INFORMATIE DIE WERKT!

Slimmer omgaan met (grote hoeveelheden) informatie
in de veiligheidsketen

Resultaten 2011 | TNO

TNO innovation for life

INFORMATIE DIE WERKT!

Slimmer omgaan met (grote hoeveelheden) informatie in de veiligheidsketen
Resultaten 2011

TNO
Oude Waalsdorperweg 63
2509 JG Den Haag

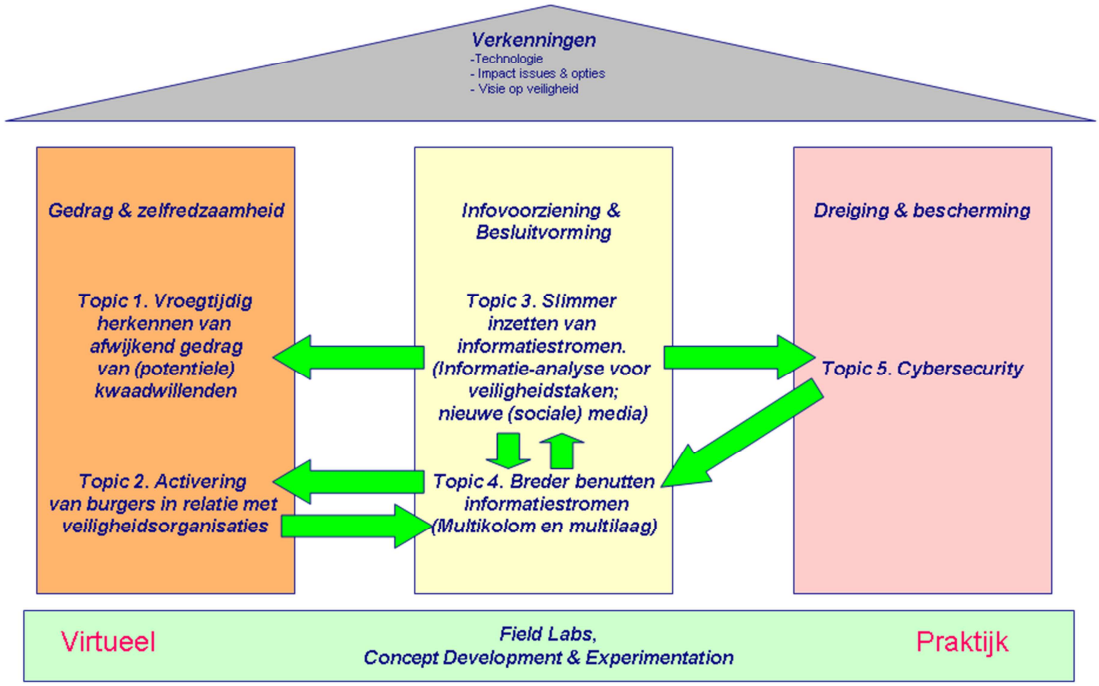
TNO.NL

Redactie
Karin de Jong | karin.y.dejong@tno.nl
Lotte van Lier | lotte.vanlier@tno.nl

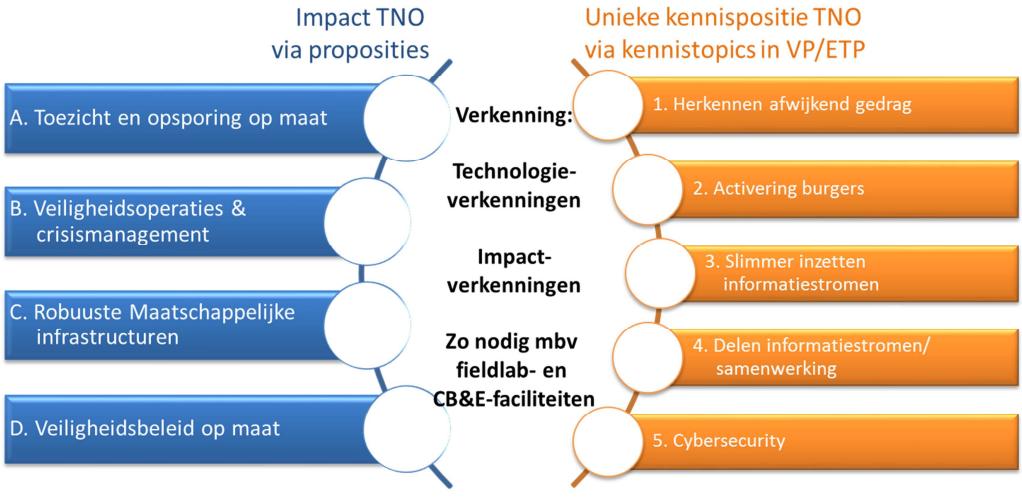
Drukwerk
Repro TNO, Den Haag

Afbeeldingen
Gettyimages.com
Noordhollandsdagblad.nl
NOS.nl
Spitsnieuws.nl
TRECVID
Wikipedia Common

Inhoudsopgave



Figuur 1 In overleg met het Ministerie van V&J, stakeholders en TNO geselecteerde topics voor VP Veilige Maatschappij 2011-2014 en hun onderlinge relatie



Figuur 2 Overzicht van proposities en topics van het programma Veilige Maatschappij

1 Inleiding

Deze publicatie neemt u in vogelvlucht mee langs de resultaten van het onderzoek dat in 2011 door TNO is verricht op het onderwerp: Slimmer omgaan met grote hoeveelheden informatie. Dit onderzoek is onderdeel van het programma Veilige Maatschappij dat TNO in opdracht van het Ministerie van Veiligheid en Justitie uitvoert.

Het is bedoeld voor alle stakeholders (zie figuur 4) op gebied van openbare orde en veiligheid. Door het lezen van dit boekje krijgt u zicht op de kennis die beschikbaar en in opbouw is en op welke wijze dit binnen het domein van openbare orde en veiligheid zijn toepassing kan vinden.

PROGRAMMA VEILIGE MAATSCHAPPIJ

In het TNO-Innovatiegebied Veilige Maatschappij is het onderzoek dat TNO uitvoert in de strategieperiode 2011-2014 gericht op de volgende impact:

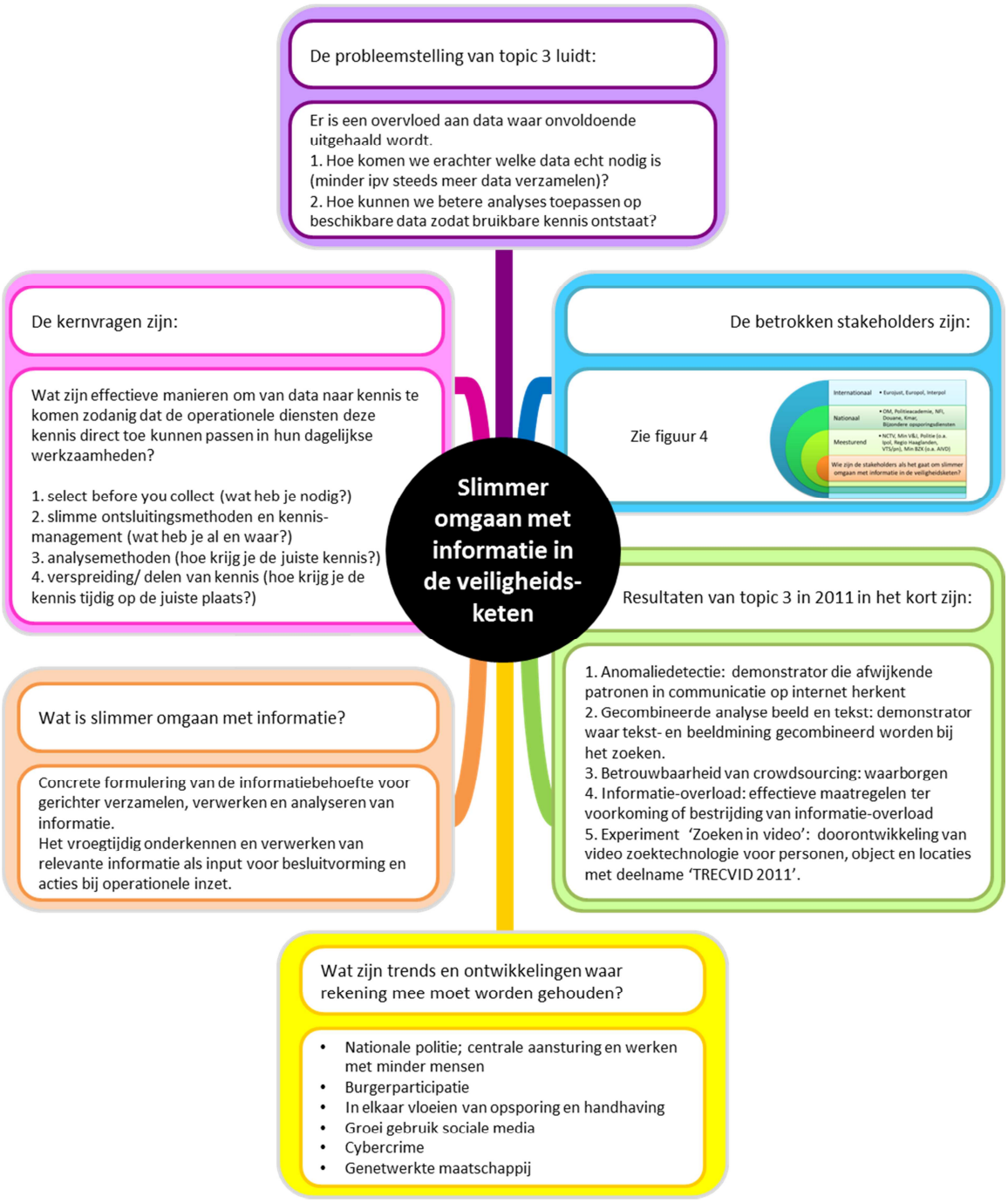
1. Effectievere/ efficiëntere veiligheidsoperaties en rampenbestrijding door innovatie van informatievoorziening, besluitvorming en organisatie van de samenwerking tussen organisaties;
2. Effectiever/ efficiënter optreden van veiligheidsorganisaties door innovatie van doctrine, materieel, uitrusting en competenties;
3. Vergroting veiligheid in openbare ruimte door effectief betrekken van burgers/bedrijven;
4. Beperking schade ten gevolge van rampen/crises (overstromingen, CBRNe-incidenten, uitval vitale infrastructuren) door snel en adequaat optreden en door kosteneffectieve pro-actieve maatregelen.

Om deze impact te bereiken zijn vier proposities opgesteld (A –D in figuur 2). In overleg met regievoerder Min V&J en behoeftestellers zijn hier vijf topics aan gekoppeld (1-5 in figuur 2) die ieder voor zich van belang zijn voor meerdere proposities. Tezamen vormen de topics het programma Veilige Maatschappij.

Deze publicatie richt zich op topic 3. Het onderwerp betreft het slimmer inzetten van beschikbare informatie en nieuwe (sociale) media voor operationele veiligheidstaken. Een belangrijke focus is de verwerking van grote hoeveelheden aanwezige informatie, waarnemingen en meldingen tot bruikbare kennis. Dit topic richt zich met name op de onder punt 1 en 2 genoemde velden waarin de impact zichtbaar moet worden (Toezicht en opsporing op maat (A) en veiligheidsbeleid op maat (D)).

LEESWIJZER

In deze publicatie vindt u een compacte beschrijving van de resultaten van het onderzoek dat in 2011 rondom topic 3 is verricht. Eerst worden in hoofdstuk 2 de onderzoeksvragen beschreven. In hoofdstuk 3 komt het belang van de samenwerking met stakeholders aan de orde. De onderzoeksvragen zijn in 5 deelprojecten uitgewerkt. Per deelproject treft u in hoofdstuk 4 de belangrijkste resultaten aan. Tot slot is er in hoofdstuk 5 een korte vooruitblik naar de plannen voor 2012.



Figuur 3 Slimmer omgaan met informatie in de veiligheidsketen

2 Slimmer omgaan met (grote hoeveelheden) informatie in de veiligheidsketen

In de veiligheidsketen is het noodzakelijk dat grote hoeveelheden informatie snel verwerkt worden tot bruikbare kennis. De hoeveelheid beschikbare informatie neemt de afgelopen jaren explosief toe, waardoor doorgaan op de huidige (veelal niet-geautomatiseerde) wijze van informatieontsluiting een te grote inzet van de menskracht zou vergen. Het is de vraag in hoeverre meer informatie op dit moment ook meer kans op een geslaagde aanhouding of interventie opleveren. Meer is niet altijd beter en wegen de kosten voor het verzamelen en analyseren van informatie wel op tegen de resultaten? De uitdagingen waarmee de spelers in de veiligheidsketen worden geconfronteerd hebben betrekking op een overvloed aan informatie (informatie-overload), een tekort aan menselijke verwerkingscapaciteit (personele krapte) en de noodzaak om proactief te analyseren en te beslissen (sneller voorin de keten komen). Daarnaast is er aansluiting nodig op de paradigma verschuiving die nieuwe en sociale media als vorm van communicatie met zich meebrengen; communicatie vindt in steeds mindere mate hiërarchisch of lineair plaats, maar juist diffuus door en met het publiek.

