Organisatieonderdeel Eenheid
District
Afdeling

Contactpersoon [redacted]
Functie Teamleider OSINT
Bezoekadres Europaweg 45
2711 EM Zoetermeer
Telefoon 070-345[redacted]
E-mail [redacted]@klpd.politie.nl
Internet www.politie.nl

Datum 18 juli 2013
Status Politie vertrouwelijk/ definitief
Blad 1/29

Artikel 10, lid 2, onder e

Project monitoren 'dreigtweets'

Van methode naar techniek

Eindevaluatie juli 2013

Auteurs
[redacted]

art. 10 lid 2 onder e

Wijzigingsregister

Versie	*Datum*	*Omschrijving*	*Conceptevaluatie*
1.1	12-6	Raamwerk eindevaluatie project monitoren 'dreigtweets'	[redacted]
1.2	13-6	Raamwerk opgezet met inhoud projectplan	
1.3	18-6	Eerste aanzet eindevaluatie	
1.4	20-6	Concept eindevaluatie: gedeeld met Radboud Universiteit en Tardis Research	
1.5	26-6	Concept eindevaluatie aangepast: gedeeld met OSINT	
1.6	1-7	Oplevering definitieve versie	
1.7	2-7	Definitieve versie aangepast nav wijzigingen concept CLST	
1.8	2-7	Gereviseerde en toegevoegde tabellen resultaten DTM	
1.9	3-7	Tekstuele wijzigingen nav toevoegen tabellen	
2.0	4-7	Tabellen aangepast	
2.1	5-7	Tekstuele wijzigingen nav aangepaste tabellen	
2.2	17-7	Wijzigingen nav opmerkingen Staf Korpsleiding Nationale Politie	
2.3	30-7	Wijzigingen nav opmerkingen Staf Korpsleiding Nationale Politie	

Inhoud

	Voorwoord	**5**
1	**Inleiding**	**7**
1.1	Probleem- en doelstelling	8
1.2	Leeswijzer	8
2	**Onderzoekskader**	**9**
2.1	Onderzoekskader	9
2.2	Onderzoeksaanpak	10
2.3	Juridische kaders	10
2.4	WOB-verzoeken	11
3	**Resultaten**	**13**
3.1	Data acquisitie	14
3.2	Radboudmodule	14
3.3	LOODS-model	15
3.4	Triage	16
3.5	Annotatie	16
3.6	Evenement-relevantie	17
3.7	Algemene resultaten	18
3.8	Samenvatting	19
4	**Organisatie**	**21**
4.1	Inrichting	21
4.2	Samenwerking extern werkproces	21
4.3	Samenwerking intern werkproces	23
4.4	Fasering	23
5	**Randvoorwaarden**	**25**
5.1	Proces	25
5.2	Inbedding in de organisatie	25
5.3	Communicatie	25
5.4	Financiering	26
6	**Conclusies en aanbevelingen**	**27**
6.1	Conclusies	27
6.2	Aanbevelingen	28

Voorwoord

Bij deze ontvangt u de eindevaluatie van het project monitoren 'dreigtweets'. Het project is een gezamenlijke inspanning van de landelijke eenheid, team OSINT, de Radboud Universiteit, Centre of Language and Speech Technology en Tardis Research. De samenwerking is gebaseerd op een idee van team OSINT om op een meer gedragswetenschappelijke basis het internet te kunnen monitoren op dreigingen gericht tegen personen, objecten, diensten en evenementen in het rijksdomein (stelsel bewaken en beveiligen).

Het experiment dat hier wordt gepresenteerd is het vervolg op een eerder experiment, uitgevoerd in 2012. De eerdere fase richtte zich vooral op de ontwikkeling van een methode, waarbij in deze fase de doorontwikkeling naar techniek centraal stond. De publiek-private samenwerking binnen dit experiment is relatief nieuw voor de nationale politie. Een samenwerking die het beste van twee, of eigenlijk drie werelden integreert: de wetenschap, de techniek en de praktijk.

Hoewel gehoopt wordt op een succes, blijft het onderzoeken van de verwachtingen altijd spannend tot aan de start van het daadwerkelijke experiment, respectievelijk de testen op Koninginnedag 2013 en Nationale Herdenking 2013. Zoals bij alle experimenten is het resultaat afhankelijk van vele factoren. Aan de basis mogen we echter trots zijn op de behaalde resultaten, variërend van de samenwerking zo'n korte tijd tot de interesse die wij ook binnen andere eenheiden hebben gewekt met de resultaten.

Een neveneffect van de resultaten is dat het project tevens kan worden beschouwd als een sociale innovatie. Immers, de DreigTweetMonitor (DTM) heeft een vernieuwing gebracht in de wijze waarop het werk binnen OSINT is georganiseerd, en wel op een zodanige wijze dat zowel de arbeidsproductiviteit als de kwaliteit van de arbeid ermee gebaat zijn. Het reduceren van berichten die daadwerkelijke relevantie hebben met het oog op nationale evenementen, is opvallend.

Zoals gezegd is er de interesse ontstaan in het experiment. Na Project X Haren bestaat er veel aandacht voor de wijze waarop 'dreigingen' kunnen worden geduid. Hiermee hebben we expliciet ervaring opgedaan. Eén van de vragen die OSINT is gesteld, is het opstellen van een afhandelingsprotocol. In samenwerking met CCT en het Kennis- en Leercentrum van de landelijke eenheid zullen we met het oog op aanstaande Prinsjesdag en de Nuclear Safety Summit (NSS) in 2014 een dergelijk protocol kunnen produceren.

art. 10 lid 2 onder c, d en g
art. 11 lid 1

Met betrekking tot het laatste dan, de NSS vraagt bij uitstek om de implementatie van de DTM. De DTM bevat in potentie alle aspecten nodig voor het gestructureerd en realtime monitoren van dreigingen, waaronder een goed werkend werkproces. In combinatie met de behoefte in het land tot detectie en afhandeling van dreigingen zou de DTM in eerste instantie binnen de landelijke eenheid, voor de NSS kunnen worden uitgerold, om vervolgens te delen met alle eenheden.

1 Inleiding

In 2011 is het team Open Source INTelligence (OSINT) van de nationale politie, landelijke eenheid (voorheen KLPD) gestart met het monitoren van open bronnen op (potentiële) dreigingen tegen personen, objecten, diensten en evenementen in het rijksdomein in het kader van stelsel bewaken en beveiligen (SBB). De kern van het werk bestaat uit het herkennen van een voornamelijk tekstuele dreiging op social media, zoals Twitter. Intussen is het volume aan berichten over bijvoorbeeld een bekend persoon 'Wilders' of tijdens een evenement 'Koninginnedag' dusdanig hoog (meer dan 100.000 berichten via een open bron, in dit geval Twitter) dat het individueel en handmatig detecteren van dreigingen is uitgesloten. Binnen de nationale politie bestaat daarom de behoefte meer grip te krijgen op uitingen via sociale media, met nadruk op het kunnen filteren van relevante dreigtweets.

art. 10 lid 2 onder c en d

Voor het uitvoeren van OSINT wordt continu gezocht naar nieuwe methoden en technieken, zoals webmonitoring en anomaliedetectie, die de taakstelling van het team OSINT en de nationale politie kunnen ondersteunen. Innovatie- en ontwikkelingsmogelijkheden worden zowel binnen de politieorganisatie en de overheid zelf verkend, als ook in samenwerking met commerciële en wetenschappelijke instellingen. Vanuit dit perspectief is in 2011 het experiment monitoren 'dreigtweets' ontstaan, mede dankzij hulp van opdrachtgever (en financier) Ministerie van Veiligheid en Justitie, in het bijzonder de NCTV. Een coöperatief project van de nationale politie/ overheid, een wetenschappelijk instituut en een commerciële partij.

art. 10 lid 2 onder c e d

De nationale politie wil met minder menskracht een betere filtering op dreigingen op social media. In het kader van het experiment is aan het Centre of Language and Speech Technology (CLST), onderdeel van de Faculteit der Letteren van de Radboud Universiteit Nijmegen, verzocht een methode te ontwikkelen voor de detectie van tekstuele dreigingen op Twitter. Vanuit de wens om op professionele wijze om te gaan met het experiment en vanuit een concurrerende marktpositie is in 2012 door de NCTV besloten om IT-adviseur Tardis Research te betrekken voor het ontwikkelen van een goede data acquisitie en een gebruiksvriendelijke interface.

Het monitoren van dreigingen op de bron Twitter is een afgewogen keuze geweest:
- Als linguïstische dreigingsdetectie werkt op berichten die maximaal 140 tekens bevatten, is het waarschijnlijk ook mogelijk voor teksten met meer leestekens;
- Twitter laat real-time data acquisitie toe, in tegenstelling tot andere media;
- Twitter is een open bron, met de focus op het zo breed mogelijk verspreiden van berichtgeving en nieuws.

art. 10 lid 2 onder c, d, en g.

