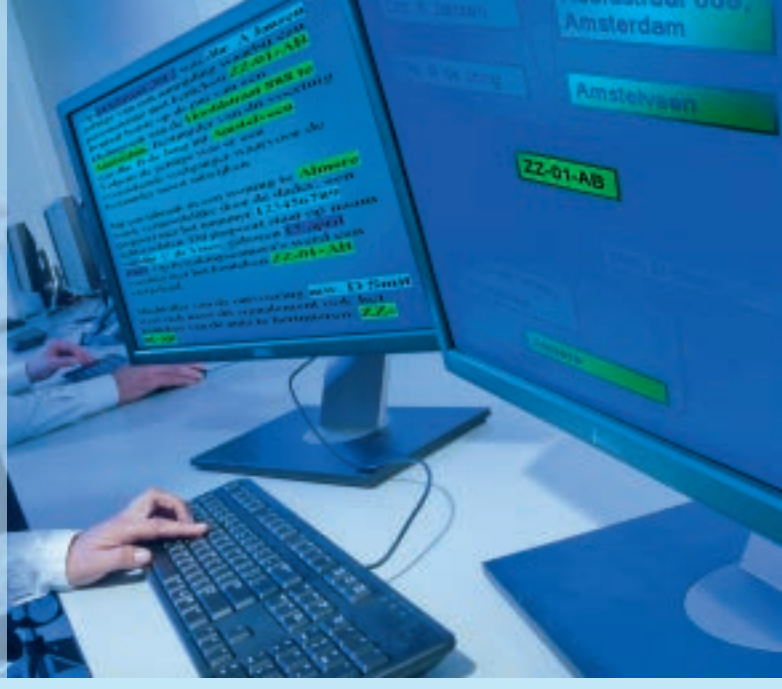
Kecida haalt met software uit alle data automatisch een aantal 'entiteiten'. "Dat zijn bijvoorbeeld persoonsnamen, organisaties,

>>

>> locaties, telefoonnummers en rekeningnummers. De politie kan die gebruiken om te zoeken naar verbanden, of om weer nieuwe entiteiten aan te geven." Soms levert de politie ook voorbeelden van interessante e-mails. Kecida bouwt daarvan een model dat de belangrijkste kenmerken vaststelt en past dit toe op de overige mails, om zo soortgelijke berichten te zoeken en op volgorde van relevantie te zetten.

Snel De tijdsduur van een data-analyseonderzoek varieert. Een snelle netwerkanalyse op basis van e-mail als indicatief onderzoek kan soms in een paar dagen, een langduriger ondersteuning in een grootschalig onderzoek duurt soms jaren. "Als iets snel moet, zijn we snel, maar we overleggen altijd met de projectpartner. Wil deze meteen een uitgebreid rapport erbij, of heeft hij liever zo snel mogelijk een uitdraai met de gegevens, omdat die nodig zijn om de





opsporingsrichting te bepalen? Dat laatste is een *quick win,* dat kan binnen één of twee weken."

Bij een kinderpornozaak in Rotterdam was ook haast geboden. Van Voorst: "Mijn collega's en ik wilden heel snel weten of de verdachte contacten had met mogelijke kindermisbruikers. Die wil je namelijk zo snel mogelijk uit het circuit hebben, zodat zij geen slachtoffers meer maken. Op iemand die zelf geen kinderen misbruikt, kun je later actie ondernemen. Omdat we zagen dat er heel veel contacten waren, hebben wij niet meteen het hele sociale netwerk in kaart gebracht, maar ons gefocust op de vraag: Welke van deze contacten kan een kindermisbruiker zijn? Iemand kan in een e-mail of een chat spreken over kindermisbruik, of daar een indicatie voor geven. Maar het kan ook zijn dat er relatief nieuwe bestanden worden uitgewisseld, nieuw of onbekend materiaal. Je gaat af op een combinatie van factoren. Samen met Kecida bedenk je eerst een aantal indicatoren die aangeven welke contacten in dat netwerk je als extra verdacht moet aanmerken. Zij kunnen dat heel mooi bekijken met analyses. Daar komt dan bijvoorbeeld uit: 'Wij hebben vijfhonderd contacten gevonden en zoveel aan die contacten gerelateerde informatie. Op basis van die informatie zou je moeten beginnen met deze tien.' Dat is geen hard gegeven op basis waarvan je iemand meteen moet aanhouden, maar wel een belangrijke aanwijzing in de goede richting."