In 2011 heeft dit zich vertaald in een vijftal projecten:

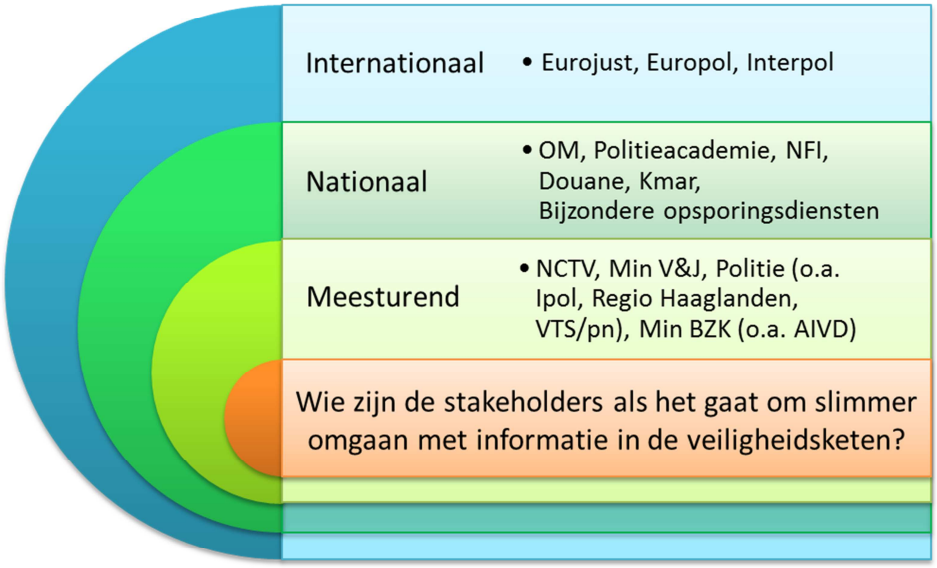
1. **Informatieanalysetechnieken:** Op welke wijze kunnen nieuwe ontwikkelingen op gebied van informatieanalysetechnieken (video content analysis, datamining, semantic annotation, collaborative tagging, etc.) een bijdrage leveren aan het verbeteren van het informatieproces van de domeinen handhaving en opsporing?

2. **Social Media en betrouwbaarheid:** Op welke wijze kan nieuwe media een rol spelen in het slimmer omgaan met informatie in de domeinen handhaving en opsporing?

3. **Informatie–overload:** Welke ontwikkelingen (gesteld in de vorm van werkhypothesen) zijn veelbelovend voor het verminderen van de informatie-overload bij de uitvoerende spelers en het vergroten van de efficiëntie van de processen binnen de domeinen handhaving en opsporing?

4. **Privacy**

5. **Internationale benchmark video content analyse**



Figuur 4 Overzicht van de betrokken stakeholders

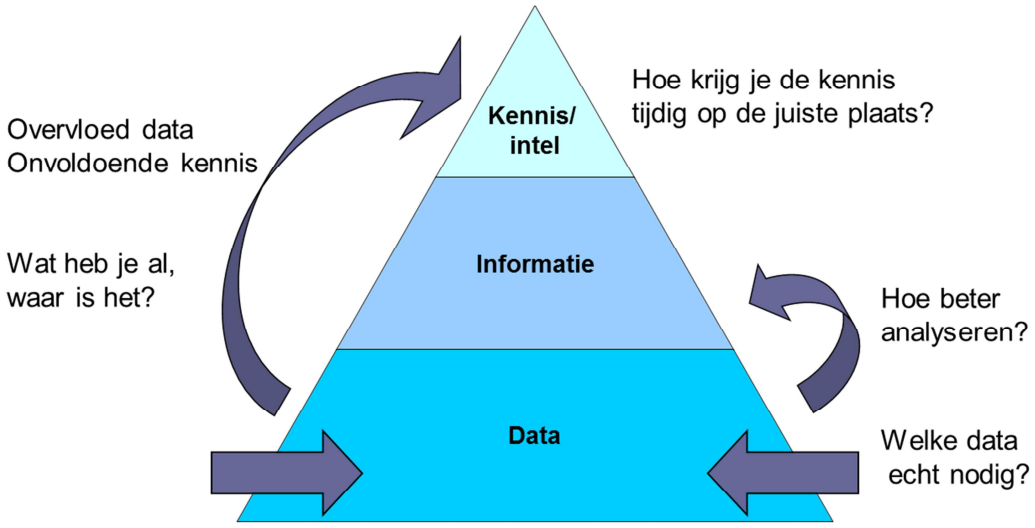
3 Partners

INTENSIVERING SAMENWERKING

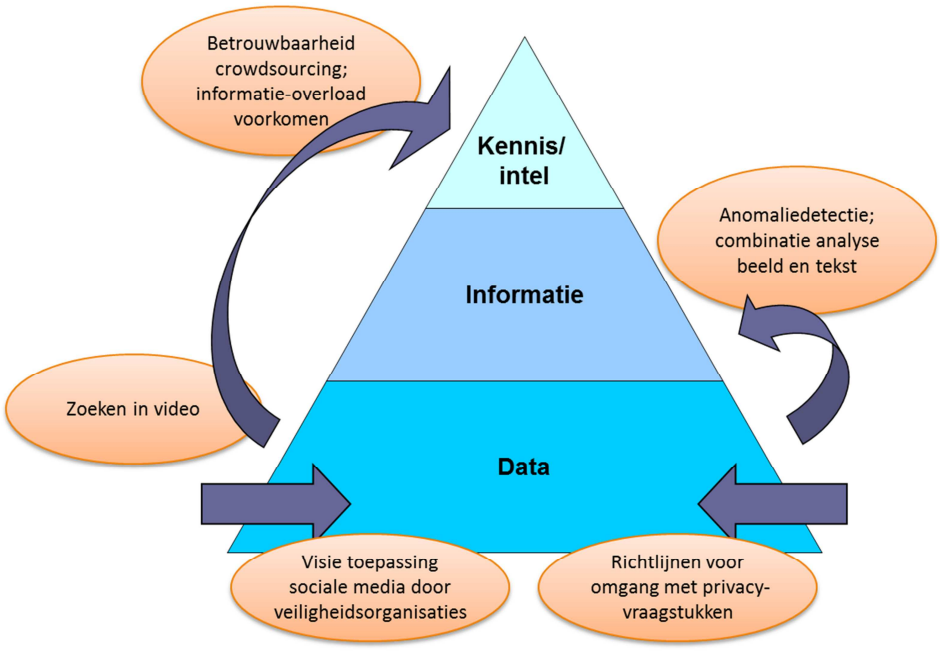
Als gevolg van de activiteiten binnen topic 3 is de samenwerking tussen TNO en de diverse stakeholders versterkt: zo zijn er contacten gelegd met het IRN (internet recherchenetwerk) wat heeft geleid tot samenwerking bij het opstellen van een voorstel voor analysetooling voor het IRN; met het PAC (programma aanpak cybercrime) waarbij de projecten zijn afgestemd en opgelijnd over en weer en het programma HDIef (Digital Fingerprinting) van de NCTV. Daarnaast is een samenwerkingsovereenkomst gesloten tussen TNO en KLPD/IPOL waarin diverse initiatieven zijn opgestart voor gezamenlijk onderzoek. Ook zijn diverse gesprekken gevoerd met de politieacademie en het KLPD om mogelijkheden voor verdergaande samenwerking in 2012 te verkennen.

KENNISDELING

Het netwerk dat is en wordt opgebouwd maakt het mogelijk dat de kennis die binnen het onderzoeksprogramma wordt opgebouwd snel aan de relevante stakeholders beschikbaar kan worden gesteld en kan worden getoetst of dit ook voldoet aan de behoeften.
Op 12 oktober 2011 is door SMVP/CCV een congres georganiseerd rondom het thema publiek-private samenwerking (wie maakt Nederland veiliger?). Op dit congres heeft TNO vanuit topic 3 een workshop verzorgd rondom het onderwerp toepassing van sociale media bij opsporing en handhaving.



Figuur 5 De Informatiepiramide: data wordt verrijkt tot kennis



Figuur 6 Plaats van de werkpakketten in de informatiepiramide

4 Resultaten 2011

De vijf projecten die in 2011 zijn uitgevoerd lijken zeer divers van aard. Ze hebben echter allen een plaats in de informatiepiramide (zie figuur 6).
Binnen topic 3 zijn vijf projecten gestart: Informatieanalysetechnieken, Social Media en betrouwbaarheid, Informatie-overload, privacy en internationale benchmark video content analyse. In dit hoofdstuk zijn per project het doel, de resultaten en de toegevoegde waarde voor veiligheidsprofessionals beschreven.

INFORMATIE DIE WERKT!

4.1 INFORMATIEANALYSETECHNIEKEN

Contactpersoon: Henri Bouma

Hoe vind je “een speld in de hooiberg” als je daarbij niet weet hoe de speld eruit ziet? Dit is het probleem van diverse stakeholders die elke dag worden geconfronteerd met grote hoeveelheden data. Er is zoveel data dat hierin de bruikbare informatie (b.v. ‘iets’ afwijkends) niet of moeilijk kan worden gevonden.
De behoefte binnen de domeinen handhaving en opsporing is om in grote hoeveelheden data efficiënt individuen en groepen te kunnen vinden op basis van kenmerkende eigenschappen, om de kenmerkende eigenschappen te scheiden van overige eigenschappen, om gerelateerde zaken in kaart te brengen, om afwijkingen te detecteren, en om trends en veranderingen te monitoren. TNO ontwikkelt daarvoor toegesneden technologie die het proces en de mens zo goed mogelijk ondersteund.

Hierbij is speciale aandacht voor de volgende zaken:

- Impact: Zoeken naar oplossingen die nauw aansluiten bij de hierboven geformuleerde behoeften en een aantal daarvan te tonen met een proof-of-concept demonstrator.
- Innovatief: Onderscheiden van reeds bestaande producten en er op richten om (op termijn) toegevoegde waarde te creëren voor de gebruikers.
- Herbruikbaar: Ontwikkelen van technieken die nu en in de toekomst herbruikbaar zijn voor meerdere stakeholders om hun behoeften te vervullen.
- Multi-modaal: Aantonen dat de combinatie van informatie uit verschillende modaliteiten veel extra kennis kan opleveren door niet alleen gebruik te maken van tekst maar ook van beelden.

ONDERZOEKSVRAAG EN WERKHYPOTHESE

De belangrijkste onderzoeksvraag voor dit project is: op welke wijze kunnen ontwikkelingen op gebied van informatieanalysetechnieken (e.g., video content analyse, data mining, etc) een bijdrage leveren aan het verbeteren van het informatieproces van de domeinen handhaving en opsporing.

De centrale werkhypothese is dat automatische informatieanalysetechnieken de effectiviteit verhogen van handhaving en opsporing. Hierbij is gefocusseerd op de volgende twee aspecten:

- Anomaliedetectie: Afwijkingen / anomalieën detecteren op het internet.
- Beeldanalyse: Aantonen dat de combinatie van tekst met beeld op het internet extra kennis oplevert.

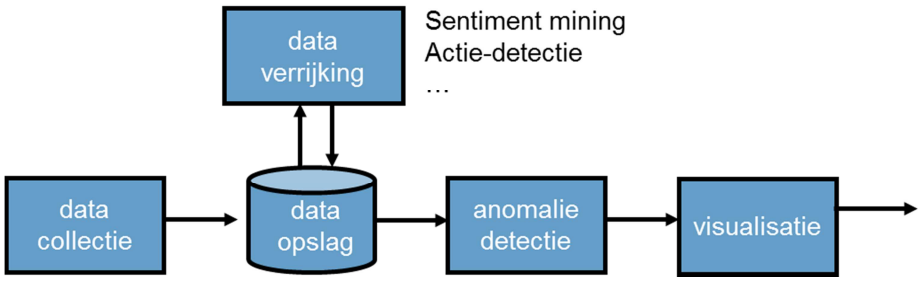
Als database / databron is gekozen voor (een aantal sites op) het internet om de volgende vijf redenen:

1. Het internet is een vrij toegankelijke openbare bron.
2. Het is een enorme multi-modale database die veel informatie bevat (een hooiberg).
3. Om database specifieke implementatie- en ontwikkelproblemen te vermijden zodat herbruikbare modules kunnen worden ontwikkeld.
4. Om fragmentatie te voorkomen en samenwerking met het project dat zich richt op sociale media mogelijk te maken.
5. Omdat diverse stakeholders geïnteresseerd zijn in technieken voor internetobservatie.

Specifiek zijn er middelen ontwikkeld om berichten op Twitter te analyseren.