Als start van het project monitoren dreigtweets is in 2012 gestart met een conceptuele ontwikkelfase. Op Prinsjesdag 2012 heeft een operationele test plaatsgevonden welke is geëvalueerd en vervolgens dienst heeft gedaan als startpunt voor doorontwikkelingen. De evaluatie is eind 2012 gedeeld met de subsidieverstrekker. Op Prinsjesdag 2012 zijn gedurende de test ongeveer 400.000 twitterberichten gefilterd op dreiging. In negen diensturen werden 15.000 tweets met (een op Prinsjesdag toegespitste) dreigingswaarde 'real-time' geselecteerd. Vervolgens zijn er ruim 4.000 tweets nader gescreend op 'ernst' en 'waarschijnlijkheid', waarvan er uiteindelijk 135 nader zijn onderzocht. Op basis van de operationele test op Prinsjesdag kan geschat worden dat er tijdens evenementen dagelijks 35.000 Nederlandstalige dreigingen met betrekking op het evenement worden geuit via Twitter. Dit volume aan berichten is nog te hoog voor een passende (handmatige) verwerking binnen het politie-

art. 10 lid 2 onder c, d en g

te hoog voor een passende (handmatige) verwerking binnen het politie-

intelligenceproces (Referentiekader bedrijfsprocessen politie 2012: verzamelen en verwerken van veiligheidsinformatie). Tijdens de uitvoering van het experiment monitoren 'dreigtweets' is het concept en de werkwijze van de operationele test op Prinsjesdag verder ontwikkeld en onderzocht met het oog op een optimalisatie van zowel het bestaande concept en werkproces alsook de uiteindelijk beschikbare 'dreigtweetmonitor' (DTM).

1.1 Probleem- en doelstelling

Berichten verspreiden via internet is een maatschappelijk toenemend verschijnsel en is erg makkelijk. Negatief beladen berichten met serieuze bedoelingen en mogelijk ernstige maatschappelijke gevolgen, vormen hierin een niet te negeren aandeel. Binnen het veiligheidsdomein heeft de nationale politie een (morele) plicht om de ontwikkelingen in relatie tot open bronnen en dreigingen op het internet zo goed mogelijk te monitoren en waar nodig in te grijpen. Politieorganisatieonderdelen zoals de Dienst Koninklijke en Diplomatieke Beveiliging (DKDB) en het Stelsel Bewaken en beveiligen (SBB) hebben baat bij de mogelijkheid om internetberichtgevingen te kunnen filteren op relevante dreiging. Een veel gebruikt sociaal medium is Twitter. Voor het project monitoren 'dreigtweets' is een definitie (opgesteld door het CLST en overgenomen in het het projectplan) van een (relevante) dreigtweet als volgt geformuleerd:

art. 10 lid 2 onder c en d

> *Onder dreigtweet wordt een tweet verstaan, waarin de auteur **expliciet** zijn **intentie** uit om een **gewelddadige actie** te plegen jegens personen, objecten, diensten of evenementen.*

art. 10 lid 2 onder c, d en g

Als doelstelling heeft het project monitoren 'dreigtweets' het onderkennen en controleren van communicatie via sociale media op mogelijke Nederlandstalige bedreigingen binnen het rijks- en later decentraal domein door de nationale politie. Het ontwikkelen van een dreigtweetmonitor, of kort gezegd de DTM, is daarvoor noodzakelijk. De dreigtweetmonitor zal ingezet worden op dreigingsdetectie en gehost binnen het IRN-netwerk, zodat aansluiting bij bestaande beveiliging en forensische borging worden gegarandeerd. De dreigingsrapportages worden via de bestaande werkprocesroutes naar de ontvangers doorgestuurd.

art. 10 lid 2 onder c en d

art. 10 lid 2 onder c en d

1.2 Leeswijzer

In dit document wordt de eindevaluatie van het project monitoren 'dreigtweets' beschreven. Na hoofdstuk 1, de inleiding, zal aandacht worden besteed aan het onderzoekskader (hoofdstuk 2). De resultaten worden besproken in hoofdstuk 3. daarna wordt ingegaan op de organisatie van het project in hoofdstuk 4, gevolgd door het evalueren van de randvoorwaarden (hoofdstuk 5) en de conclusies en aanbevelingen in hoofdstuk 6.

Als aanvullende opmerking is het goed als de lezer zich realiseert dat het project monitoren 'dreigtweets' is aangemerkt als experiment. Aan de hand van de eindevaluatie kan gekeken worden of doorontwikkelingen met betrekking tot het geteste instrument en mogelijk inbedden binnen de (werkprocessen van de) nationale politie wenselijk is.

2 Onderzoekskader

2.1 Onderzoekskader

Het project monitoren 'dreigtweets' heeft bestaan uit het experimenteel testen van een aantal onderdelen:

art. 10 lid 2 onder c, d en g

1. het aanmerken van potentieel relevante dreigtweets door een linguïstisch dreigingsfilter;
2. het 'real-time' zichtbaar maken van de potentieel relevante dreigtweets door een gebruiksvriendelijke interface, waarbij ook data acquisitie en analyse- en workflowmogelijheden mogelijk zijn; en
3. het inbedden van de testdagen in het dagelijkse werkproces en de taakstelling van de nationale politie.

Deze onderdelen zijn gezamenlijk door verschillende samenwerkende partijen opgepakt. De linguïstische dreigingsfilter (1) is opgeleverd door het CLST van de Radboud Universiteit. De interface van de dreigtweetmonitor (en de daarbijbehorende data acquisitie, en analyse- en workflowmogelijkheden) (2) is opgeleverd door IT-ontwikkelaar Tardis Research. Het inbedden van de dreigtweetmonitor in het werkproces en de taakstelling (3) van de nationale politie, tijdens de voorafgesproken evenementen, is uitgevoerd door het team en de teamleider van het team OSINT Landelijke Eenheid.

Het instrument dat in de eerdere conceptuele ontwikkelfase is getest op Prinsjesdag 2012 was het uitgangspunt voor de verdere doorontwikkeling van de dreigtweetmonitor binnen dit experiment/project. In de tijdslijn kan dit als startpunt van het experiment worden gezien. Het eindpunt wordt bepaald door de te monitoren evenementen Koninginnedag 2013 en Nationale Herdenking 2013, op respectievelijk 30 april 2013 en 4 mei 2013.

Wanneer de omstandigheden het zouden toelaten (in de vorm van een spontaan te monitoren evenement/dreigingssituatie) werd tijdens het project bekeken of er mogelijk een tussentijdse 'test' met de tot dan toe ontwikkelde mogelijkheden uitgevoerd kon worden. Deze tussentijdse test vormde in dat geval de nieuwe opmaat voor de officiële testmomenten binnen het experiment, Koninginnedag 2013 en Nationale Herdenking 2013.

De evaluatie van het experiment geldt als eindresultaat. In deze eindevaluatie staat de ontwikkeling van de dreigtweetmonitor beschreven. Alle 'bijvangst' van het primaire doel (niet beoogde maar wel logische en geïmplementeerde wijzigingen of resultaten) wordt, waar nodig, ook beschreven in deze rapportage. De rapportage wordt ondersteund door twee technisch-specifieke deelrapportages, opgeleverd door het CLST van de Radboud Universiteit en door Tardis Research.

Het experiment is geslaagd wanneer na 4 mei 2013 blijkt dat de ontwikkelde dreigtweetmonitor Nederlandstalige twitterberichten afdoende filtert op onder andere dreiging in relatie tot een te monitoren evenement. De filteringsmethode vindt zijn oorsprong binnen wetenschappelijk taalonderzoek en wordt in het bestaande werkproces van de nationale politie ingebed. De dreigtweetmonitor maakt op basis van real-time data acquisitie, analyse en workflow in een gebruiksvriendelijke interface mogelijk.

art. 10 lid 2 onder c, d en g

2.2 Onderzoeksaanpak

Vanuit de conceptuele testfase is een evaluatie gedaan van de tot dan toe ontwikkelde 'dreigtweetmonitor'. Op basis van het experiment en de stand van zaken zijn vervolgens afspraken gemaakt over de verantwoordelijkheden en eindresultaten per betrokken partij (CLST, Tardis, team OSINT).

Met het oog op het experiment is het interessant vergelijkingen te trekken met andere methoden en technieken waar de politie momenteel gebruik van met maakt. Hierbij is van belang voor welk doel deze methoden en technieken zijn ingezet. Crowd sourcing kent hele andere uitgangsprincipes dan dreigingsdetectie, maar kan wel signalen van dreiging oppakken. De uitdaging in dit project is de beide processen te koppelen. Binnen het project en het experiment is gekozen voor Twitter als te monitoren platform. Ook hier dient in de vergelijking met andere methoden en technieken rekening te worden gehouden.

art. 10 lid 2 onder c en d

2.3 Juridische kaders

Bij het in gang zetten van het project monitoren 'dreigtweets' is de juridische grondslag van het monitoren van dreigtweets onderzocht door juridische zaken van de Landelijke Eenheid. De grondslag voor het monitoren ligt in art. 3 Politiewet, vooral omdat de toegepaste ontologie is gericht op potentiële en acute dreigingen gericht tegen een evenement. Het monitoren is gericht op de goede uitvoering van de politietaken, en met het oog op het evenement het handhaven van de rechtsorde.

art. 10 lid 2 onder c en d

De verwerking van de 'dreigtweets' valt onder art. 8 Wet politiegegevens. De naar voren gekomen serieuze dreigtweets zijn conform de bestaande werkwijze van het team OSINT van de Landelijke Eenheid afgehandeld, waarbij de dienstdoende (SGBO-)officier, de betrokken regio (achteraf) middels een bij team OSINT opgemaakte en geregistreerde internetrapportage in kennis wordt gesteld.

art. 10 lid 2 onder c en d

De politie heeft tot taak in ondergeschiktheid aan het bevoegd gezag en in overeenstemming met de geldende rechtsregels te zorgen voor de daadwerkelijke handhaving van de rechtsorde en het verlenen van hulp aan hen die deze behoeven (art. 3 Politiewet). Deze verantwoordelijkheid, ook bij evenementen, wordt nog eens benadrukt in de handleiding Bewaken en beveiligen, nationale evenementen (NCTV, 2011). Van belang is dat het karakter van het evenement intact blijft. De werkwijze heeft primair tot doel het evenement op een veilige manier te laten verlopen voor alle aanwezigen; de genodigden, de toeschouwers/het publiek en de hulpverlenende en handhavende instanties.