Xiraf De uitslag van Kecida was er in deze zaak binnen een paar dagen. Van Voorst: "We hadden een paar dagen winst doordat wij binnen de eenheid Rotterdam zelf het systeem Xiraf hebben staan. Xiraf indexeert data en maakt ze doorzoekbaar. Het NFI is eigenaar van dat systeem en biedt die dienst aan. Bij complexere zaken verdient het zeker aanbeveling om je data bij het NFI aan te bieden. Maar bij ons hoefde dat niet apart."

Elk onderzoek is maatwerk, volgens Menno Israël. "Wij maken een soort road map voor de informatieanalyse: deze informatie heb je, dat kan er uitkomen, maar dan moet je eerst dit nog doen." Dat kan zelfs resulteren in het ontwerpen van nieuwe hard- en/of software voor meer specifiek onderzoek op een bepaald gebied. "Voor het onderzoek naar de carrouselfraude bijvoorbeeld hebben we een 'carrouseldetector' gebouwd. Die bekijkt alle transacties en selecteert elke cirkel, elke kleine carrousel. De basis is een samenspel van traditionele zoekalgoritmes met kunstmatige intelligentie én het nodige maatwerk. We denken in termen van functionaliteit voor het vinden van bepaalde informatie: wat werkt het beste gegeven de opdracht? Dat kan soms heel iets anders zijn dan de projectpartner denkt."

Actie "Met *big data analytics* voorziet Kecida in behoeften op operationeel, strategisch en beleidsniveau", zegt Israël. "Het geeft inzicht in grote hoeveelheden data en eventuele patronen van activiteit. Dat komt neer op reconstrueren wat er is gebeurd, op basis van de sporen die iemand achterlaat. Dit inzicht maakt duidelijk wat de beste actie is en ondersteunt die beslissing."

Soms is het mogelijk aan de hand van data verschillende mogelijke scenario's te schetsen. "Bijvoorbeeld als er een moord is gepleegd", zegt Israël. "Je hebt een dader en een slachtoffer op dezelfde plaats en tijd, plus een moordwapen. Met dat soort gegevens kun je een berg informatie doorzoeken en dat levert een aantal mogelijke scenario's op. Dat is nog in ontwikkeling, maar we passen het soms al toe."

Container Patronen en verbanden kunnen herkennen in grote hoeveelheden data spreekt de afdeling Digitale Opsporing in Rotterdam ook wel aan. "De hoeveelheid informatie, bijvoorbeeld bij TGO's, is de laatste jaren explosief gestegen doordat we steeds meer bronnen aanboren", zegt Robert van Voorst. "Denk aan smartphones, tablets, sociale netwerken en dergelijke. Daardoor krijg je ook een informatieoverflow: je ziet door de bomen het bos niet meer. Ik vergelijk het met een grote zeecontainer helemaal vol papier, die je moet onderzoeken. Je raapt hier en daar een papiertje op. Is het wat, dan bewaar je het; zo niet, dan gooi je het weg. Maar eigenlijk zou je van tevoren iemand die container in willen sturen die zo snel mogelijk alles rangschikt in emails, chats enzovoorts. En die dan de meest relevante tien voor je klaarzet. Het zou mooi zijn als we kleinschalige tools kunnen standaardiseren, die collega's kunnen gebruiken zonder dat ze direct naar het NFI toe hoeven."

Israël waarschuwt in dat opzicht echter voor te veel optimisme, omdat intelligente data-analyse 'complexer is dan het soms lijkt'. Van Voorst beaamt dat: "Maar het NFI en wij denken wel na over wat je zelf kan doen en wat maatwerk moet zijn van Kecida. In ons vakgebied is het altijd maatwerk. Maar wat nu nog heel exclusief is, kan over tweeënhalf tot vijf jaar bij wijze van spreken generiek zijn. En dan is het mooi als Kecida kan zeggen: 'Daar hebben we nu een standaard tool voor, dus dat kun je voortaan zelf. Dat leren wij uit.' Dan kunnen zij zich weer richten op nieuwe ontwikkelingen."

Redactie.blauw@politieacademie.nl

Voor meer informatie:
kecida@nfi.minvenj.nl
PKN > Forensische opsporing > Digitale opsporing