ANOMALIEDETECTIE

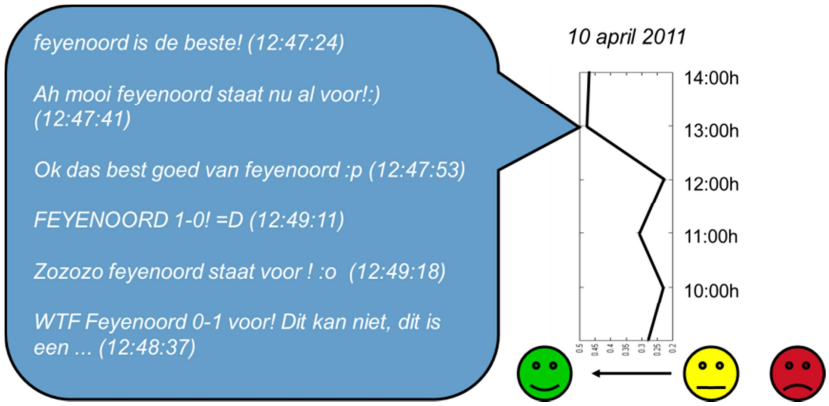
Voor anomaliedetectie is een verwerkingsketen opgezet die bestaat uit datacollectie, dataopslag, dataverrijking, anomaliedetectie en visualisatie.



Figuur 7 Verwerkingsketen voor anomaliedetectie

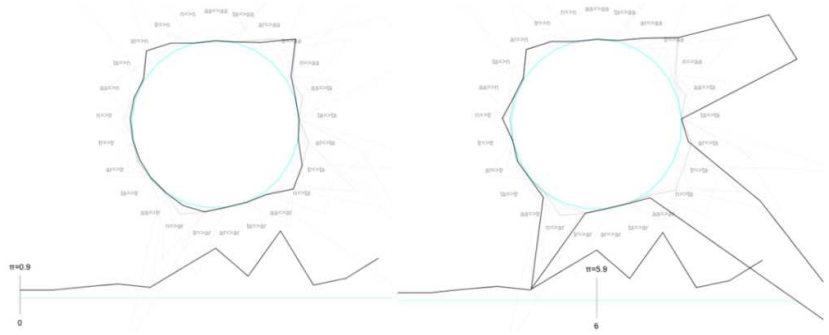
Om de ontwikkeling en validatie op relevante data te laten plaatsvinden is een tool gemaakt die zowel de afgelopen periode als de komende periode data van Twitter verzameld en opslaat. Hierdoor is het mogelijk om na het plaatsvinden van een incident nog te analyseren of dit te voorspellen was.
Het detecteren van anomalieën is uitdagend omdat: afwijkingen zeldzaam zijn, sommige afwijkingen later normaal worden, de grens tussen normaal en afwijkend soms dun is, en omdat er vaak verschillende perspectieven op data mogelijk zijn (wat is de beste context?). Daarom is het systeem geen ‘black box’; de analist definieert zelf ‘meetvariabelen’, waarna het systeem de interactie tussen deze variabelen onderzoekt.

De data bestaat uit tekst en metadata, zoals activiteit en relaties. Daarnaast worden nog kenmerken toegevoegd op basis van tekst interpretaties, zoals bijvoorbeeld sentiment analyse. Er worden automatisch clusters en correlaties berekend. Een anomalie wordt herkend als een abrupte verandering in het correlatiepatroon van meerdere variabelen. Hieronder staat een voorbeeld van een plotselinge verandering tijdens een voetbalwedstrijd.



Figuur 8 Sentimentanalyse

Een ander voorbeeld is een rel tussen Turken en Koerden in Amsterdam in oktober 2011. De twitterberichten tonen aan dat er een grote toename in activiteit is geweest (o.a. retweets en reacties) in deze context.



Figuur 9 Visaulisatie van actviteit

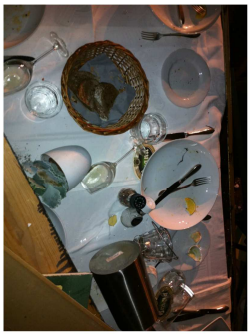
In de bovenstaande figuren wordt de output van het systeem getoond. De linker afbeelding is voorafgaand aan de rel en de rechter afbeelding is tijdens de rel. De afwijking van het gemiddelde per variabele wordt getoond in een Spider diagram, waarbij de cirkel normaal is. Daaronder wordt de anomaliecurve getoond, die de afwijkingen combineert tot één score voor elk tijdsmoment.
Veranderingen en afwijkingen kunnen met dit systeem worden geanalyseerd. Ook kunnen veranderingen van een individu, groep of onderwerp worden waargenomen over tijd en ten opzichte van anderen in dezelfde context worden beoordeeld.

BEELDANALYSE

De geanalyseerde berichten hebben een multimodaal karakter: naast tekst kunnen ook afbeeldingen en filmpjes met geluid worden toegevoegd. Het complete bericht vormt een uiting waarvan we de betekenis willen analyseren. Als aanvulling op tekstgebaseerde analyse, hebben wij onderzocht of extra informatie uit bijgevoegde afbeeldingen kan worden geëxtraheerd. In dit onderzoek is er vanuit gegaan dat de tekst van een bericht zelf niet beschrijvend hoeft te zijn, maar dat er naar een betekenisvol bijgevoegde afbeelding wordt gerefereerd. Ter illustratie volgen twee voorbeelden.

In het volgende fictieve bericht bestaat de tekst uit: "dit willen wij weer doen" samen met een plaatje van de Twin Towers.

dit willen wij weer doen



ongeluk

ongeluk

Figuur 10 Een fictieve bericht en twee berichten van twitter die verwijzen naar een ongeluk

De tekst zelf bevat weinig informatie voor sentimentanalyse of anomaliedetectie: het is een verwijzing naar het plaatje. De afbeelding zelf heeft wel een lading, en kan relevant zijn in een bepaalde context.

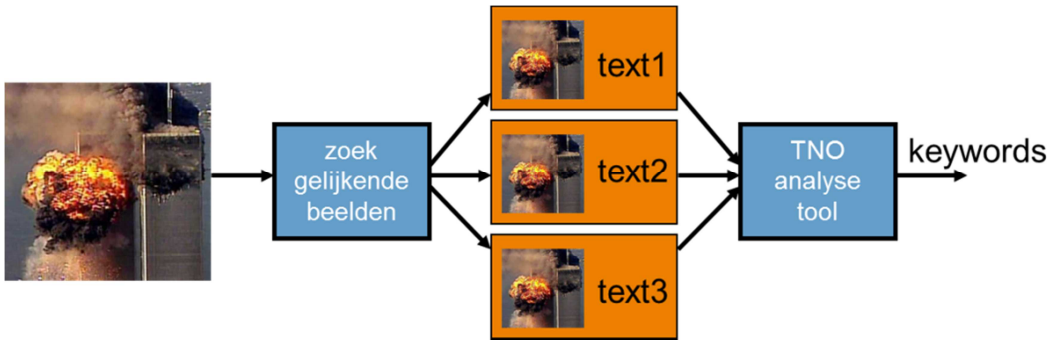
Een ander voorbeeld zijn twee berichten van twitter met de tekst: "ongeluk!". Eén bevat een foto van een verkeersongeval, de ander van een tafel met omgevallen servies. De eerste is voor de stakeholders mogelijk van interesse om getuigen te achterhalen (degene die het bericht verstuurde weet mogelijk meer), terwijl er maar weinig interesse zal zijn voor het tweede bericht.

Om in grote hoeveelheid berichten met afbeeldingen automatisch interessante gevallen te selecteren moeten we weten wat er in de afbeeldingen gebeurd. Zaken die bijzondere aandacht vragen zijn logo's, teksten in beelden, persoondetectie of -herkenning, naakte personen, shirts van een groep, avatars, maar ook abstractere zaken zoals agressie.

De bovenstaande voorbeelden tonen een tweetal aspecten die van belang voor beeldanalyse methoden voor het achterhalen van de betekenis:

- gaat het om nieuw of hergebruikt materiaal.
- gaat het om algemene herkenning van willekeurige concepten of om herkenning van te voren aangegeven objecten.

In het geval van de Twin Towers is er gebruikt fotomateriaal opnieuw geplaatst. Door gebruik te maken van een commerciële service, genaamd TinEye, kan hergepubliceerd beeldmateriaal worden gevonden op internet, mits het in hun database voorkomt. Door met deze service te zoeken naar gelijkende beelden worden ook pagina gevonden waar (een afgeleide vorm van) de afbeelding wordt gebruikt. De teksten op deze pagina's leveren met behulp van bewezen tekst-mining technieken een aantal keywords op die waarschijnlijk bij deze afbeelding horen. De keywords bij deze nieuwe pagina's worden vervolgens gebruikt om de tekstgebaseerde analyse te verbeteren.



Figuur 11 analyseproces voor extraheren van keywords

De tekst analyse extraheert alle woorden zonder tags uit de pagina's, behoudt alleen niet-stopwoorden, brengt woorden terug naar kanonieke vorm (m.b.v. Porter stemming), en maakt uiteindelijk een histogram van alle woorden. Woorden die in meerdere pagina's terugkomen, zijn waarschijnlijk verbonden met de afbeelding.

Nieuw of uniek materiaal zal niet in deze of vergelijkbare services geanalyseerd kunnen worden aangezien ze (nog) niet geindexeerd zijn. Met behulp van specifiek getrainde concept detectoren kan achterhaald worden of de afbeelding behoord tot een categorie.
Om bijvoorbeeld een set van relevante ongelukken te selecteren is er een politiewagen-detector gemaakt. Wanneer een afbeelding een (deel van een) politiewagen in beeld heeft, is deze mogelijk interessant. De gebruikte technieken voor deze strategie staan beschreven in 'videoanalysetechnieken', paragraaf 4.5.

TOEGEVOEGDE WAARDE

De twee systemen die de inhoud van hergebruikte of nieuwe afbeeldingen kunnen achterhalen, hebben beide hun sterktes en zwaktes. Het voordeel van de eerste aanpak is dat deze succesvol kan zijn zonder van te voren het aantal categorieën op te geven. Het werkt echter alleen voor bekende afbeeldingen, niet voor heel recente of unieke foto's. Een voordeel van de tweede aanpak is dat uniek en nieuw materiaal geanalyseerd kan worden. Het werkt echter alleen voor specifieke, van te voren opgegeven concepten.

Uit onze eerste tests blijkt dat er op Twitter nauwelijks hergebruikt materiaal wordt geplaatst. In onze database staan honderden afbeeldingen die allemaal uniek geschoten materiaal lijken te bevatten. De vraag is hoe belangrijk het zal zijn om hergebruikt materiaal toch te vinden indien zich dit voordoet, maar het lijkt beter om op classificatie van gedefinieerde concepten te mikken.

Op dit moment is het mogelijk om achteraf te kijken of afwijkingen in tekst of beeld in de sociale media al duidden op een opkomende onrust. De potentie van automatische anomaliedetectie is dat uit grote hoeveelheden data, afwijkingen vroegtijdig onder de aandacht worden gebracht bij de analist, die deze signalen kan interpreteren voor een verbeterde inzet bij de ophanden zijnde situatie.

› INFORMATIE DIE WERKT!

4.2 SOCIALE MEDIA

Contactpersoon: Allard Kernkamp

Sociale media zijn een hype. Toch lijken ze niet meer weg te denken, aangezien zeer veel mensen ze in korte tijd zijn gaan gebruiken. Sociale media als Hyves, Twitter, Facebook en YouTube zijn begrippen geworden die niet meer weg te denken zijn. Dit wil niet zeggen dat de toepassingen van deze sociale media zullen veranderen. Twitter bijvoorbeeld was een begrip waar niemand 4 jaar geleden nog van had gehoord. Onderzoek van Comscore wijst aan dat Twitter inmiddels het hoogste bereik ter wereld realiseert in Nederland. De chatdienst bereikt 22,3 procent van de bevolking van 15 jaar en ouder. Dit zal in een razend tempo doorgroeien. Het is hoogstwaarschijnlijk dat er binnen 2 jaar een nieuw sociaal medium zal zijn ontstaan dat zeer interessant is voor het grote publiek. Dit doet echter niets af aan de vraag hoe je verschillende vormen van sociale media kunt gebruiken in het veiligheidsdomein. Sociale media bieden namelijk in potentie interessante informatie waar organisaties in het veiligheidsdomein voor handhaving, opsporing en crisisbeheersing wat mee kunnen. Een aantal 'early adaptors' binnen de politie, de brandweer en de GHOR zien de mogelijkheden ook , waardoor initiatieven met sociale media als paddenstoelen uit de grond schieten.

SOCIALE MEDIA
Sociale media zijn een verzamelbegrip voor online platformen waar de gebruikers, met geen of weinig tussenkomst van een professionele redactie, de inhoud verzorgen. Tevens is er sprake van interactie en dialoog tussen de gebruikers onderling.

Waar dan te beginnen? En welke onderwerpen zijn vooral interessant voor de politie, brandweer of GHOR in handhaving, opsporing of crisis-beheersing?
Binnen dit project is voor een tweetal aspecten gekozen om nader uit te werken. Dit zijn het fenomeen Crowdsourcing en het bepalen van de betrouwbaarheid van gegevens van sociale media. Uit diverse interviews en uit evaluaties (bijvoorbeeld onderzoek naar de crash van de Turkish Airlines of de brand bij Moerdijk) is namelijk gebleken dat hulpverleners niet gewend zijn op een juiste wijze om te gaan met informatie afkomstig van sociale media die relevant kan zijn vooraf, tijdens of na afloop van een incident. Hoe kan dan toch een gemene deler uit de stroom van informatie worden gehaald? En hoe moet het grote publiek dan bereikt of bevraagd worden (crowdsourcing)? En hoe kan de betrouwbaarheid van de gegevens die in de sociale media de ronde gaan worden beoordeeld.