Rondom nationale evenementen bestaat een logisch verband tussen het openbare-orde-element en de veiligheid van de personen die voorkomen op de limitatieve lijst. Dreiging en effectuering daarvan op het eerste aspect kan direct invloed op de dreiging van de ander hebben en vice versa.

In beginsel wordt voor het beveiligen van nationale evenementen uitgegaan van risicobeheersing. Het uitsluiten van risico's is immers niet realiseerbaar. Het is altijd een afweging op basis van informatie, de vooraf gekozen risicotypes en weerbaarheidsscenario's die de contouren van het beveiligingsplan bepalen en tevens daarmee de risicobeheersing operationeel zichtbaar maken.

art. 10 lid 2 onder c en d

Voor dat doel biedt art. 3 Politiewet alle handvatten. Ter voorbereiding op grote en/of nationale evenementen start de politie met het verzamelen en verwerken van veiligheidsinformatie (RBP, 2012). Met de komst van het internet, in het bijzonder de sociale media, zijn vragen gerezen over de legitimiteit daarvan in relatie tot de

kerntaak van de politie. De analogie met de uitvoering in de fysieke wereld gaat op drie punten niet helemaal op. De politie verzamelt en verwerkt veel meer informatie dan in de fysieke wereld kan worden waargenomen door een individuele politiemedewerker. De zoekvragen die worden gesteld, leveren bovendien een discussie op over de mogelijke stelselmatigheid van het inwinnen van inlichtingen.
Voor het laatste is in een opsporingsonderzoek een machtiging nodig van de officier van justitie. Met andere woorden, er moet al sprake zijn van een onderzoek en van (een) verdachte(n), en moeten politie en justitie van oordeel zijn dat het verdedigbaar is dat zij voor de periode van het onderzoek bepaalde aspecten van iemands privéleven schenden. Tot slot kan op het internet ver terug worden gekeken in het verleden naar onderwerpen, personen en/of groepen.

art. 11 lid 1
art. 10 lid 2 onder c en d

De lopende discussie heeft ertoe geleid dat uit voorzorg politie en justitie een gezamenlijk standpunt hebben ingenomen over de periode van en tot wanneer OSINT onder de werking van art. 3 Politiewet haar politietaken kan en mag uitvoeren. Het betreft hier een consensusbesluitvorming in optima forma.

Indachtig de SGBO-structuur levert OSINT haar bevindingen aan de HIN, die oordeelt of en hoe maatregelen worden getroffen. Over die beslissing en onder welke juridische paraplu de opvolging plaatsvind is per casus verschillend. Bij acute dreigingen tijdens het evenement gaat het echter om het voorkomen van de dreiging en valt daarmee logischerwijs onder art. 3 Politiewet en behoeft geen bijzondere bevoegdheden van politiemedewerkers. Dat is waar het experiment zich in hoofdzaak op richt.

art. 10 lid 2 onder c en d

2.4 WOB-verzoeken

De politie heeft de nauwlettende aandacht van het journaille en maatschappelijk betrokken personen. Door middel van WOB-verzoeken kunnen vragen worden gesteld over bepaalde ontwikkelingen van en/of activiteiten door de nationale politie. WOB-verzoeken vormen een relatieve bedreiging voor deze exploratieve poging controle te krijgen op dreigingen in open bronnen. Communicatie inzake WOB-verzoeken geschiedt via juridische zaken. Tijdens het experiment zijn er geen WOB-verzoeken ingediend.

3 Resultaten

De belangrijkste resultaten van dit project monitoren 'dreigtweets' zijn afkomstig van de testmomenten Koninginnedag 2013 en Nationale Herdenking 2013. Eerst moet opgemerkt worden dat Koninginnedag 2013 een extra waarde als evenement heeft meegekregen, vanwege de troonswisseling. Om het evenement Koninginnedag (lees troonwisseling) goed te kunnen monitoren is besloten om ook de voorafgaande koninginnenacht te monitoren om zodoende nog eventuele laatste wijzigingen voor de testmomenten door te kunnen voeren. Kortom: dit leverde na afloop van het experiment een extra testmoment, namelijk drie, met bijbehorende resultaten: Koninginnenacht (KN), Koninginnedag (KD) en Nationale Herdenking (NH).

De resultaten (zowel de uitkomsten als het werkproces) laten zich het beste beschrijven in de chronologische volgorde van de werking van de dreigtweetmonitor (zie ook onderstaande tabel 1) en de gehanteerde werkmethodiek;

***Tabel 1**:schematische weergave van de filtering van twitterberichten naar serieuze dreigtweets.*

Real-time Twitter-stream		
	Acquisitie	
Originele Nederlandse tweets		
	Radboudmodule	
Geclassificeerde 'dreigtweets'		
	LOODS-dossiers	
Evenementgerelateerde 'dreigtweets'		
	Frontoffice en Backoffice	art. 10 lid 2 onder c, d en g
Geannoteerde, serieuze dreigtweets		
	Opvolging in operatieën	

Eerst vindt data acquisitie van Twitterberichten (real-time twitterstream) plaats, waarbij de originele Nederlandstalige berichten worden gefilterd. Vervolgens worden de Nederlandstalige tweets door de linguïstische dreigingsfilter van de Radboud Universiteit (de Radboudmodule, of de RBM) gefilterd. Daarna kunnen twitterberichten via de interface selectief gefilterd worden volgens het recherchekundige IRS/ LOODS-model. Dit betreft een aantal categorieën (locaties, objecten, omstandigheden, daders en slachtoffers) die in relatie tot het evenement op dat moment van toepassing zijn. Na het filteren volgens deze stappen vindt binnen het werkproces triage (sortering van de dreigtweets) door een experimentele frontoffice plaats, gevolgd door de annotatie, met mogelijk de benodigde (operationele) opvolging door een experimentele backoffice. In het experiment is rekening gehouden met de mogelijke inbedding in een FOBO-structuur.

Het versturen van een serieuze dreigtweet kan resulteren in een uiteindelijke informatierapport of, een proces-verbaal, en op grond daarvan een beslissing tot een maatregel, bijvoorbeeld de aanhouding van een 'dreiger'. Hierbij wordt opgemerkt dat het voorafgaande proces van data acquisitie tot aan de annotatie door de backoffice in de meeste gevallen niet langer dan vijftien minuten in beslag neemt. Deze snelheid is noodzakelijk voor het bepalen van de ernst en waarschijnlijkheid van bepaalde dreigingen. Hierbij moet onder meer worden gedacht aan de positie of locatie van de auteur van het bericht in relatie tot de bedreigde (persoon, object of dienst). Onderstaand volgt een uitgebreidere beschrijving van het verloop van de verschillende stappen in het werkproces.

art. 10 lid 2 onder c, d en g

3.1 Data acquisitie

Binnen data acquisitie vindt de eerste filteringsstap plaats. Van alle twitterberichten van de real-time-streaming worden de Nederlandstalige berichten (berichten verstuurd in het Nederlandse domein) als input gebruikt in de dreigtweetmonitor. Retweets, twitterberichten die worden doorgestuurd door volgers zonder iets te veranderen aan de inhoud van de boodschap (RT) worden niet worden meegenomen. Alleen originele of aangepaste RT's gelden als resultaat na de eerste filteringsstap. Na dit proces vindt het filteren door de Radboudmodule en de LOODS-componenten plaats. Op basis van handmatige beoordeling wordt uiteindelijke relevantie toegekend aan de dreigtweets. Over het residue van de verschillende filteringsstappen kan kort het volgende worden geconcludeerd:

art. 10 lid 2 onder c, d en g

Tabel 2: schematische weergave filteringsstappen

Aantal	*KN*	*KD*	*NH*	*Totaal*
Ned. tweets	-	6-7Mio	4-5Mio	-
Data Acquisitie	-	464.604	906.749	-
Radboudmodule	-	2.734	4.393	-
IRS/ LOODS	488	270	1.039	1.797

Tijdens de troonswisseling (Koninginnenacht en Koninginnedag samen) bleven er tijdens Koninginnenacht 758 te beoordelen 'dreigtweets' over.

Tijdens Koninginnedag (KD) werden 464.604 berichten aangeboden aan de RBM. Door de RBM werden 2.734 dreigingen gedetecteerd, waarvan – op grond van de LOODS-filter - 270 dreigingen zijn verwerkt die verband hielden met Koninginnedag.

art. 10 lid 2 onder c, d en g

Tijdens Nationale Herdenking (NH) gold dat voor de 4 miljoen tot 5 miljoen twitterberichten er bijna een miljoen geacquireerde, Nederlandse twitterberichten overbleven, die zijn aangeboden aan de RBM. Door de RBM werd op 4.393 dreigingen gedetecteerd, waarvan – op grond van de LOODS-filter - 1.039 berichten zijn verwerkt die verband hielden met Nationale herdenking.