CROWDSOURCING
User generated content, ofwel informatie geproduceerd door burgers en andere belanghebbenden, verschaft veel actuele informatie. Maar naast een bron van intelligentie, kan deze bron ook aangesproken worden met vragen. Wanneer dit gedaan wordt spreekt men doorgaans over crowdsourcing. Hoewel de term crowdsourcing pas in 2006 door Jeff Howe werd geïntroduceerd als samenvoeging van outsourcing met de crowd, is het fenomeen niet nieuw.

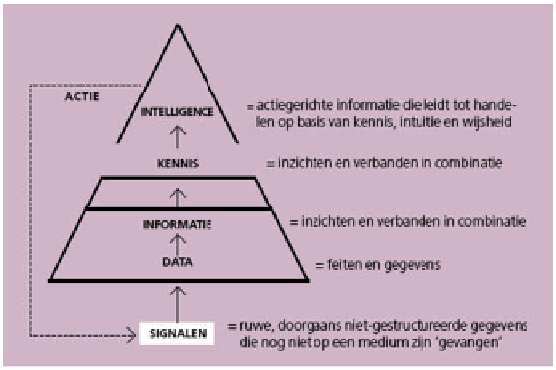
VISIE OP SOCIALE MEDIA

Aan het begin van het onderzoek en uit diverse interviews met veiligheidsregio's, korpsen, de politieacademie, enzovoorts, bleek dat er geen gemeenschappelijke visie bestond hoe sociale media te gebruiken. Veel professionals die wat met het onderwerp te maken hebben, hadden zogezegd een eigen stokpaardje of waren vanuit hobby of professionele interesse bezig met het onderwerp.

Hierdoor ontstond binnen TNO de ambitie om een duidelijk beeld te schetsen. De bedoeling van deze TNO visie op gebruik van sociale media voor het veiligheidsdomein is het voeren van een discussie met de betrokken partijen en zo te komen tot een gedeeld beeld voor de toekomst. Om de visie te verwezenlijken is uit meerdere projecten binnen TNO een bijdrage geleverd. Deze visie wordt eind 2011 uitgebracht en is primair bedoeld voor beleidsmakers en –ontwikkelaars in het veiligheidsdomein. Daarnaast biedt het hopelijk voor alle andere professionals in het veiligheidsdomein een bron van inspiratie.

CROWDSOURCING VOOR OPSPORING

Om het onderwerp Crowdsourcing verder te verdiepen, is literatuuronderzoek gedaan en zijn interviews gehouden bij de veiligheidsregio's. Dit heeft geresulteerd in een artikel waarbij crowdsourcing met name is bezien in het licht van Opsporing. Reden voor een focus op dit domein is dat Handhaving en Crisisbeheersing reeds binnen andere TNO onderzoeken worden opgepakt. Een focus op Opsporing heeft daardoor de meeste toegevoegde waarde.
Het artikel gaat in op verleden, heden en toekomstig gebruik van crowdsourcing en spitst dat toe op de opsporingspraktijk. Hoe de politie burgers effectief kan inzetten om oplossingen en ideeën voor problemen te verkrijgen om zo de kwaliteit van beleid, uitvoering of controle te verhogen. In het artikel wordt ook de relatie met het Nationaal Intelligence Model gelegd; eind 2009 heeft de Raad van Hoofdcommissarissen namelijk bepaald dat alle korpsen eind 2012 'informatiegestuurd' moeten werken[1]. En crowdsourcing kan hier een deel van de oplossing bieden. Wel moet opgemerkt worden dat de bak met data die interessant kan zijn voor opsporing, nog geen intelligence betekent. Zie hiervoor figuur 12. Naar verwachting wordt het artikel eind 2011 aangeboden ter publicatie.



Figuur 12 Van data naar kennis en intelligence

[1] http://www.rijksoverheid.nl/documenten-en-publicaties/rapporten/2009/05/04/informatiegestuurde-politie.html

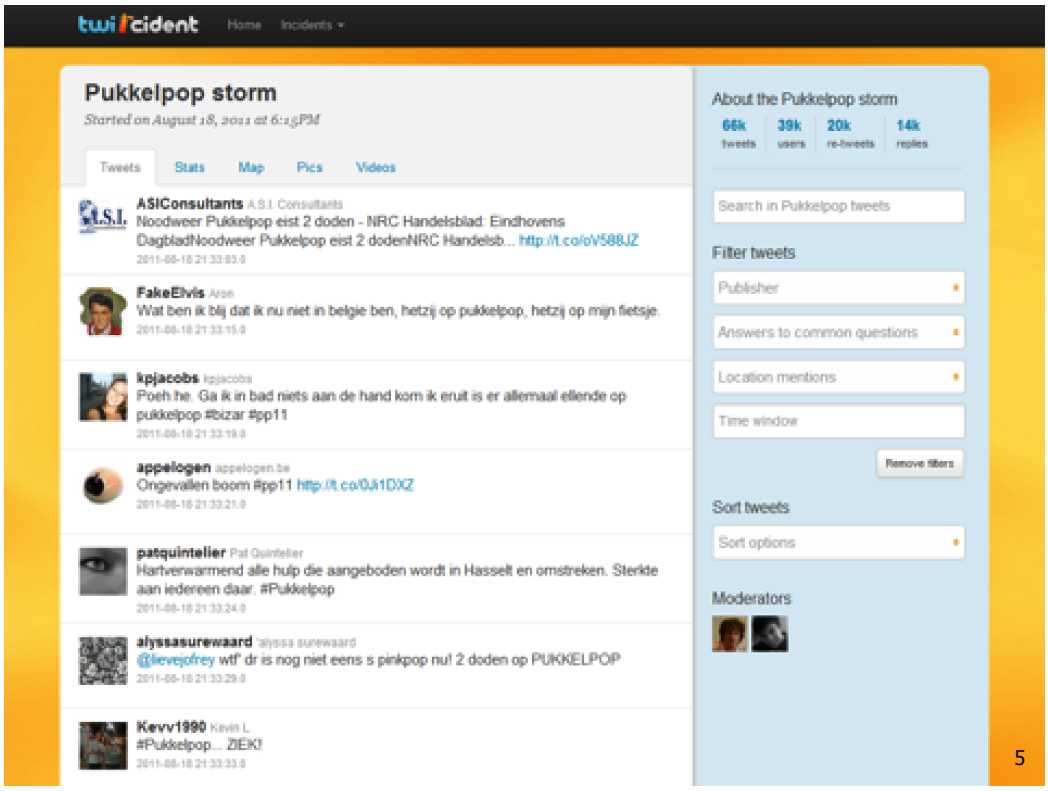
BETROUWBAARHEID VAN GEGEVENS

Er is geen eenduidig antwoord te geven over hoe de betrouwbaarheid van informatie op sociale media bepaald kan worden. De noodzaak tot een bepaalde maat van betrouwbaarheid van bijvoorbeeld een zwak signaal uit sociale media dat kan duiden op een incident is bijvoorbeeld minder dan de maat van betrouwbaarheid van gegevens die benodigd zijn voor opsporing.

Om de betrouwbaarheid van sociale media te bepalen, is vanuit dit project een samenwerking gestart met topic 2 (focus op burgerparticipatie), waarbij een experiment wordt uitgevoerd om te bepalen wat de toegevoegde waarde is van de informatie van burgers uit sociale media bij incidentmeldingen. En wat er dan vervolgens mee gedaan kan worden, gerelateerd aan de betrouwbaarheid van deze informatie. Dit experiment wordt eind 2011 afgerond en is gebaseerd op een tweetal recente casussen (incidenten) waarbij de twitterdata achteraf geanalyseerd en gewaardeerd is op basis van de informatieve elementen uit de berichten en de betrouwbaarheid van de zenders. Dit is vergeleken met de ervaringen van de professionals uit de veiligheidsregio's die te maken hadden met de afhandeling van deze incidenten, toen deze zich afspeelden.

Daarnaast is gedurende het jaar informatie naar voren gekomen uit een afstudeeronderzoek (*Exploiting Twitter to fulfill information needs duringIncidents - Richard Stronkman*), waarin op een tweetal manieren is geprobeerd om uit Twitterdata betrouwbaarheid van gegevens af te leiden. Hieruit is gebleken dat met name reputatie of autoriteit van de zender op een bepaald onderwerp, een grote rol speelt en (in het geval van Twitter) tot betrouwbare resultaten kan leiden. Denk bijvoorbeeld aan het aantal volgers dat iemand heeft op Twitter, of het feit dat een persoon in het veiligheidsdomein zelf werkt.

Echter niet elk sociaal medium werkt zoals Twitter met zenders en volgers. Daarmee is voor een deel antwoord gegeven op de betrouwbaarheidsproblematiek. In de toekomst zal enerzijds aandacht moeten worden besteed aan hoe deze betrouwbaarheidscriteria uit te breiden naar andere sociale media en anderzijds hoe men op een snelle en efficiënte wijze de juiste berichten boven water haalt als elke seconde telt.



Figuur 13 Twitcident

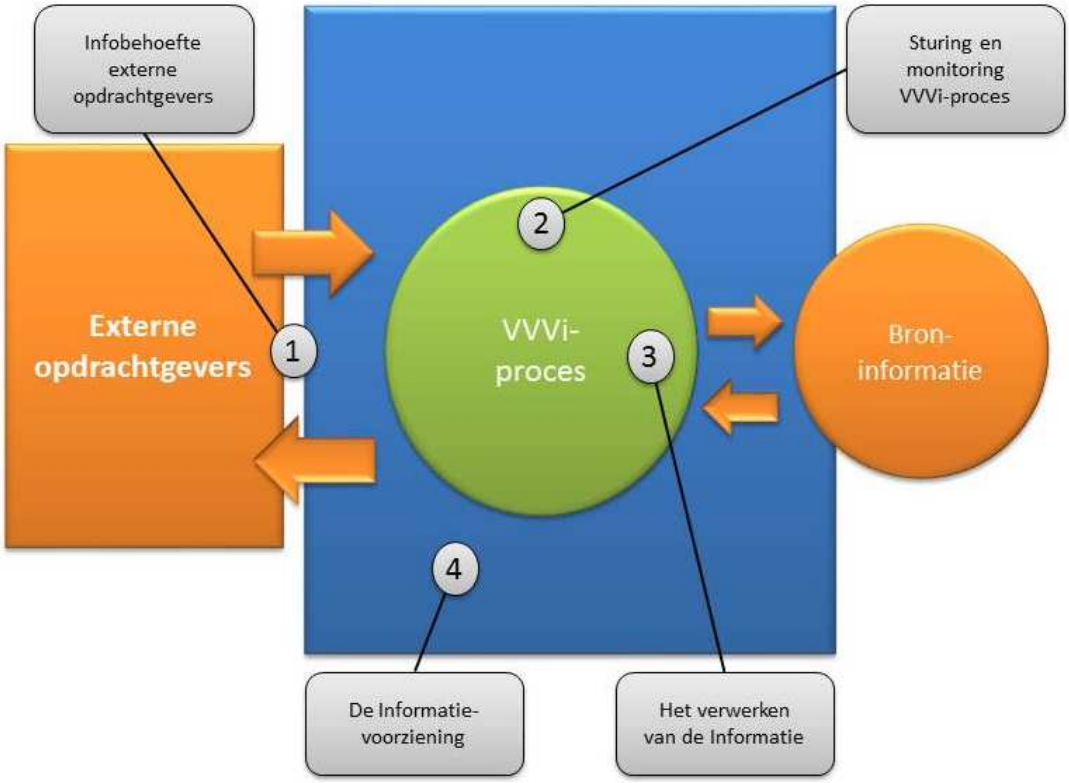
TNO heeft het pionierswerk van Richard Stronkman gewaardeerd en hem betrokken bij de verdere ontwikkeling van een sociale media monitoringstool binnen een aanpalend project (Burgermeldingssignalering) op gebied van sociale media en handhaving. Zie bovenstaande figuur voor een voorbeeld uit een Beta-versie van de tool Twitcident (ontwikkeld door Richard Stronkman, TUDelft en gefinancierd door TNO) van hoe in relevante tweets gezocht en gemakkelijk gefilterd kan worden.

CONCLUSIES

Sociale media zijn niet meer weg te denken uit de huidige maatschappij. Voor het veiligheidsdomein betekent het dat niets doen, geen optie meer is. TNO heeft daarom haar visie verwoord hoe sociale media in het veiligheidsdomein gebruikt zouden kunnen worden, door een blik in de toekomst van social media toepassingen in de Openbare Orde en Veiligheid te werpen. Daarvoor zijn er diverse interviews met eindgebruikers geweest, is een expertworkshop gehouden, een enquête uitgezet en is een SWOT analyse gemaakt. Bovendien zijn benodigde technologische en sociale innovaties in ogenschouw genomen, en zijn dilemma's die gepaard gaan met de grootschalige veranderingen door social media benoemd. Ook is een eerste handreiking gedaan om social media in (het beleid van) een organisatie in te bedden.

Dit project heeft via deze werkzaamheden bijgedragen aan een drietal resultaten:

1. Een bijdrage aan een TNO visie(boek) op toepassing van sociale media voor het veiligheidsdomein, primair bedoeld voor beleidsmakers in handhaving, opsporing of crisisbeheersing;
2. Een artikel over crowdsourcing gericht op opsporing;
3. Een bijdrage aan het experiment dat primair door topic 2 (Burgerparticipatie) is uitgevoerd in relatie tot het bepalen van de betrouwbaarheid van informatie uit sociale media.



Figuur 14 Belangrijke elementen in het informatieproces

4.3 VOORKOMEN VAN INFORMATIE-OVERLOAD DOOR BETERE STURING OP HET PROCES

Contactpersoon: Kees den Hollander

Eén van de denkbeelden die speelt bij het begrip informatie-overload is dat er veel te veel informatie beschikbaar is, en dat er oplossingen gezocht moeten worden in de richting van het snel (automatisch) vinden, ophalen en combineren van de juiste informatie om die vervolgens te (laten) gebruiken. Dit is echter maar een deel van het probleem en de oplossingsrichtingen.

Binnen dit project is vooral gekeken naar het informatievoorzieningsproces en is de focus gelegd op het proces van informatiestromen en verwerking en de beschikbaarheid en juistheid van de (start)informatie. De factoren mens en techniek komen in algemene zin aan bod. De wijze waarop de mens met grote hoeveelheden informatie en overload omgaat is in dit project niet meegenomen (dit staat voor 2012 op het programma) en de techniek die kan ondersteunen bij het snel en slim doorzoeken van informatie wordt in een ander project ontwikkeld (zie het project Informatie-analysetechnieken).

AANPAK

Dit project is uitgevoerd in nauwe samenwerking met KLPD/IPOL. Zij zijn bereid geweest om als onderzoeksobject te fungeren. Op basis van een vooraf opgestelde vragenlijst zijn vertegenwoordigers van units/teams van IPOL en de Dienst Projecten en Informatievoorziening (Informatiemanagers KLPD) geïnterviewd over o.a. ingaande en uitgaande informatiestromen, informatiebehoeften van anderen/op te leveren producten aan anderen en eigen informatiebehoeften/te gebruiken producten van anderen.

AANBEVELINGEN

De resultaten van het onderzoek zijn verwoord in een rapportage aan IPOL[2]. De generieke problemen en oplossingsrichtingen die op basis van het onderzoek bij IPOL naar voren zijn gekomen en die ook in andere organisaties worden gesignaleerd, worden hier beschreven.

De aanbevelingen richten zich op vier categorieën (zie figuur 14):

1. Goed inzicht in de informatiebehoeften van (externe) opdrachtgevers (klantvraag)
2. Heldere inrichting van sturing en monitoring van het inhoudelijke (VVVi-)proces (vanaf informatiebehoefte tot product/advies).
3. Efficiënte (snelheid/ tijdigheid) en effectieve (juistheid/voldoen aan klantvraag) verwerking van informatie
4. Gestroomlijnde informatievoorziening.

[2] Dit rapport is niet openbaar; meer informatie via KLPD/IPOL unit onderzoek en ontwikkeling te verkrijgen

Rondom deze vier aanbevelingen kan in generieke zin een aantal aandachtpunten worden aangedragen. Deze aandachtspunten kunnen voor organisaties die met grote hoeveelheden informatie moeten werken interessante aanknopingspunten bieden ter evaluatie en verbetering van het eigen proces.

1. GOED INZICHT IN DE INFORMATIEBEHOEFTEN VAN (EXTERNE) OPDRACHTGEVERS (KLANTVRAAG)

Het verwerken van informatie gebeurt om een bepaalde reden. Er is iemand binnen of buiten de organisatie die de resultaten van de informatieverwerking (analyse/ veredeling/ ordening) wil toepassen om zijn of haar werk uit te voeren.

Dit betekent dat een eerste vereiste is dat de informatiebehoeften van opdrachtgevers helder moeten zijn voor degenen die de informatieverwerking doen. Een goed intakeproces is hiervoor cruciaal.

Een belangrijk element in het intakeproces is prioriteitstelling. In de meeste situaties zijn er meerdere opdrachtgevers tegelijk die ieder hun eigen prioriteiten hebben en veelal onderling niet afstemmen. Het gesprek aangaan over prioriteiten met opdrachtgevers en tussen opdrachtgevers onderling geeft duidelijkheid bij alle partijen. Voor de informatieverwerkende partij zorgt het voor structuur in de volgorde van uit te voeren werkzaamheden en gaat er minder tijd verloren met het ad hoc communiceren met opdrachtgevers over wanneer hun resultaat wordt geleverd. Bij de prioriteitstelling tijdens het intakeproces is het verder aan te bevelen om een classificatie van opdrachtgevers (urgentie, legitimiteit van de informatiebehoefte, kans op vervolgvragen, etc.) op te bouwen en om te werken met duidelijke criteria (inclusief opschorten en stoppen van ‘lopend’ werk).

Als het gaat om ongevraagd advies (vraag-anticiperend werken) dan is een goed inzicht in de eigen informatiepositie van groot belang. Als de eigen informatiepositie goed op orde is (informatie is betrouwbaar, herleidbaar, etc.) kan van daaruit proactief advies worden gegeven aan diverse partijen voor wie de informatie van belang kan zijn. Het is hierbij van belang om vooraf de te verwachten meerwaarde van het beoogde informatieproduct vast te stellen. Ander cruciaal element hierbij is dat de ‘terugkoppellus’ op orde is. Hiermee bedoelen we dat het belangrijk is voor de informatieverwerker om te weten welke toegevoegde waarde de informatie heeft gehad voor degenen die het hebben ontvangen. Door deze terugkoppeling goed te organiseren weet de informatieverwerker steeds beter welke adviezen hij aan bepaalde partijen dient te verstrekken (leercyclus).

Binnen de informatieverwerkende organisatie dient er inzicht te zijn in (de belasting van) de eigen capaciteit. Op basis hiervan kunnen afspraken worden gemaakt met opdrachtgevers over levertijd en de hoeveelheid aan te nemen werk (prioriteiten).
Voor landelijke diensten speelt verder dat het van belang is om bij de prioriteitsstelling in overleg met opdrachtgevers bewust prioriteit te geven aan de opdrachten die een meerwaarde hebben t.o.v. wat de regionale informatieverwerkers kunnen opleveren. Ook hierbij spelen ervaringen uit de eerder genoemde ’terugkoppellus’ een belangrijke rol.

Aanbevelingen:

- Maak duidelijke afspraken met opdrachtgevers over prioriteiten op basis van duidelijke criteria en een classificatie van opdrachtgevers
- Benadruk onderlinge afstemming tussen opdrachtgevers
- Voer een goede intake uit, gericht op een heldere informatiebehoefte
- Maak vooraf inschatting over verwachte meerwaarde van informatieproducten
- Organiseer de 'terugkoppellus'
- Zorg voor inzicht in de (beschikbare) capaciteit

2. HELDERE INRICHTING VAN STURING EN MONITORING VAN HET INHOUDELIJKE (VVVI-) PROCES (VAN INTAKE INFORMATIEBEHOEFTE TOT PRODUCT/ADVIES).

Om in staat te zijn effectief te werken naar tevredenheid van klanten is het van belang om inzicht te hebben in de in behandeling van de klantvragen en te weten wat de status is van de behandeling van de diverse klantvragen. Indien er nieuwe vragen binnenkomen kan op basis van dit inzicht snel worden ingeschat of een dergelijke vraag al in behandeling is danwel of deze in te passen is (gelet op beschikbare capaciteit en prioriteit).

Hiervoor kan een zogenaamde *CCIRM*[3]-functionaliteit geïntroduceerd worden. Deze functie behelst het volgende:

- De IRM-functionaliteit adviseert of bepaalt prioriteiten en ziet toe op de status en de voortgang van de in bewerking zijnde en op te leveren informatieproducten
- De CCM-functionaliteit versterkt de eigen informatiepositie door het sturen (*tasken*) en uitzetten van verzoeken tot informatie of middelen (*asken*) op basis van de informatiebehoeften, dus niet alleen door het opbouwen en onderhouden van contacten met bronnen

Aan de hand van verschillende criteria voor het prioriteren, monitoren en bepalen van de kwaliteit van producten (ook voor hergebruik), kan de sturing en coördinatie worden ingevuld. Het gaat hierbij enerzijds om het meetbaar maken van de productie aan de hand van het aantal opgeleverde producten (zoals bijvoorbeeld vermeld in het jaarplan), maar ook om het inzichtelijk maken van de toegevoegde waarde van de producten voor de opdrachtgevers.

Aanbevelingen:

- Introduceer een CCIRM-functionaliteit
- Ontwikkel criteria om kwantitatief en kwalitatief zicht te krijgen op de geleverde prestaties

3. EFFICIËNTE EN EFFECTIEVE VERWERKING VAN INFORMATIE

Het primaire proces van organisaties die te maken hebben met grote hoeveelheden informatie is informatieverwerking. Voorbeelden zijn IPOL, BZK, informatie-units bij Politie

[3] CCIRM: Collection Coordination and Information Requirement Management

(intelligence/analyseprocessen), maar ook belastingdienst, Koninklijke Marechaussee en Douane (controlerende processen).

Het op een effectieve en efficiënte manier verwerken van de informatie is daarmee cruciaal voor het goed presteren van de organisatie. Een aantal onderwerpen is hierbij van belang:

Kwaliteit van broninformatie:
Informatie bij de bron moet consistent en volledig ingevoerd worden. Dit is de cruciale eerste stap teneinde in staat te zijn om verderop in het VVVi-proces zoveel mogelijk geautomatiseerd de informatie te kunnen verwerken en analyseren. Inspanningen om informatie te "repareren" en te verrijken (vanwege in kwalitatief en kwantitatief opzicht ontoereikende broninformatie) verstoren het primaire proces in hoge mate. Enerzijds moet men in staat zijn om aan bronnen de behoefte aan informatie met de bijbehorende kwaliteitscriteria duidelijk te maken. Anderzijds dient nadrukkelijker te worden gestuurd op het consistent en gedisciplineerd invoeren van informatie bij de bron.

Excelleren in het combineren van informatie en kennis:
Bij het opleveren van informatieproducten kan het van belang zijn om nieuwe trends, fenomenen, onderwerpen, etc. te signaleren voor opdrachtgevers. Analisten met diverse expertises kunnen bij elkaar gebracht worden met als taak zich te richten op het combineren van informatie van verschillende expertises en bronnen (*all source analysis*). Het over de grenzen van een specifieke expertise heen kijken, levert vaak interessante nieuwe invalshoeken op.

Van analyseorgaan naar adviesorgaan:
Om een goede toegevoegde waarde te kunnen leveren richting opdrachtgevers is inleving in de rol van de opdrachtgever en de context waarin deze rol wordt vervuld van groot belang. Daarmee kan een beter beeld verkregen worden over de juiste informatiebehoefte en de inhoud en vorm van het product dat geleverd moet worden, inclusief het leveren van de juiste adviezen.

Aanbevelingen:
- Ontwikkel kwaliteitscriteria voor en communiceer deze (maak afspraken) met leveranciers van broninformatie
- Kijk ook over de grenzen van specialismen heen om te komen tot nieuwe inzichten/ invalshoeken
- Investeer in het leren kennen van de wereld waarin je opdrachtgevers werken en stem hier de inhoud en vorm van het product / advies op af

4. GESTROOMLIJNDE INFORMATIEVOORZIENING

De interne informatiehuishouding van een informatieverwerkende organisatie moet op orde zijn. Als de ondersteuning hierin onvoldoende wordt ingevuld, zal dit direct ten koste gaan van de effectiviteit en efficiëntie van het primaire proces. De interne informatiehuishouding is het werkgebied van de informatiemanager. Deze definieert in samenwerking met de informatieverwerkers de wijze waarop in informatiebehoeften zal worden voorzien. De Informatiemanagementfunctie kent een aantal taken:

- Het in kaart brengen van de integrale workflow en bijbehorende (interne) informatiebehoeften
- Bedenken hoe hieraan (via diverse informatiekanalen) zou kunnen worden voldaan
- Monitoren of hieraan wordt voldaan

Daar waar de CCIRM-functionaliteit zich richt op de inhoud van de externe informatieproducten, heeft de informatiemanagementfunctie vooral betrekking op het proces van de informatievoorziening.
Om de informatiemanagement-functie goed uit de verf te laten komen dient deze binnen een informatieverwerkende organisatie te worden belegd, bij voorkeur als rechterhand van het lijnmanagement.