Vastgesteld kan worden dat de RBM een sterke reductie van het tweets verzorgd, die relevant zijn voor dreigingsdetectie in het algemeen. Voorts levert het IRS/ LOODS-model de mogelijkheid tot het nader filteren op relevantie met het evenement.

art. 10 lid 2 onder c, d en g

Deze sterke reductie van de te beoordelen Twitterberichten door de automatische filtering, maakt het mogelijk om uiteindelijk handmatige beoordeling van mogelijke dreigtweets gerelateerd aan het evenement te kunnen doen.

3.2 Radboudmodule

Na het acquireren van de originele Nederlandstalige twitterberichten vindt er (zoals zojuist gemeld) filtering plaats door de Radboudmodule. Deze module voert een linguïstische analyse uit op de content van een tweet. Deze linguïstische analyse heeft een objectieve, wetenschappelijke basis en wordt uitgebreid beschreven in de deelevaluatie van het CLST. De Radboudmodule wijst automatisch een score toe op basis van aanwezigheid van dreigingscomponenten in het bericht, en plaatst deze vervolgens in één van de onderstaande categorieën:

Tabel 3: klasse en omschrijving Radboudmodule

art. 10 lid 2 onder c, d en g

Klasse	*Omschrijving*
0	Berichten bevatten geen elementen die dreiging suggereren en worden niet door de filter gelaten en dus niet zichtbaar als resultaat;
1-19	Berichten bevatten elementen die dreiging suggereren, maar ook sterke aanwijzingen dat de dreiging niet serieus genomen hoeft te worden;
20-49	Berichten bevatten elementen die dreiging suggereren, maar de linguïstische dreigingsfilter kan de suggestie niet bevestigen;
50-99	Berichten bevatten elementen die dreiging suggereren, ondersteund door sterke aanwijzingen vanuit de linguïstische dreigingsfilter.

De berichten die voldoen aan de beoordeling groter dan 0, dus 1 of hoger, de laatste drie categorieën, worden als resultaten in de dreigtweetmonitor getoond.
Onderstaande tabel laat de classificering van de aantallen verwerkte berichten zien.

Tabel 4: schematische weergave classificering berichten Radboudmodule

Klasse	*KN*	*KD*	*NH*	***Totaal***
Totaal	-	**2.734**	**4.393**	-

art. 10 lid 2 onder c, d en g

3.3 LOODS-model

Om te voorkomen dat heel Twitter op dreigingen gescreend moet worden en om te voorkomen dat de front- en backoffice te maken krijgen met andere dreigingen dan dreigtweets gericht op het te monitoren evenement, is er de mogelijkheid om de dreigtweets specifiek verder te filteren. De filtering vindt plaats aan de hand van het LOODS-model waarin specifiek gerelateerde informatie over het evenement centraal staat. In de interface worden handmatig filters (dossiers) gebouwd gericht op de voor het evenement van belang zijnde locaties, objecten, omstandigheden, daders en slachtoffers.

Daarnaast wordt een dossier aangemaakt met trefwoorden van belang voor evenementgerelateerde dreigingen. Per dossier zijn de (evenement-)gerelateerde trefwoorden, bijnamen, afkortingen, foutgeschreven of andere schrijfwijzes van namen verzameld. In het bijzonder voor Koninginndag is in nauwe en intensieve samenwerking met Eenheid Amsterdam een uitgebreid dossier tot stand gekomen. De dossiers zijn in de dreigtweetmonitor apart te gebruiken en te allen tijden aan en uit te zetten, zodat tijdens het monitoren het leggen van dwarsverbanden tussen de verschillende dossiers en mogelijke dreigingen mogelijk is. Voor het adequaat vullen van de dossiers is een gedegen voorbereiding nodig, waarbij (politie-)ervaring, expertise en voorkennis met betrekking tot het evenement onontbeerlijk is.

art. 10 lid 2 onder c, d en g

Tabel 5: schematische weergave classificering berichten IRS/ LOODS

Klasse	*KN*	*KD*	*NH*	***Totaal***
1-19	12	38	43	**93**
20-49	37	153	342	**532**
50-99	439	79	654	**1.172**
Totaal	**488**	**270**	**1.039**	**1.797**[1]

[1] Met behulp van webmonitoringtools zijn op Koninginnedag ruim 80.000 berichten waargenomen die verband hielden met locatie, objecten, omstandigheden en potentiële slachtoffers.

art. 10 lid 2 onder c, d en g

3.4 Triage

Aan de hand van de bovenbeschreven categorisering en filtering op basis van de LOODS-dossiers vindt in de frontoffice vervolgens triage plaats. De triage wordt uitgevoerd door het team OSINT welke als taakstelling heeft om onder andere dreigingen op het internet te detecteren. De triage bestaat uit een handmatige beoordeling van de gecategoriseerde berichten, waarbij de dreigtweets die een hogere classificatie door de RBM krijgen, eerder worden beoordeeld. De kleurcodering wordt tijdens triage slechts toegekend op basis van de *platte inhoud* van het bericht. Tijdens triage kent de medewerker frontoffice aan ieder bericht een kleurcodering koppelt: groen, geel, oranje of rood. De betekenis geldt als volgt:

Tabel 5: Gebruik van kleurcodering tijdens triage.

Klasse	*Omschrijving*
1	Geen dreiging
2	Dreiging, lage prioriteit
3	Dreiging, gemiddelde prioriteit
4	Dreiging, hoge prioriteit

art. 10 lid 2 onder c, d en g

Bij het toekennen van een kleurcode wordt de beoordeling kort toegelicht in de interface. Wanneer een 'dreigtweet' in deze fase toch geen dreiging blijkt te bevatten, wordt deze als groen geannoteerd en verdwijnt deze uit de interface. Een dreigtweet moet namelijk contextueel geinterpreteerd worden en daarnaast moet er een verband zijn tussen de mogelijke dreiging en het evenement. De automatische filtering binnen de DTM kan namelijk geen contextuele interpretatie waarnemen. Die contextuele interpretatie (of 'wereldkennis' zoals genoemd door CLST) is bijvoorbeeld een humoristische, cynische of sarcastische bedoelde inhoud van een tweet die door de Radboudmodule als dreigtweet wordt aangemerkt. Maar ook als-dan-constructies, logica, antisemitische of holocaust-revisionistische opmerkingen, reacties en quotes in de dreigtweet vallen hieronder. Als voorbeeld worden 'dreigtweets' die humoristisch zijn en geen dreiging bevatten, tijdens triage beoordeeld als groen met de toelichting 'humor'.

art. 10 lid 2 onder c, d en g

Bovendien zijn er als 'dreigtweet' gedefinieerde Twitterberichten die ondanks de filtering via de evenementgerelateerde LOODS-dossiers, geen verband hebben met het te monitoren evenement. Ook deze zullen tijdens triage als groen worden beoordeeld en gescheiden worden van de voor het evenement relevante 'dreigtweets'.

Een aanvullende opmerking hierbij is dat tijdens de triage groen beoordeelde dreigtweets 'verdwijnen' uit de interface, maar dat ze worden weggezet in de (log)data van de dreigtweetmonitor. Door deze technische mogelijkheid is analyse van de als 'groen'-beoordeelde dreigtweets achteraf mogelijk.

art. 10 lid 2 onder, c, d en g

3.5 Annotatie

Na triage vindt annotatie in de backoffice plaats. Dit betreft de berichten waaraan een gele, oranje en rode kleurcode is toegekend. De medewerkers van de backoffice zoomen dieper in op het bericht, en relateren de inhoud van de boodschap aan de auteur (leeftijd, geslacht), de locatie, twittergeschiedenis, foto's, etc. Daarnaast wordt gekeken naar mogelijke maatschappelijke context, zoals fixatie en frustratie en waar nodig naar de sociale en fysieke omgeving van de auteur.

Tabel 6: Overzicht van de door de DTM aangeboden berichten ter beoordeling van de front- backoffice-medewerkers.

Klasse	*KN*		*KD*		*NH*		*Totaal*	
	Triage	**Annotatie**	**Triage**	**Annotatie**	**Triage**	**Annotatie**	**Triage**	**Annotatie**
1-groen	437	11	242	6	929	29	1.608	46
2-geel	41	12	27	28	94	53	162	93
3-oranje	9	-	1	1	11	7	21	8
4-rood	1	4	-	1	5	8	6	13
Totaal	**488**	**27**	**270**	**36**	**1.039**	**97**	**1.797**	**160**

art. 10 lid 2 onder c, d en g

Op KN zijn er 27 (6%), op KD 36 (13%) en tijdens NH 97 (9%) berichten in de backoffice door nader 'onderzoek' beoordeeld ('annotatie'). Van alle berichten is met andere woorden 9% (n=160) door de backoffice nader beoordeeld op serieus te nemen dreiging en mogelijk vervolgonderzoek.