Aanbevelingen:
- Richt de informatiemanagement-functie op het optimaliseren van de interne informatiehuishouding van een organisatie
- Positioneer informatiemanagement in de organisatie als rechterhand van het lijnmanagement

Figuur 15 Aanbevelingen voor sturing van informatieprocessen

CONCLUSIES

De aanbevelingen zijn vooral gericht op het verbeteren van processen en werkwijzen om daarmee beter om te kunnen gaan met de informatiestromen in de organisatie.

In het algemeen geldt dat behalve voor het implementeren van deze verbeteringen ook moet worden geïnvesteerd in de factoren mens en techniek.

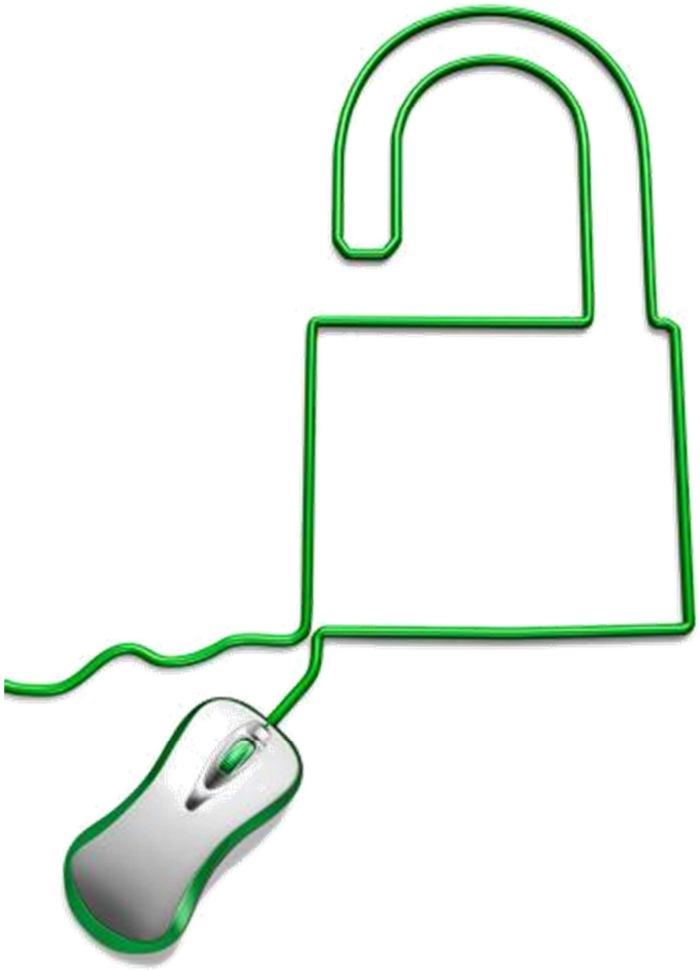
Op het gebied van de mens (personeel) dient vooral te worden geïnvesteerd in het vergroten van het informatiebewustzijn van de informatieverwerkende medewerker. Informatiebewustzijn heeft hier betrekking op kennis over de plaats en verwachte bijdrage in de totale informatieverwerkende keten. Verder dient te worden geïnvesteerd in de informatieverwerkende vaardigheden van het personeel, zowel bij werving, selectie en aanname van nieuw personeel als bij het opleiden en onderhouden van de kennis van personeel.

Op het gebied van techniek dienen de verbeterde processen en werkwijzen, zoals aanbevolen in de voorgaande paragrafen, op adequate wijze te worden ondersteund door tooling. Denk hierbij aan tooling voor de CCIRM-functionaliteit, tooling voor het analyseproces (datamining, ontsluiten van meerdere bronnen), tooling voor de ondersteuning van de workflow en tooling ter ondersteuning van samenwerking.

Om deze aanbevelingen daadwerkelijk te implementeren in een organisatie is het van belang om de oplossingsrichtingen verder uit te werken voor de specifieke omstandigheden in die bepaalde organisatie in overeenstemming met de relevante omgeving. Het beproeven van aanbevelingen in een experimentele/ gecontroleerde omgeving (bijvoorbeeld in de vorm van een pilot) is een beproefde werkwijze. Bij succes kan daadwerkelijk tot implementatie worden overgegaan.

Tezamen vormen deze activiteiten een sterke basis voor het oplossen en voorkomen van information-overload, omdat bewust gestuurd wordt op het effectief en efficiënt genereren van informatieproducten op basis van informatie die daarvoor benodigd en relevant is.

› INFORMATIE DIE WERKT!



4.4 PRIVACY BIJ HET SLIMMER OMGAAN MET GROTE HOEVEELHEDEN INFORMATIE

Contactpersoon: Jaap-Henk Hoepman en Gabriela Bodea

In het onderzoeksvoorstel (2011 – 2014) van topic 3 wordt aandacht voor privacy als een belangrijk onderdeel benoemd. Er is voor gekozen om de focus te leggen op de technologische kant van hoe om te gaan met privacy problematiek bij het werken met grote hoeveelheden informatie. De juridische, ethische en organisatorische kant vallen buiten het kader van het topic. Binnen TNO wordt in andere projecten hier wel onderzoek naar gedaan[4].
De centrale vraag rondom privacy vanuit topic 3 is:
Hoe kunnen privacy en vertrouwelijkheid de vereiste aandacht krijgen bij de nieuwe informatieanalysemethoden voor ondersteuning van veiligheidstaken?

Deelvragen hierbij zijn:

- Wat zijn de verwachte ontwikkelingen op gebied van privacy en welke impact heeft dit voor het gebruik van sociale media, burgerparticipatie etc… in operationele veiligheidstaken?
- Welke specifieke privacy-enhancing technologies zijn er en komen eraan?
- Hoe kun je tegelijkertijd de (maatschappelijke) veiligheid verhogen en de privacy beter beschermen door systemen zo in te richten dat je alleen informatie over overtreders te zien krijgt?
- Vertrouwelijkheid van informatie vormt vaak een blokkade om informatie te delen. Is dit zo of zijn er ook andere blokkades die dit mede veroorzaken? Wat kun je daar aan doen? Hoe en in welke mate kan technologie een rol spelen bij vraagstukken rondom vertrouwelijkheid?

In de opzet voor 2011 was geen specifiek project opgenomen op gebied van privacyvraagstukken bij het slimmer omgaan met grote hoeveelheden informatie. Het idee was om het onderwerp als onderdeel van alle deelprojecten mee te nemen. In de loop van het jaar is toch de behoefte ontstaan (mede vanuit de deelprojecten) om meer zicht te krijgen op de vraagstukken in het veiligheidsveld (op de werkvloer en bij beleidsmakers) op gebied van privacy bij handhaving en opsporing. Hiertoe wordt een aantal interviews gehouden. Het resultaat hiervan is bij het uitbrengen van dit boekwerk nog niet beschikbaar. Desondanks willen we toch even stilstaan bij enkele basisprincipes op gebied van privacy die voor de werkzaamheden in topic 3 van belang zijn.

[4] O.a. binnen het Europese project Virtuoso (FP7)

UITDAGINGEN EN DILEMMA'S

Veiligheid en privacy (beide fundamentele rechten van mensen) worden vaak als tegenovergestelde polen neergezet. De uitdaging is echter om beide gelijktijdig op een hoger plan te brengen. Een verantwoord gebruik van technologie kan ook leiden tot een verbetering van onze privacy, juist doordat deze technologie het mogelijk maakt contexten te scheiden en privacybescherming te zien als meer dan alleen het beschermen van de confidentialiteit. Hiervoor is innovatie noodzakelijk. Een voorbeeld van een innovatieve privacybeschermende techniek is *revocable privacy* (zie kader).

REVOCABLE PRIVACY

Veiligheid en privacy worden ten onrechte als elkaars tegenpolen gezien. Hierdoor gaan oplossingen voor (maatschappelijke) veiligheid ten onrechte ten koste van de privacy. En wordt bij de ontwikkeling van privacybeschermende technologieën te weinig rekening gehouden met redelijke veiligheidswensen. Het *revocable privacy* concept probeert deze kloof te overbruggen (Hoepman 2008). De crux is om zowel de security als de privacy leidend te laten zijn in de architectuur van een systeem.

Een systeem implementeert *revocable privacy* als de architectuur van het systeem garandeert dat gegevens over individuen slechts dan beschikbaar komen als aan een aantal vooraf gedefinieerde voorwaarden is voldaan. Zolang de gebruikers van het systeem zich aan de regels houden, dan is hun privacy gegarandeerd. Als je het revocable privacy principe toepast op bijvoorbeeld rekeningrijden, dan is gegarandeerd dat de overheid niet weet waar je gereden hebt, tenzij jij je rekening niet betaald hebt.

Diverse maatschappelijke ontwikkelingen[5] hebben er voor gezorgd dat informatie alomtegenwoordig is, praktisch onvergankelijk is, en eenvoudig te doorzoeken en te combineren is. Dit heeft gevolgen voor de zogenaamde informationele privacy. Daarmee wordt bedoeld het recht om te bepalen welke informatie over onszelf door anderen verzameld wordt en hoe deze informatie vervolgens gebruikt wordt. Het aspect van privacy dat gaat over het beschermen en uitoefenen van controle op de informatie over jezelf wordt ook wel dataprotectie genoemd. Met name deze informationele privacy is van belang bij het onderwerp waar we ons in dit topic mee bezig houden.

Enkele uitdagingen hierbij zijn:

- Informatie, ook persoonlijk informatie, is eenvoudig te bewaren, te kopiëren en te verspreiden. Iedereen doet daar aan mee door gebruik te maken van sociale netwerken zoals Facebook, Hyves en Twitter. Niet iedereen is zich echter bewust van de privacyrisico's. Bedrijven als Google en Facebook rekken de grenzen van de privacy bewust op omdat zij verdienen aan het delen van zoveel mogelijk informatie.
- Het toezicht houden door inzet van slimme sensorsystemen en camera's wordt een onderdeel van het dagelijkse leven. Op steeds meer plekken wordt ervan gebruik gemaakt om misstanden te kunnen signaleren: denk aan parkeergarages, scholen,

[5] Zie voor meer achtergrondinformatie http://wiki.science.ru.nl/privacy

winkelcentra, kantoorgebouwen etc.... Hoe acceptabel is dat en hoe bewust zijn we ons hiervan? Wat mogen we wel/niet met de op deze manier verzamelde informatie?

- Risico op onjuiste conclusies/interpretaties op basis van onvolledige of onjuiste data; Door het steeds meer koppelen van systemen en databestanden wordt het zicht op de bron en betrouwbaarheid van de data steeds lastiger en de kans op onjuiste conclusies groter.

Bij de mogelijke technieken die gebruikt kunnen worden om privacy te beschermen kunnen ruwweg twee benaderingen worden gevolgd:
A priori bescherming, gericht op het voorkomen van een link tussen persoon en gegevens. Immers, zonder persoonlijke gegevens is er ook geen sprake van een inbreuk op de informationele privacy.
A posteriori bescherming, gericht op hoe eenmaal verzamelde persoonsgegevens daadwerkelijk gebruikt worden.

MANIEREN OM TE KOMEN TOT SYSTEMEN MET RESPECT VOOR PRIVACY

Er zijn verschillende benaderingen en methoden ontwikkeld waarmee systemen ontwikkeld kunnen worden waarbij het respect voor privacy vanaf het begin wordt gewaarborgd.

1. Principes voor het ontwerpen van systemen
2. Toepassing van *Privacy Enhancing Technologies* (PET's)
3. Privacy by design

Hieronder worden deze drie kort toegelicht.

1. ONTWERPPRINCIPES

Er zijn een aantal ontwerpprincipes die kunnen helpen bij het ontwikkelen van een privacy-vriendelijk systeem:

- *Select before you collect*: Volgens dit principe mag je niet eerst alle mogelijke informatie verzamelen om vervolgens te kijken welke informatie je echt nodig hebt. In plaats daarvan moet je meteen beoordelen of je een bepaald data item nodig hebt, en mag je het alleen verzamelen en verder verwerken als dit inderdaad zo is (Jacobs, 2005).
- *Context scheiden*. Dit principe vereist dat gegevens over een persoon uit de ene context niet gebruikt worden in een andere context, of gecombineerd worden met gegevens over deze persoon uit een andere context. Informatiestromen uit verschillende contexten moeten dus gescheiden gehouden worden.
- *Anonimiseren.* Dit principe gaat nog een stap verder en vereist dat alle gegevens die er voor zorgen dat de informatie tot een persoon herleidbaar is wordt verwijderd.

2. HET GEBRUIK VAN *PRIVACY ENHANCING TECHNOLOGIES* (PET'S)

Er is de afgelopen decennia veel onderzoek verricht naar PET's. Dit onderzoek heeft bouwstenen opgeleverd die veelal gebaseerd zijn op cryptografie (met name het versleutelen en weer ontsleutelen van gevoelige informatie).Voorbeelden zijn het gebruik van pseudoniemen (werkt goed bij scheiden van contexten) en *Mixing*. Bij *Mixing* worden berichten niet direct van zender naar ontvanger gestuurd, maar via een aantal speciale routers (mixers) gestuurd, die een aantal van zulke berichten een korte tijd vasthouden en daarna in willekeurige volgorde (en met andere sleutels versleuteld) doorsturen. Zo is het verband tussen inkomende en uitgaande berichten verstoord, en is het veel moeilijker geworden om een bericht over het Internet op zijn pad van afzender naar ontvanger te volgen.

3. PRIVACY BY DESIGN

Privacy By Design[6] is een aanpak die erop gericht is privacy en eisen ten aanzien van data protectie als integraal onderdeel van nieuw te ontwerpen systemen en technologieën mee te nemen. Hierbij wordt niet alleen gekeken naar technische beschermingsmaatregelen, maar ook naar organisatorische en procesmatige maatregelen. Bovendien wordt de bescherming van persoonsgegevens hier integraal, over de hele keten, beschouwd. Hierbij is het

DE 7 PRINCIPES VAN PRIVACY-BY-DESIGN TOEGEPAST OP CAMERATOEZICHT EN MONITOREN AFWIJKEND GEDRAG[7]:

- Proactive not Reactive; Preventative not Remedial – Identificeerbare stukjes data zoals gezichten en kentekens weg filteren uit de beelden, zodat anderen ze niet alsnog kunnen identificeren. Tracks van mensen pas vrijgeven als op individuele gedragingen afwijkingen zijn geconstateerd. Losse observaties kunnen pas gekoppeld worden, indien een individuele observatie daar aanleiding toe geeft.
- Privacy as the Default – Als je geen afwijkend gedrag vertoond, dan gebeurt er verder niets met jouw persoonsgegevens.
- Privacy Embedded into Design – We maken er een punt van om geen identificeerbare stukjes data, zoals gezichten, te gebruiken om gedrag over tijd en plaats te koppelen.
- Full Functionality – Positive-Sum, not Zero-Sum – Indien het gedrag van mensen aanleiding geeft, dan blijft het mogelijk daar nader naar te kijken. Dit wordt niet onterecht onmogelijk gemaakt door verkeerde design-choices.
- End-to-End Security – Lifecycle Protection – Indien er geen afwijkend gedrag is vertoond, dan worden de beelden verwijderd na verloop van de bewaartermijn.
- Visibility and Transparency – De lijst van gedragingen is te controleren door bijvoorbeeld CBP of het parlement.
- Respect for User privacy/ keep it user-centric – Dit beginsel is wat vager omschreven, en is lastig te vertalen naar situaties waarin de "gebruiker", in casu de geobserveerde, niet vanzelfsprekend zelf interactie heeft met het systeem.

[6] A. Cavoukian, 2009
[7] J. van Rest, TNO 2011

noodzakelijk om vooraf een zogenaamde *privacy impact assessment* (PIA) uit te voeren om te bepalen welke privacy risico's kunnen ontstaan door het gebruik van het te ontwikkelen systeem. Zo wordt bij het ontwerpen volgens *privacy by design* direct invulling gegeven aan de wettelijke bepalingen zoals die in Nederland in de Wet bescherming persoonsgegevens (Wbp) staan. Voorbeelden zijn het expliciet maken van het doel van de informatieverwerking, het bepalen of de verwerking proportioneel is, er voor zorgen dat de verzamelde gegevens afdoende beveiligd zijn, en nagaan of aan de mogelijkheid tot inzage en correctie is voldaan.

In 2012 zal in het project 'Select before you collect' onderzocht worden op welke wijze dit principe in het veiligheidsdomein invulling kan krijgen. Daarnaast wordt op basis van de lopende inventarisatie naar privacyvraagstukken in het veld bezien welke activiteiten we kunnen opstarten.

INFORMATIE DIE WERKT!

4.5 VIDEOANALYSETECHNIEKEN

Contactpersoon: John Schavemaker

Het automatisch kunnen herkennen van logo's, gezichten, objecten of andere beelden in videomateriaal van youtube tot in beslag genomen PC's, kan zinvol gebruikt worden in het analyseproces van bijvoorbeeld radicaliserings- en kindermisbruikvideo's. Het zoeken in grote videodatabases op basis van een visueel voorbeeld heet "Visual instance search".
Dit onderzoeksgebied (in het Engels: "*query by example*") is momenteel zeer sterk in ontwikkeling, denk bijvoorbeeld aan de nieuwe functionaliteit van Google image search diein juni 2011 is gelanceerd. Het herkennen van gezichten in consumenten applicaties zoals Picasa en Facebook schept hoge verwachtingen die in de context van CCTV en lage kwaliteit internetvideo's (user-generated content) niet in de praktijk kunnen worden waargemaakt: gezichtsherkenning is resolutie gebonden.

In een aantal projecten met het MKB heeft de TNO de afgelopen jaren o.a. gewerkt aan het zoeken naar logo's op het internet, zinvol in de context van beeldmerkbescherming. Binnen dit project is onze ambitie om beter te kunnen begrijpen en uitleggen onder welke condities "*query by example*" technieken effectief kunnen worden ingezet.

TRECVID

Om wetenschappelijk stappen verder te komen met het inzichtelijk maken onder welke condities je kunt zoeken in grote video bestanden is de internationale benchmark TRECVID in het leven geroepen:

"The main goal of the TREC Video Retrieval Evaluation (TRECVID) is to promote progress in content-based analysis of and retrieval from digital video via open, metrics-based evaluation. TRECVID is a laboratory-style evaluation that attempts to model real world situations or significant component tasks involved in such situations."

TRECVID organiseert elk jaar een benchmark op een zestal taken waar tientallen gerenommeerde internationale onderzoeksinstellingen, bedrijven en universiteiten aan meedoen. Deze taken zijn geïnspireerd op de typische dagelijkse taken van eindgebruikers in het veld van internet surveillance en recherche:

- Zoeken naar bekende dingen
- Semantisch indexeren
- Multimedia kopie detectie
- Detectie van gebeurtenissen in video's van camera's op luchthaven
- Detectie van gebeurtenissen in multimedia
- Specifieke dingen zoeken (Instance Search)

TNO neemt in de persoon van Prof. Wessel Kraaij al jaren deel in de organisatie van TRECVID; daarnaast heeft TNO de afgelopen jaren meegedaan aan de instance search taak.

INSTANCE SEARCH TAAK

De instance search taak betreft in 2011 het zoeken naar personen, objecten en locaties in 250 uur BBC TV archiefmateriaal. Onderstaande figuur laat voorbeelden zien waar mee gezocht moet worden, voor elke query zijn meerdere plaatjes beschikbaar:

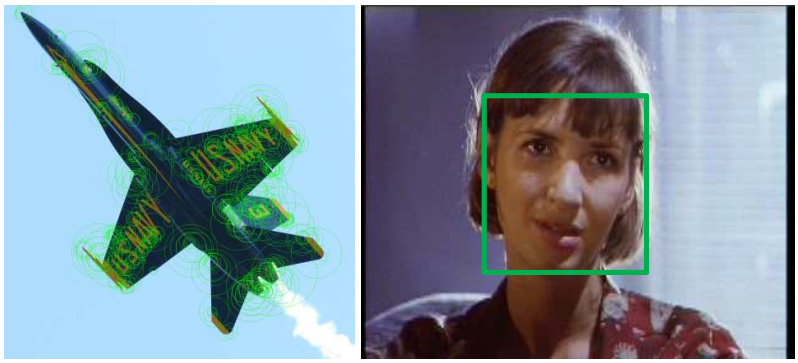
Figuur 16 voorbeeld queries van personen, locaties en objecten

TNO AANPAK INSTANCE SEARCH

TNO heeft voor drie verschillende aanpakken gekozen. Alle aanpakken worden gebouwd uit open-source componenten.

1. TNO-SURFAC: aanpak met zoeken in alle SURF punten
2. TNO-BOWCOL: aanpak met bag-of-words, globale kleur, gezichts- en huidskleurdetectie
3. TNO-SURFEIG: aanpak met gezichtsherkenning (Eigenface)

Ten eerste wordt van alle video's elke seconde een videoframe uitgekozen. Voor elk videoframe worden SURF punten, gezichten, kleur en huidskleur gedetecteerd en in XML formaat weggeschreven.



Figuur 17 (links) SURF punten in groen, (rechts) gezichtsdetectie (http://blueangels.navy.mil/)

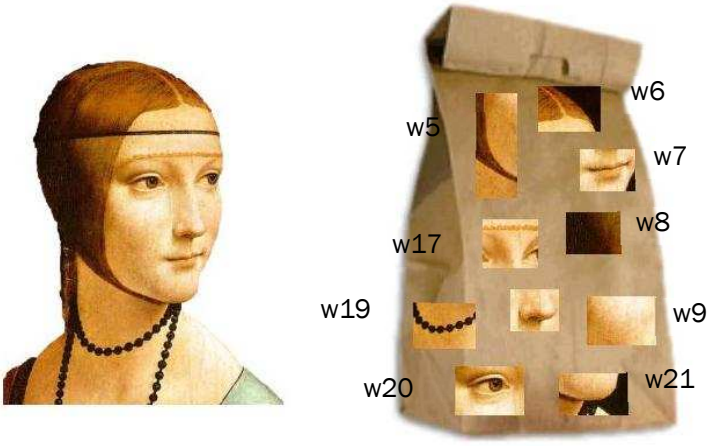
Ten tweede volgt één van de drie aanpakken:

1. TNO-SURFAC

Alle SURF punt beschrijvingen worden doorzocht met de methode van Lowe[8] (verschil tussen 1e en 2e resultaat) door eerst alle beschrijvingen naar een aantal grote binaire files om te

[8] Lowe, David G. (1999). "Object recognition from local scale-invariant features". Proceedings of the International Conference on Computer Vision. 2. pp. 1150–1157

zetten en dan de files met beschrijvingen één voor één met benaderingsmethoden te doorzoeken.



Figuur 18 bag of visual words

2. BAG-OF-WORDS

In deze aanpak worden beeldkenmerken van een videoframe vertaald naar een tekstfile met woorden. Als een gezicht in het videobeeld is gedetecteerd wordt het woord 'face' toegevoegd. Ook worden de kleuren globaal beschreven met termen als bijvoorbeeld 'HighOrange' (veel oranje) of 'LowBlack' (beetje zwart) in de tekst. Daarnaast wordt uit een woordenboek met zogenaamde visuele woorden (stukjes beeld) de woorden gekozen die in het beeld zijn gevonden: in ons woordenboek 'w5', 'w6', ..., 'w249'.

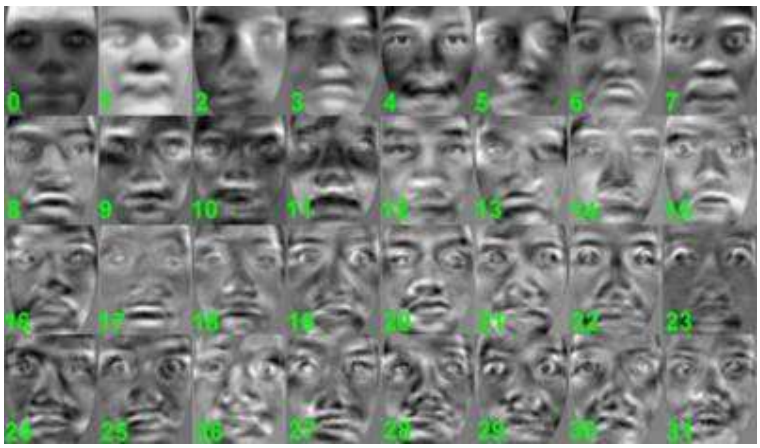
<DOC>
<DOCNO>1000_mpg_item3_sample2</DOCNO>
<TEXT>
w5 w6 w7 w8 w9 w17 w19 w20 w21 w22 w24 w25 w26 w27 w28 w29 w30 w34 w42 w51 w52 w59 w60 w62 w64 w66 w67 w68 w69 w71 w73 w76 w77 w78 w81 w84 w86 w89 w90 w96 w97 w113 w114 w117 w120 w130 w171 w177 w180 w181 w182 w189 w192 w196 w197 w203 w206 w207 w222 w224 w241 w249
face LowOrange MediumOrange HighOrange LowYellow MediumYellow LowBlack
</TEXT>
</DOC>

3. EIGENFACE

voor elk gedetecteerd gezicht in de video's wordt een beschrijving (100 getallen) uitgerekend met de Eigenface[9] projectietechniek (gevonden gezicht is gewogen som van 100 databasegezichten) en binair opgeslagen. Een query met een persoon (en dus gezicht)

[9] L. Sirovich and M. Kirby (1987). Low-dimensional procedure for the characterization of human faces. Journal of the Optical Society of America A, 4, 519–524

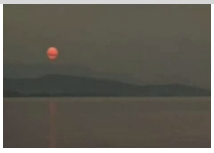
wordt gedaan door opnieuw de Eigenface (100 getallen) uit te rekenen en met approximate nearest-neighbor te vergelijken met alle gevonden gezichten.