Berichten met een rode annotatie zullen direct worden doorgecommuniceerd van de frontoffice naar de backoffice om met prioriteit te worden onderzocht. Uit de tabel blijkt dat dit in totaal 13 keer heeft plaatsgevonden. Van deze 13 berichten zijn uiteindelijk 8 informatierapporten opgemaakt.

art. 10 lid 2 onder c, d en g

Wanneer verder internetonderzoek de beoordeling van de dreigtweet ondersteunt en de dreiging ernst en waarschijnlijkheid wordt toegedicht, wordt deze onmiddellijk voorgelegd aan de officier van dienst informatie (OVDi) of HIN (in geval van een SGBO). De HIN is binnen de staf verantwoordelijk voor de verstrekking van potentiële dreigingen aan de algemeen commandant, die op zijn of haar beurt weer verantwoordelijk is voor de te treffen maatregelen. In acute gevallen wordt de OVDi en/ of HIN eerst telefonisch op de hoogte gesteld. In alle gevallen wordt deze melding uiteindelijk begeleid door een informatierapport over de dreigtweet, de onderzochte omstandigheden en de relatie tot een bedreigde object of persoon. Ook de berichten met een lage (gele) of gemiddelde (oranje) prioriteit kunnen naderhand nog leiden tot een informatierapport en vervolgonderzoek, bijvoorbeeld omdat wordt geoordeeld dat het een weliswaar niet acute, maar wel een strafbare bedreiging betreft. Deze misdrijven kunnen ambtshalve ook aanhouding buiten heterdaad (art 285 WvS) worden geadresseerd.

3.6 Evenement-relevantie

Zoals beschreven in paragraaf 3.4 en 3.5 bepaalt uiteindelijk handmatige beoordeling of een als dreigtweet gedefinieerd Twitterbericht in de context (wereldkennis of evenement-relevantie) op dat moment interessant genoeg is om tot vervolgonderzoek over te gaan.

Vanuit de monitoring op dreigtweets en het bepalen van de relevantie van de dreiging op de evenementen KN, KD en NH is een overzicht gemaakt van de voor het evenement relevantie dreigtweets afgezet tegen het aantal te beoordelen dreigtweets vanuit triage en annotatie.

art. 10 lid 2 onder c, d en g

Tabel 7: Overzicht van de te beoordelen tweets en het aandeel van voor het evenement geldende relevante dreigtweets.

Klasse	*KN*	*KD*	*NH*	*Totaal*
Totaal	488	270	1.039	1.797
Relevant	37	235	448	720
Percentage	8%	87%	43%	40%

Het totaal aantal berichten per evenement betreft de berichten die ter beoordeling zijn aangeboden aan de front- en backoffice-medewerker(s). Ook is te zien welke van deze berichten door de frontoffice-medewerker(s) zijn beoordeeld als relevant voor het evenement (dus een gele, oranje of rode categorisering kennen). Tijdens Koninginnenacht blijkt het percentage als relevant beoordeelde berichten met 20% aan de lage kant, met als verklaring dat binnen de DTM veel dreigtweets werden geduid, die geen of te weinig betrekking hadden op het evenement. Uit de direct daaropvolgende (technische) analyse binnen de DTM bleek dat na de data acquisitie op KN werd getrechterd op grond van de LOODS-dossiers en niet eerst op het dreigingsfilter. Door dit te wijzigen (eerst filtering door de Radboudmodule, daarna selectie op basis van de LOODS-dossiers in zowel technisch opzicht als in het werkproces) was het percentage van als relevant beoordeelde berichten tijdens KD veel hoger, namelijk 87%.

art. 10 lid 2 onder c, d en g

Het percentage als relevant beoordeelde berichten op Nationale Herdenking is 43%. Dit is lager dan het percentage op KD, maar vanuit het experiment goed te verklaren: Vooral tijdens NH is gebleken dat de voorbereidingstijd om de LOODS-dossiers te vullen te kort was. Hierdoor konden de LOODS-dossiers niet voldoende worden gevuld met de kennis en expertise vanuit de organisatie. Het gevolg was dat er op bredere schaal is gefilterd dan specifiek voor het evenement mogelijk was geweest. Door het niet adequaat inzetten van de LOODS-dossiers, trad er vervuiling op: meer dan de helft van de te beoordelen dreigtweets was niet relevant voor NH. Bovendien heeft Nationale Herdenking als evenement een ander karakter dan Koninginnenacht/-dag en wordt er vanuit de nationale politie en overige eenheden (uitgezonderd Eenheid Amsterdam) weinig tot geen voorbereidingen getroffen op dergelijke nationale evenementen in relatie tot het internet. De ervaring hiermee is nog onvlodoende.

art. 10 lid 2 onder c, d en g

3.7 Algemene resultaten

Uit bovenstaande blijkt nogmaals dat het verzamelen en gebruiken van kennis en expertise van (politie)professionals die ervaring hebben met evenementgerelateerde dreigingen, onmisbaar zijn voor de meest optimale inzet van de dreigtweetmonitor.

art. 11 lid 1

Ook mag geconstateerd worden dat het aantal *false positives* laag genoeg is waardoor de DTM in het werkproces kan worden ingevoegd. Een onzekerheid binnen het werkende systeem is wel dat er tijdens het experiment te weinig wetenschap is verzameld over de berichten die NIET door de DTM als te beoordelen berichten worden aangeboden (*false negatives*). In de deelrapportage van het CLST wordt hier nader op ingegaan (sectie Recall). Om dit aspect goed te kunnen doorzien zal verder onderzoek nodig zijn.

art. 11 lid 1
art. 10 lid 2 onder c, d en g

Verder is het belangrijk te vermelden dat de resultaten die worden besproken in de deelevaluatie van CLST afwijkingen vertoont met de hierbeschreven resultaten. In grote lijnen komen ze wel overeen, maar voor de verschillen is een aantal redenen op te noemen:

- Naar beide partijen zijn de originele bestanden gestuurd vanuit Tardis (vanuit de historie gegenereerd door de software). Het betreft drie bestanden die zijn samengevoegd en gegevens bevatten die op basis van het IRS/LOODS-model zijn gefilterd en door front- en backoffice zijn beoordeeld. Hierin zijn ook 178 berichten meegenomen die volgens de Radboudmodule geen dreiging bevatten;
- De door OSINT gebruikte tijdsperiode/testmomenten voor de Troonswisseling is opgedeeld geweest in Koninginnenacht en Koninginnedag. In de evaluatie van CLST is alleen Koninginnedag meegenomen;

art. 11 lid 1
art. 10 lid 2 onder c, d en g

- Koninginnenacht (KN) is door OSINT afgebakend tot 29-4-2013 om 23:59 uur, maar bevat ook enkele berichten die bij wijze van test zijn beoordeeld ter voorbereiding op Koninginnedag en Nationale Herdenking;
- KD is afgebakend van 30-04-2013 om 00:00 uur tot 30-04-2013 om 00:00 uur. De eerste dreiging, met een dreigingsscore van 59 wordt gezien op 30 april 2013 om 04:29 uur en de laatste, met een dreigingsscore van 20 wordt gezien op 30 april 2013 om 22:02 uur;
- NH van 04-05-2013 om 00:00 uur tot 05-05-2013 om 00:00 uur. De eerste dreiging, met een dreigingsscore van 59 wordt gezien op 4 mei 2013 om 04:29 uur en de laatste, met een dreigingsscore van 55 wordt gezien op 4 mei 2013 om 19:57 uur. Ook voor Nationale herdenking geldt dat er enkele berichten zijn beoordeeld uit de periode tussen Koninginnedag en Nationale Herdenking. Dit ter voorbereiding op Nationale Herdenking zelf.
- OSINT en Tardis hebben gefocust op de dreiging als ook op de relatie met het evenement, waar CLST alleen heeft gekeken naar de dreiging.

art. 11 lid 1
art. 10 lid 2
onder c, d en
g

Kortom de bestaande verschillen kunnen worden verklaard door operationele definitieverschillen, dienstverbanden, en gekozen tijdsperiodes om de resultaten te presenteren. Vooralsnog zijn de resultaten grotendeels hetzelfde.

3.8 Samenvatting

Gekeken naar de manier waarop de DTM in te zetten is als hulpmiddel voor het monitoren van het internet (in dit geval Twitter) op bedreigingen tijdens een evenement, is het project monitoren 'dreigtweets' succesvol. Ondanks de genoemde onvolkomenheden zijn de gemeten hitratio's tevredenstellend. Een verwerking van alle aandachtbehoevende tweets op de testdagen, tijdens de nationale evenementen is haalbaar gebleken. De dreigtweetmonitor werkt naar wens en kan snel (real-time) en adequaat relevante dreigtweets extraheren zodat individuele toetsing van inhoud en context mogelijk is. Het filteringsproces waarbij na de data acquisitie meteen gebruik wordt gemaakt van de Radboudmodule, ondersteunt de ambitie om in de toekomst niet alleen te kijken naar evenementgerelateerde dreigingen, maar ook generiek naar dreigingen te zoeken. Door het inbedden in het werkproces van de nationale politie kan waar nodig gerichte opvolging in de operatieën plaatsvinden.

art. 11 lid 1
art. 10 lid 2
onder c, d en
g

4 Organisatie

4.1 Inrichting

Twee belangrijke condities voor succes van de organisatie zijn de inrichting en de mate waarin de organisatie het juiste evenwicht in de voortgang vindt tussen voortdurende aanpassingen en stabiliteit. Het experiment vraagt om samenwerking en vertrouwen tussen verschillende partijen met verschillende doelstellingen, waarbij het bereiken van succes onlosmakelijk verbonden is met de door de partijen gevormde organisatie.