Figuur 19 De eerste 31 Eigengezichten in de database (http://www.shervinemami.info/faceRecognition.html)

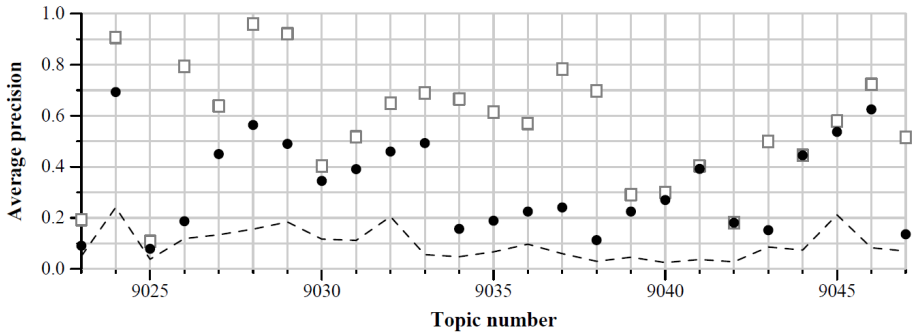
INSTANCE SEARCH RESULTATEN

In onderstaande tabel worden een drietal Instance Search zoekresultaten getoond voor een locatie, object en een persoon. Ieder zoekresultaat is met een andere aanpak uitgevoerd om de meerwaarde van elke aanpak te laten zien: SURF punten voor precies zoeken, bag-of-words voor generiek zoeken, Eigenface voor zoek naar (gezichten) van personen.

QUERY	RESULT	RESULT	RESULT
 LOCATION	 SURF punten	 SURF punten	 SURF punten
 OBJECT	 bag-of-words	 bag-of-words	 bag-of-words
 PERSON	 Eigenface	 Eigenface	 Eigenface

Figuur 20 Instantsearch zoekresultaten

De volgende figuur laat een kwantitatieve analyse zien van de zoekresultaten met SURF punten. De zwarte punten zijn de TNO resultaten, de stippellijn de gemiddelde resultaten voor alle deelnemende partijen en de vierkantjes de beste resultaten van alle partijen. De SURF punten aanpak heeft hier meer dan gemiddeld gepreseteerd.



Figuur 21 De gemiddelde (stippellijn), beste (vierkant) en TNO (zwarte stippen) score voor de SURF aanpak

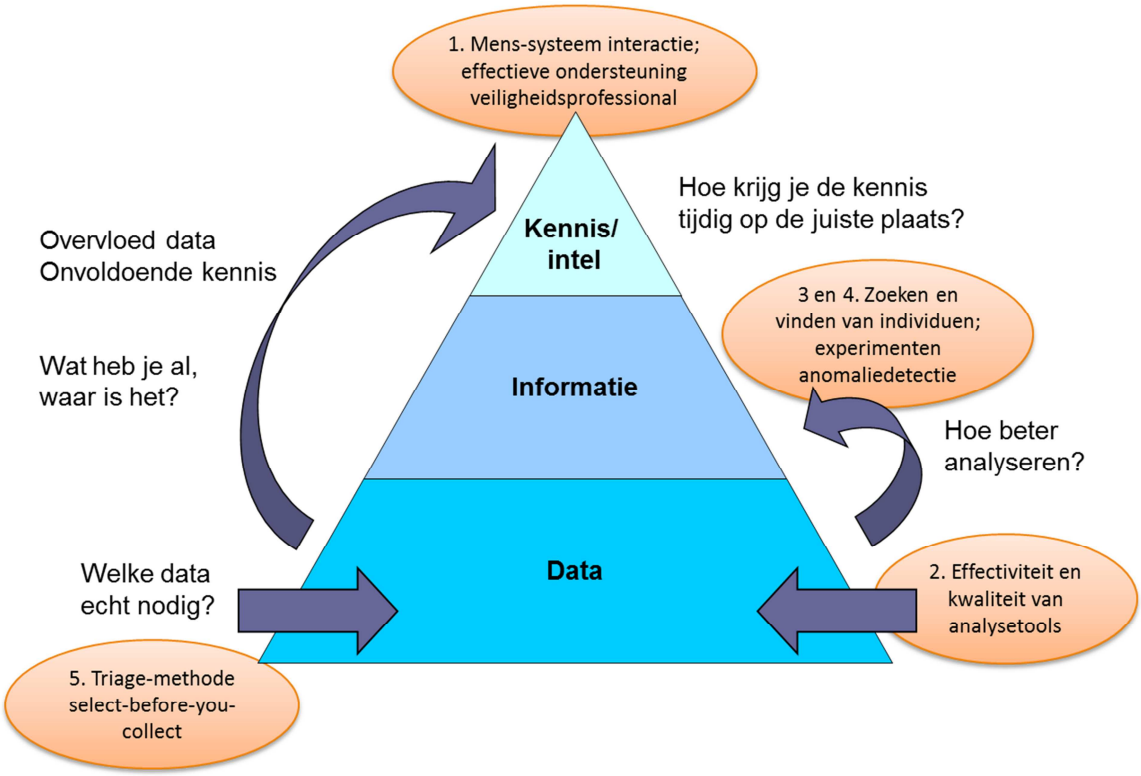
CONCLUSIES

De drie gekozen aanpakken hebben ieder potentie voor verschillende toepassingen.

De eerste aanpak (domweg alle SURF punten doorzoeken) geeft zeer precieze zoekresultaten die bijna exact gelijk zijn aan het zoekplaatje of een deel daarvan (foto of schilderij aan de muur). Voor bijvoorbeeld doorzoeken van KP materiaal in de zedenrecherche is dit zeer interessant om verbindingen tussen materiaal te leggen. Deze aanpak generaliseert echter niet: bij een zoekopdracht van een zonsondergang krijg je alleen die specifieke zonsondergang terug; geen andere die daar op lijken. Verder kosten deze zoekopdrachten relatief veel tijd: 30 tot 60 minuten voor 250 uur video.

De tweede aanpak (bag-of-words) generaliseert juist zeer sterk. Voor een zoekopdracht van een zonsondergang, bevatten de resultaten niet alleen de originele zonsondergang maar ook andere die daar op lijken. Een groot voordeel is dat deze techniek snel is en interactief gebruikt kan worden.

De derde aanpak laat zien dat met een eenvoudige (Eigenface) beschrijving van een gezicht zeer interessante resultaten voor zoeken naar personen in video worden geboekt. Voorwaarde is wel dat in het zoekbeeld het gezicht goed in beeld is (van voren en met veel resolutie). Omdat de beschrijving klein is kan er snel en interactief mee gezocht worden.



Figuur 22 De onderwerpen voor 2012 en hun plek in de informatiepiramide

5 Doelstellingen 2012

In dit hoofdstuk worden de speerpunten voor 2012 benoemd. Op basis van de bereikte resultaten in 2011 en gesprekken met de stakeholders is voor 2012 gekozen voor een vijftal onderwerpen (zie figuur 22). Ten tijde van het opstellen van deze publicatie worden de onderwerpen in overleg met de stakeholders uitgewerkt tot concrete projecten. Hierbij is het streven wederom zoveel mogelijk samen met één of meerdere stakeholders een onderwerp uit te werken. Deze manier van werken stelt ons in staat om snel zicht te krijgen op de toepasbaarheid van de ontwikkelde kennis.

FOCUS

De projecten voor 2012 betreffen de volgende onderwerpen:

1. Hoe gaan veiligheidsprofessionals die met grote hoeveelheden informatie werken te werk? Op welke wijze kan deze groep het beste worden ondersteund met technologie, training/opleiding en slimme processen (zoals gedistribueerde taken, bijv. telepolitie)? De focus ligt hierbij op mens-systeem interactie.
2. Wat is de meerwaarde (kwalitatief en kwantitatief) van de inzet van instrumenten / tools voor het verwerken van grote hoeveelheden informatie, en hoe kan dit worden gemeten?
3. Hoe kan in de grote hoeveelheid data (b.v. van social media) automatisch relevante afwijkingen (anomalieën) worden gedetecteerd? Welke kenmerken zijn hiervoor het meest waardevol? Hoe kunnen opkomende incidenten vroegtijdig worden gesignaleerd?
4. Hoe kan relevante informatie over individuen automatisch worden gecombineerd en geanalyseerd (bv. passagierslijsten, of op internet). Niet alleen netwerkanalyse op naam/nummer, maar ook op basis van kenmerken, of als afwijking. Wat is de echte identiteit (veel mensen gebruiken meerdere identiteiten) en verandert de identiteit?
5. Hoe kan select-before-you-collect worden gebruikt om te voorkomen dat teveel data binnenkomt en hoe kan de juiste prioriteit worden aangegeven, zoals bij Triage? Door het ontwikkelen van methoden en technieken die het mogelijk maken het verwachte toekomstig nut van informatie te bepalen (conform risico gebaseerde analyse of bijvoorbeeld proportionaliteit kwantificeren) kan het “select before you collect”- principe op een zo laag mogelijk niveau worden toegepast zodat grote hoeveelheden (onbruikbare) informatie ten behoeve van handhaving en opsporing worden voorkomen.

De onderwerpen sociale media en privacy zullen bij al deze onderwerpen een rol blijven spelen. Op dat vlak zullen dus ook nog activiteiten worden verricht. De exacte vorm ervan wordt begin 2012 vastgesteld als duidelijk is op welke wijze bovengenoemde vijf onderwerpen nadere uitwerking zullen krijgen.

BEOOGDE RESULTATEN

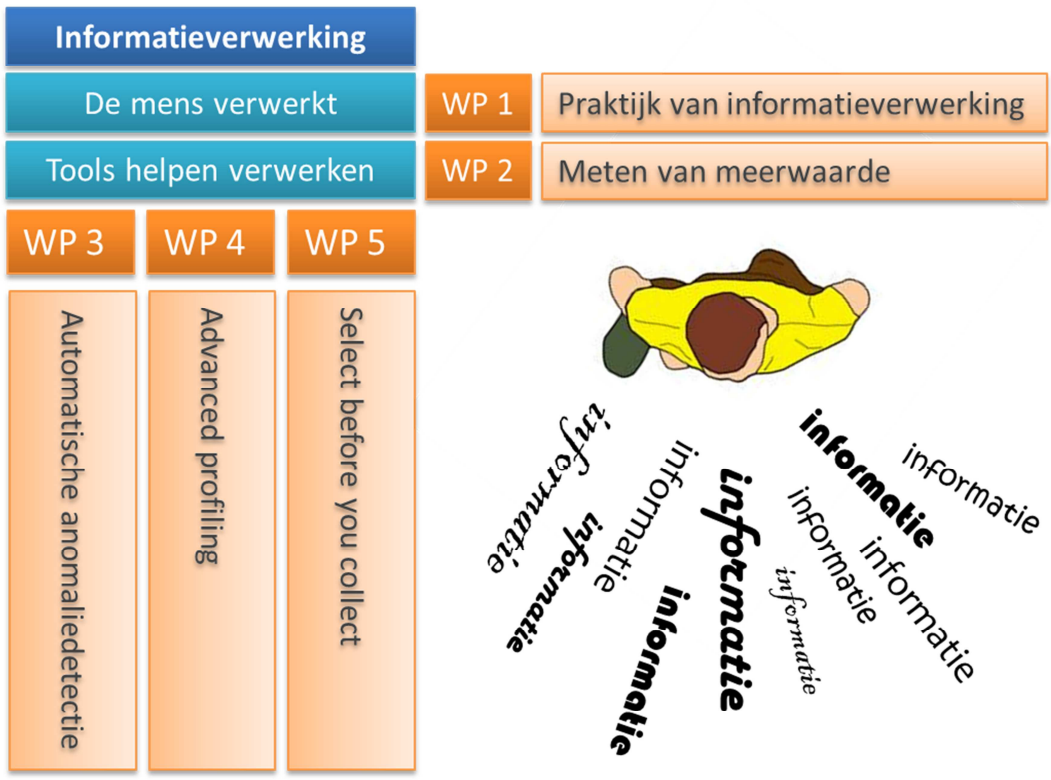
We verwachten bij het uitwerken van deze thema's tot concrete handvatten te komen voor stakeholders aan de hand waarvan ze beter in staat zijn om hun eigen informatieverwerking in goede banen te leiden en te stroomlijnen. De resultaten zijn ondersteunend voor het verder professionaliseren van het informatiegestuurd werken van de veiligheidsorganisaties. Het vormt de basis voor de processen intelligence, handhaving, opsporing en crisis-beheersing.

Concrete toepasbaarheid ligt vooral op het domein van de toezichtcentrales/meldkamers (Nationale Politie, KMar), analysewerkzaamheden (o.a. BZK, NCTV, IPOL), (internet) recherchewerkzaamheden (Nationale Politie, KMar) en ondersteuning van de informatie-functie in het front-office/back-office concept (Nationale Politie).

INTENSIVERING SAMENWERKING

In 2012 wordt de samenwerking met de huidige partijen gecontinueerd en verder uitgebouwd. Daarnaast is het de intentie om de samenwerking met de Nationale Politie te verstevigen.

Op het vlak van de internationale samenwerking zijn een aantal projecten lopend (vb. VIRTUOSO) en staan diverse projecten in de steigers (bijvoorbeeld een FP7 project rondom social media in het veiligheidsdomein en een FP7 project rondom anomaliedetectie).



Figuur 23 Samenhang tussen de werkpakketten