De projectorganisatie van het project monitoren 'dreigtweets' was een samenwerkingsverband van de overheid werkzaam binnen het veiligheidsdomein (nationale politie, Landelijke Eenheid), een wetenschappelijk instituut (CLST, Radboud Universiteit Nijmegen) en een commerciële IT-partner (Tardis Research). Vanuit elke 'tak van sport' werd een verantwoordelijke benoemd voor een goed verloop van de samenwerking. De wens bestond om een onafhankelijke projectleider aan te wijzen die de overview kon houden. Uiteindelijk is door de opdrachtgever besloten om de aangewezen verantwoordelijke vanuit de nationale politie ook als eindverantwoordelijke aan te stellen. Het in het projectvoorstel voorgestelde team is als gevolg met één wijziging gelijk gebleven; [redacted] heeft niet aan het project meegewerkt.

art. 10 lid 2 onder e

- Projecteindverantwoordelijke/-leider overheid – [redacted], team OSINT, LE
- Projectmedewerker/algemene coördinatie – [redacted], team OSINT, LE (stagiaire)
- Projectmedewerker overheid – [redacted], team OSINT, LE
- Projectverantwoordelijke CLST – [redacted], CLST, Radboud Universiteit
- Projectmedewerker CLST – [redacted], CLST, Radboud Universiteit
- Projectverantwoordelijke Tardis – [redacted], Tardis Research

4.2 Samenwerking extern werkproces

Gedurende het werkproces is er veelvuldig overleg gepleegd. In eerste instantie (tijdens de beginperiode van het project) is er door de projecteindverantwoordelijke veel contact geweest met de subsidieverstrekker en de opdrachtgever. Door interne wijzigingen binnen de nationale politie, traden in deze periode ook verschuivingen op in het opdrachtgeversschap en de manier waarop gebruik gemaakt kon worden van de verstrekte subsidie. De opstart van het project en als gevolg de voortgang in het werkproces verliep stroef. Lange tijd bestond er bij de betrokken partijen onzekerheid over de doorgang van het project en de subsidiëring, waarbij de projecteindverantwoordelijke veelvuldig gebruik heeft moeten maken van het vertrouwen van de samenwerkende partijen. Dit proces heeft een aantal keren geleid tot escalaties waarmee de projectverantwoordelijke uiteindelijk continuïteit van het project en het officieel starten van het werkproces heeft afgedwongen bij de opdrachtgever en de subsidieverstrekker.

art. 11 lid 1
art. 10 lid 2 onder g

Door de geografische verspreiding van de betrokken partijen (Den Haag, Zoetermeer, Nijmegen) hebben afspraken en overleggen grotendeels telefonisch plaatsgevonden. Een aantal keer heeft de groep ook fysiek vergaderd (een contractbespreking en twee voorbereidingsbijeenkomsten). Vanuit de projecteindverantwoordelijke is er diverse malen vinger aan de pols van het project gelegd en waar nodig zijn adequate maatregelen getroffen om de voortgang van het project te garanderen. Bij (interne) escalaties is te allen tijde de opdrachtgever op de hoogte gebracht van de situatie en heeft de projectverantwoordelijke zorg gedragen voor een gepaste oplossing.

De gekozen fysieke inrichting, namelijk ontwikkelen vanuit drie geografische locaties vergde een grote mate van zelfstandigheid en deelverantwoordelijkheid voor hetgeen van de deelnemende partijen verwacht werd. In de basis is dit goed verlopen. Echter, terugkijkend, was het efficiënter geweest om ook de op te leveren delen en de inrichting van het project meer te stroomlijnen op gezamenlijke voortgang, deelbaarheid en communicatie. De onafhankelijkheid is de gezamenlijke verantwoordelijkheid voor het project niet altijd ten goede gekomen. art. 11 lid 1

Vanuit het projectplan was de vooraanname gedaan dat het projectteam maandelijks zou rapporteren over de stand van zaken. Deze rapportage werd vanwege de korte doorlooptijd en deadlines in overeenstemming met alle partijen telefonisch en per email uitgevoerd. De mate van contact met Tardis Research was bij aanvang van het experiment intensiever (bijna op dagelijkse basis) dan met het CLST. In het bijzonder de geografische nabijheid van Tardis Research (Den Haag) en de benodigde (opstart)samenwerking op technisch vlak verklaren dit.

Al deze aspecten, de doorlooptijden, de technische vereisten, als ook de beschikbare tijd van de afzonderlijke projectmedewerkers, het intensievere contact van de projectleider met Tardis Research hebben af en toe tot spanning tussen de samenwerkende partijen geleid. Achteraf bezien, zou het voor de onderlinge verhoudingen beter zijn geweest om formele contactmomenten in een vaste frequentie te laten plaatsvinden. In het projectplan werd omschreven dat benodigde bemiddeling tussen de twee externe partijen (Tardis en Radboud) niet onder de verantwoordelijkheid van de projectverantwoordelijke zou moeten vallen. Om het experiment tot een succesvol einde te begeleiden heeft interventie door de projecteindverantwoordelijke toch een aantal keren moeten plaatsvinden. Toch kan geconstateerd worden dat, ondanks de ontstane spanningen, beide partijen hun professionaliteit voorrang hebben gegeven. Zo is in de dagen dat de Radboudmodule in de DTM werd ingebed, intensief samengewerkt om dit deel van het project tot een goed einde te brengen. art. 11 lid 1

In de voorbereiding en tijdens het draaien van de testen op Koninginnedag en Nationale Herdenking zijn diverse meedenkers/ potentiële afnemers van de dreigtweetmonitor betrokken geweest in de ontwikkelingen van het experiment. Een van de partijen (DKDB) is ook fysiek aanwezig geweest tijdens de testdagen. Zij hebben de kans gekregen om een 'real-time' demonstratie van de dreigtweetmonitor te zien. 10 lid 2 onder c en d

Op diverse momenten is er door de samenwerkende partijen contact gezocht met de projectverantwoordelijke over de ontwikkelingen in de 'randen' van het project. De Radboud Universiteit krijgt de kans om een wetenschappelijke publicatie te doen over hun bijdrage aan de dreigtweetmonitor. De bescherming van auteurs- en ontwikkelrechten valt binnen het project niet onder de verantwoordelijkheden van de projectleiding. Alle partijen hebben dit voor hun eigen bijdrage aan het project moeten bewaken.

Tijdens de uitvoering van het project was een gestelde voorwaarde dat de dreigtweetmonitor uiteindelijk volledig (onafhankelijk van ontwikkelpartijen) te gebruiken is binnen het werkproces van de nationale politie. Met het oog op internetgebruik houdt dat concreet in dat de dreigtweetmonitor bijvoorbeeld moet kunnen draaien op het IRN-netwerk of soortgelijk politienetwerk. Dit netwerk is goedgekeurd voor het veilig en vertrouwd gebruiken van internet voor opsporing en art. 10 lid 2 onder c, d en g

handhaving. De projectmedewerkers van team OSINT, Landelijke eenheid hebben voor het kunnen voldoen aan deze voorwaarde veel tijd en aandacht besteed. Het systeem waarop de dreigtweetmonitor draaide bleek in voorbereiding op de testdagen instabiel waardoor het projectteam pas zeer kort voor de testdagen daadwerkelijk voorbereidende handelingen ten aanzien van de dreigtweetmonitor kon plegen.

art. 11 lid 1
art. 10 lid 2
onder g

4.3 Samenwerking intern werkproces

Om de dreigtweetmonitor te kunnen inzetten was er sprake van een intern (operationeel) werkproces. Voor het adequaat monitoren van mogelijke dreigtweets werd er een frontoffice en een backoffice in de uitvoering van het werk ingericht. De filterstappen triage en annotatie (zie hoofdstuk 3, alinea's triage en annotatie) zijn hoofdzakelijk 'mensenwerk' welke op de werkvloer moet worden ingericht. De projectverantwoordelijke van team OSINT heeft ervoor gekozen om de front- en backoffice fysiek in dezelfde ruimte te plaatsen. Hierdoor verliep de communicatie tussen deze twee organisatieonderdelen zeer goed. Ook bleek het inzetten van een coördinator naast medewerkers van de front- en backoffice succesvol. De coördinator onderhield contact met de dienstdoende officier van het draaiende SGBO en kanaliseerde waar nodig de communicatie tot heldere en bondige meldingen. Zo konden de medewerkers in de frontoffice en de backoffice zich volledig focussen op de aan hen gestelde taken. Tijdens en vlak na de testdagen van de dreigtweetmonitor bleek dat de bestaande SGBO-structuren (voor Koninginnedag en Nationale Herdenking) moesten wennen aan het hebben van een extra informatiestroom vanaf het internet. Vanuit het team OSINT moest actief gecommuniceerd worden om de ernstige dreigtweets ook op te pakken binnen de bestaande werkprocessen tijdens de evenementen.

art. 11 lid 1
art. 10 lid 2
onder c, d en g

art. 10 lid 2
onder c, d en g

4.4 Fasering

Het experiment heeft een aantal fases doorlopen die gebonden waren aan tijdsgerelateerde evenementen. De uiteindelijk afrondende testdagen lagen vast, namelijk Koninginnedag en Nationale Herdenking, en het daadwerkelijk uitvoeren van de testen heeft op deze dagen ook plaatsgevonden. Hieronder volgt een korte beschrijving van het verlopen van de fasering:

- September 2012, Prinsjesdag: een ruwe versie van de monitor 'dreigtweets' is getest. De verzamelde gegevens vormen een opmaat voor de doorontwikkeling van de uiteindelijke dreigtweetmonitor.
- December 2012: Het projectplan monitoren 'dreigtweets' wordt goedgekeurd door de opdrachtgever en het projectteam gaat de monitor verder ontwikkelen. Enige onzekerheid over de voortgang bestaat vanwege interne organisatieveranderingen binnen de nationale politie en daarmee de goedkeuring van het project.
- Maart 2013: Het projectplan monitoren 'dreigtweets' wordt officieel goedgekeurd door de nationale politie en het projectteam gaat verder met de ontwikkelingen.
- Medio April 2013: de potentiële gebruikers (waaronder de DKDB en de iTrechter) van de monitor stellen diverse zoekvragen en/of lijsten samen om te komen tot de juiste berichten met het oog op de troonswissleing en NH 2013.
- April 2013, daags voor Koninginnedag: De zo goed als definitieve monitor 'dreigtweets' wordt in voorbereidende werking gesteld (tijdens Koninginnenacht), om op Koninginnedag 'officieel' getest te kunnen worden. Bij deze operationele exercitie zijn beide samenwerkende partijen aanwezig geweest, om waar nodig ondersteuning te verlenen.
- Mei 2013, Nationale Herdenking: De definitieve 'dreigtweetmonitor' wordt opnieuw voorbereid en in werking gesteld. Via IRN wordt ook op regionaal niveau gebruik gemaakt van de monitor (toetsing op afstand).

art. 10 lid 2
onder c en d

- Juni 2013: De evaluatie van het project monitoren 'dreigtweets' wordt in samenwerking met de betrokken partijen uitgevoerd en gerapporteerd, met name voor de opdrachtgever als ook de subsidiegever. In de korpsleiding zal op basis hiervan worden besloten of implementatie van de dreigtweetmonitor in het werkproces van de nationale politie bij dreigingen op evenementen gewenst is.

Alle partijen waren gebonden aan het strakke tijdschema. Iedere partij heeft daarin zijn eigen verantwoordelijkheid genomen, waarbij de projectverantwoordelijke bijstuurde waar nodig. De 'onderdelen' zijn volgens de geplande fasering opgeleverd en het uiteindelijke product, de dreigtweetmonitor, heeft op de testdagen succesvol gedraaid.

5 Randvoorwaarden

5.1 Proces

De medewerkers van het projectteam zijn (wetenschappelijke) professionals binnen hun eigen werkdomein. Vanuit deze professionele houding is vanuit ieders perspectief voldoende betrokkenheid getoond. Het werkproces en ieders eigen bijdrage heeft uiteindelijk geresulteerd in een naar wens werkende dreigtweetmonitor. Ook vanuit het werkveld (organisatieonderdelen zoals DKDB en DLIO (Driebergen) is bereidheid getoond te investeren in het experiment, door met gerichte werkopdrachten aan te sluiten. De dreigtweetmonitor kan ingezet worden voor de taken binnen het rijksdomein. Kanttekening is dat het geteste proces voor sommige diensten/deelopdrachten verfijnder en specifieker moet worden. Dit is afhankelijk van tijd- en geldinvestering maar zeker realiseerbaar, en op adviesbasis van deze evaluatie zal het projectmanagement er goed aan doen om te zorgen voor een goede begeleiding van de onderlinge interne en externe projectcommunicatie. Hoe vaker de dreigtweetmonitor wordt ingezet binnen het rijksdomein, hoe soepeler en verfijnder het bijbehorende werkproces kan worden ingericht.

art. 10 lid 2 onder c en d

art. 11 lid 1
art. 10 lid 2 onder c, d en g

5.2 Inbedding in de organisatie

Verschillende diensten binnen de nationale politie hebben tijdens het project interesse getoond in de dreigtweetmonitor. Tijdens de voorbereidingen van Koninginnedag 2013 en Nationale Herdenking 2013 is ook gebleken dat de DTM en de resultaten inzetbaar zijn voor een bredere doelgroep dan alleen team OSINT. Ambtenaren die werkzaam zijn binnen het rijksveiligheidsdomein, kunnen worden voorzien van de juiste real-time informatie als het gaat over dreigingen op het internet, in dit geval Twitter.

art. 10 lid 2 onder c, d en g

Voorwaarden voor het optimaal inzetten van de DTM binnen de organisatie zijn een helder doel van de vrager, voldoende voorbereidingstijd, een evenement-professioneel en DTM-ervaren team/leidinggevende, en dat bij voorkeur vertrouwelijke informatie over mogelijke dreiging beschikbaar is om te gebruiken. Wanneer inbedding in de organisatie gewenst is, is het raadzaam ook te investeren in communicatie over de mogelijkheden van de DTM voor de organisatie. En er zal nagedacht moeten worden over (vaste) personele bezetting voor het optimaal benutten van de dreigtweetmonitor.

art. 11 lid 1
art. 10 lid 2 onder c, d en g

Bovenstaande laatste twee opmerkingen komen voort uit de ervaring dat tijdens en vlak na de testdagen van de dreigtweetmonitor bleek, dat de bestaande SGBO-structuren (voor Koninginnedag en Nationale Herdenking) moesten wennen aan het hebben van een extra informatiestroom met betrekking tot dreigingen via het internet.

art. 10 lid 2 onder c, d en g

Tardis heeft zorg gedragen voor het kunnen draaien van de dreigtweetmonitor binnen het Internet Research Netwerk (IRN). Vanuit dit gegeven is het technisch mogelijk de DTM op elke IRN-computer te kunnen draaien. Dit zal met name interessant zijn voor de teams die zich vanuit hun taakstelling dagelijks bezighouden met dreigingen en internetonderzoek. Wanneer inbedding van de dreigtweetmonitor gewenst is, zal rekening moeten worden gehouden met een mogelijk service level agreement van de partijen die een aandeel hebben in de DTM. Onderhoud, training, doorontwikkeling, helpdesk-achtige constructies vormen voorbeelden van.

5.3 Communicatie

Communicatie over het project monitoren dreigtweets heeft op meerdere manieren plaatsgevonden. Uiteraard is tijdens het uitvoeren van het project betrokkenheid geweest van diverse diensten die potentiële klant van de dreigtweetmonitor kunnen

worden. Daarbij is er door de samenwerkende partijen op ieders eigen manier gecommuniceerd over het project. De Radboud Universiteit heeft in overleg met de projectverantwoordelijke gekozen om na afronding van het project te bekijken of het mogelijk is een wetenschappelijke publicatie over hun aandeel in het project te doen. Ook Tardis heeft vanuit zijn perspectief veelvuldig gecommuniceerd over de mogelijkheden die de dreigtweetmonitor biedt, zowel binnen de politieorganisatie als daar buiten. Daarnaast zijn er diverse verzoeken vanuit de landelijke pers geweest. Over deze verzoeken is met het projectteam afgesproken dat ze via de Landelijke Eenheid en de projectverantwoordelijke worden opgepakt. Zo zijn er onder andere artikelen verschenen met betrekking tot het experiment in het magazine Nationale Veiligheid en Crisisbeheersing (februari 2013), NRC Next (februari 2013) en het blad Opportuun (februari 2013).

Als afsluiter van het experiment geldt uiteraard de evaluatie als belangrijk communicatief instrument. Hierin wordt op zo objectief mogelijke basis een realistisch oordeel gegeven over het project, waarbij logischerwijs niet alleen de loftrompetten klinken.

Tijdens de uitvoering van het experiment is een digitale map bijgehouden, waarin de historie van het project terug te vinden is. Opgeslagen versies van projectplannen en presentaties zijn hierin terug te vinden. Ook e-mail(uitwisselingen) en digitale correspondentie aangaande het project zijn door de projecteindverantwoordelijke verzameld en opgeslagen. In het projectplan was ook als deel van de documenten mogelijke gebruikershandleidingen voorzien. Echter deze zijn niet opgeleverd gedurende het project.

5.4 Financiering

Binnen het projectplan is de voorgestelde financiering goedgekeurd. Tijdens de uitvoering is het beschikbaar gestelde budget niet overschreden. Tardis heeft aangegeven dat er uiteindelijk meer facetten binnen de dreigtweetmonitor zijn ontwikkeld dan in de oorspronkelijke afspraken de bedoeling was. Het CLST heeft aangegeven dat zij meer tijd en aandacht dan begroot hebben moeten besteden aan het goed werkend krijgen van de Radboudmodule. Vanuit het team OSINT is er meer tijd en aandacht dan begroot besteed aan hoofdzakelijk het testen en terugkoppelen van de benodigde functionaliteiten voor de eindgebruiker en de bemiddelende rol zoals boven beschreven. De opdrachtgevers hebben de subsidie volledig verstrekt aan de projecteindverantwoordelijke (in dit geval de nationale politie, Landelijke Eenheid), welke op zijn beurt de gereserveerde bedragen aan de twee externe partijen (Tardis en de CLST) heeft verstrekt, dan wel zal verstrekken.

Voor het eventueel implementeren en doorontwikkelen van de dreigtweetmonitor zal een nieuw budget beschikbaar gesteld moeten worden. De financiering die in dit project benodigd was, is volledig gebruikt om de dreigtweetmonitor in de praktijk te testen met de nadruk op de technische basis. Bij het gebruik van de DTM in de werkprocessen van bijvoorbeeld de nationale politie, zal een groot deel van de financiering besteed moeten worden aan training, voorbereiding en onderhoud. Technische doorontwikkeling ten aanzien van de filterkwaliteit en het gebruiksgemak valt hier ook onder. Wanneer ook op decentraal niveau gebruik van de DTM gemaakt moet worden, zal hiervoor apart budget benodigd zijn. Ditzelfde geldt voor de mogelijkheid om de DTM in te zetten op ofwel andere sociale media-platformen of bijvoorbeeld het toevoegen van meerdere taalgebieden aan de DTM.

art. 11 lid 1
art. 10 lid 2 onder c, d en g

6 Conclusies en aanbevelingen

6.1 Conclusies

Het project monitoren 'dreigtweets' is geslaagd. De op de testdagen ingezette methoden en technieken zijn onderzocht en geëvalueerd. De resultaten zijn daadwerkelijk bruikbaar gebleken binnen politievervolgonderzoeken, waarmee het bestaansrecht van een monitoringstool als de DTM op de werkvloer is gevalideerd. De ontwikkelde dreigtweetmonitor werkt naar behoeven en de resultaten zijn succesvol. Het taalwetenschappelijk onderdeel is getest en in een deelrapportage geëvalueerd door CLST. De ontwikkelde technieken zijn door Tardis beschreven in een deelrapportage.

De DTM bestaat uit een automatische filtering van dreigingen op het internet, ingericht en voorbereid op basis van bestaande kennis en expertise op dit veiligheidsdomein, waarbij de resultaten, op wetenschappelijke basis ondersteund, worden getoond in een gebruiksvriendelijke interface, die ingezet kunnen worden in de bestaande werkprocessen binnen de nationale politie.

Uiteindelijk bepaalt handmatige beoordeling of een als dreigtweet gedefinieerd Twitterbericht in de context (wereldkennis of evenement-relevantie) op dat moment interessant genoeg is om tot vervolgonderzoek over te gaan. art. 10 lid 2 onder c en d

Door het gebruik van de DTM heeft de nationale politie de mogelijkheid om meer controle te krijgen op het volume aan serieuze dreigingen op het internet, geuit via Twitter. Het inzetten van de DTM bij de politietaak is te realiseren via IRN of soortgelijke politie-infrastructuur. Dit waarborgt het gebruik van de DTM met betrekking tot optimale beveiliging en de continuïteit binnen de politieorganisatie. art. 10 lid 2 onder c en d

Voor het testen van dreigingen op Twitter is gekozen vanwege het korte berichtgevingskarakter. Wanneer deze berichten op dreigingscomponenten te classificeren zijn, zal dit waarschijnlijk ook mogelijk zijn voor berichtgeving van meer dan 140 leestekens. art. 10 lid 2 onder c en d

De monitor kan in de toekomst wellicht ingezet worden voor wetenschappelijk onderzoek op internetgedrag, twitter en dreigingen via (andere) open bronnen op het internet. Ook het verder trainen en verfijnen van de functionaliteiten van de DTM behoren tot de mogelijkheden.

Met een dreigtweetmonitor kan de overheid haar controle op het rijksdomein via moderne middelen verbreden. Vanuit de missie, visie en strategie van de nationale politie is bewaken en beveiligen een kerntaak en zal een moderniseringsslag, zoals het gebruik van de DTM als monitoringstool, een welkome aanvulling bieden.

Naast de beschreven resultaten volgt hier een beknopt overzicht van de SWOT-analyse van de dreigtweetmonitor in het algemeen.

Tabel 7: swot-analyse DTM

Sterkte	*Zwakte*
- een tool met linguïstische filtering ontwikkeld door de taalkundige wetenschappers - sterke reductie van het aantal te beoordelen dreigtweets - makkelijk aanpasbaar voor specifiek gebruik - technische inbedding binnen KNP succesvol	- grote afhankelijkheid samenwerkingspartners - grote afhankelijkheid hardware en software - contextuele interpretatie niet te automatiseren - complexe en arbeidsintensieve voorbereiding vanuit kennis en expertise organisatie
Kansen	*Bedreigingen*
- training en onderhoud linguïstische analyse middels aanleveren beoordeling OSINT - uitbreiding naar monitoring meerdere sociale media - uitbreiding naar meerdere taaldomeinen - complexe voorbereiding verder automatiseren	- dreigingen geuit in andere taaldomeinen - cultuurveranderingen - *false negatives* niet/nauwelijks te controleren - dreigingen die door de 1e en 2e annotatie verkeerd of anders worden geïnterpreteerd

art. 10 lid 2 onder c, d en g

6.2 Aanbevelingen

De DTM wordt eigendom van de nationale politie. De samenwerkingsconstructie binnen dit project is zo opgebouwd dat de betrokken partijen (Tardis en CLST) zelf verantwoordelijk blijven voor het door hen ontwikkelde onderdeel hierin. Het succes van de DTM is op meerdere vlakken gehaald: De nationale politie heeft gezorgd voor het inrichten van het proces, het gebruik van de DTM in de bestaande werkprocessen door het inrichten van een front- en backoffice, en het aanbrengen en structureren van informatie in de vorm van de LOODS-dossiers; Tardis heeft zorg gedragen voor data acquisitie, de technische mogelijkheid om in de DTM te kunnen classificeren en annoteren, en Tardis heeft tevens gezorgd voor de presentatie van de onderzoeksgegevens in een gebruiksvriendelijke interface; en CLST heeft gezorgd voor een categorisering van de Twitterberichten met dreigingscomponenten op basis van een uitgebreide linguïstische en wetenschappelijk onderbouwde dreigingsfilter, waarbij wordt gewerkt op basis van een door hen wetenschappelijk opgebouwde definitie van een dreigtweet.

Gezien het gegeven dat verschillende partijen verschillende gedifferentieerde en onafhankelijke 'onderdelen' aan de dreigtweetmonitor hebben geleverd, en gezien de mogelijk toekomstige wensen en ontwikkelingen binnen een organisatie als de nationale politie, kan van de in dit project ontwikkelde DTM-structuur afgeweken worden. Dat kan betekenen dat slechts een (onder)deel van de bestaande dreigtweetmonitor in de bestaande politiewerkprocessen wordt doorontwikkeld. Welk deel dit betreft hangt volledig af van de wens van de opdrachtgevende partij en zal te allen tijde besproken moeten worden met de projecteindverantwoordelijke van dit project.

Om belangenverstrengeling te voorkomen en onafhankelijkheid binnen een mogelijk vervolgproject te waarborgen, wordt ten zeerste aangeraden om een projecteindverantwoordelijke aan te wijzen die geenszins betrokken is bij één van de door te ontwikkelen onderdelen van de dreigtweetmonitor. Deze eindverantwoordelijke

zal bovendien op IT-vlak technisch onderlegd moeten zijn en voldoende kennis en ervaring inzake contractmanagement moeten hebben, om te zorgen dat er binnen het project blijvend partnering en vertrouwen tussen de samenwerkende partijen kan ontstaan. Verder (wetenschappelijk) onderzoek naar dreigingen via internet en het automatisch filteren van die berichten is altijd gewenst. De vraag is in hoeverre een organisatie als de nationale politie wil investeren in een mogelijk maatschappelijke trend als geuite bedreigingen via Twitter, en alle facetten waar dit mee samenhangt: taal- en cultuurgerelateerde uitingen, andere platformen dan Twitter waarop bedreigingen worden geplaatst, mogelijk geografische pinpointing, het koppelen van gerelateerde informatie vanuit de politiesystemen, het onderzoeken van mogelijk voorspellende waardes en waarschijnlijkheid in een bericht, etc.

ar. 10 lid 2 onder c, d en g

Ook is de discussie over de juridische (on)mogelijkheden rond dit maatschappelijk vraagstuk nog vers en kan op diverse manieren uitpakken voor het wel of niet gebruiken van een monitoringstool binnen het werkproces van de nationale politie. Vanuit het project monitoren 'dreigtweets' zijn er legio aan uitbreidingsontwikkelingen en vervolgonderzoeken mogelijk. Om de DTM in zijn algemeenheid door te kunnen ontwikkelen is het raadzaam om in eerste instantie te kijken naar andere te monitoren open bronnen op het internet. Door de DTM alleen in te richten op Twitter is er voor het welslagen van het monitoren van dreigingen een bijzondere afhankelijkheid van de technische infrastructuur die de bron Twitter zelf biedt.

art. 10 lid 2 onder c, d en g

Naast de relevantie met betrekking tot de automatisering en de technische basis van de dreigtweetmonitor, zal serieus moeten worden omgegaan met werkprocessen die de handmatige beoordelingen ondersteunen: Door het werkproces in te delen in een frontoffice en een backoffice kan het monitoren van dreigtweets op de werkvloer adequaat verlopen. Hierbij mag onverlet opgemerkt worden dat het venijn in de staart zit: het uiteindelijk inschatten of er sprake is van echte (be)dreiging, blijft mensenwerk waarbij expertise onontbeerlijk blijft. Handmatige beoordelingen zoals contextuele interpretatie en logisch nadenken, zal altijd een belangrijk deel moeten blijven in het deels geautomatiseerde werkproces.

art. 11 lid 1
art. 10 lid 2 onder c, d en g