



Koninklijk Nederlands
Meteorologisch Instituut
Ministerie van Infrastructuur en Milieu

Het verbinden van sociale media en het weer

I. van der Giessen

KNMI Intern Rapport IR-2014-05

Het verbinden van sociale media en het weer

Een onderzoek naar de mogelijkheden binnen ArcGIS om real-time Twitter data in kaart te brengen.

Ingeborg van der Giessen

31-01-2014



Inhoudsopgave

	Bladzijde
Inleiding & Probleemstelling	2 - 3
Hoofdstuk 1: Analyse	4 - 6
1.1 <i>Potentieel geschikte software</i>	4 - 5
1.2 <i>Twitter</i>	6
Hoofdstuk 2: Functionaliteiten	7 - 19
2.1 <i>GeoEvent Processor binnen ArcGIS Framework</i>	7 - 8
2.2 <i>Werking GeoEvent Processor</i>	8 - 10
2.3 <i>Use Cases</i>	11 - 12
2.4 <i>Koppeling Twitter functionaliteiten</i>	13 - 14
2.5 <i>Visualisatie functionaliteiten (output)</i>	14 - 18
2.6 <i>Overige functionaliteiten</i>	19
Hoofdstuk 3: Discussie	20 - 23
3.1 <i>Software</i>	20
3.2 <i>Tekortkomingen</i>	20 - 21
3.3 <i>Bruikbaarheid GeoEvent Processor voor KNMI use cases</i>	22 - 23
Conclusie & Aanbevelingen	24 - 25
Bronnen	26

Inleiding en Probleemstelling

Berichten van sociale media kunnen bijdragen aan de verslaglegging van het weer, klimaat of seismologie. Berichtgeving over de gevolgen van slecht weer, kunnen meteorologen inzicht bieden in de impact van het weer op de samenleving. Wanneer er een grote storm over Nederland raast zou het KNMI graag inzicht krijgen in de impact van dit weer op de samenleving, bijvoorbeeld informatie over omgevallen bomen, weggewaaide daken en ontoegankelijke wegen. Op sociale media wordt op dagen met bijvoorbeeld weeralarmen veel gecommuniceerd over het weer, maar ook over de gevolgen ervan. Er worden bijvoorbeeld foto's op Twitter geplaatst van kapotte daken of omgevallen bomen.

Data afkomstig van sociale media is door de gebruikers niet perse bedoeld als geografische data, maar heeft uiteindelijk wel deze functie (Stefanidis e.a., 2013). De hoeveelheid data is enorm groot en het is een uitdaging om deze in kaart te brengen, zodat de meteorologen op het KNMI de informatie in één beeld overzichtelijk hebben. Uitdagingen liggen in het weergeven van *real-time* data, het selecteren van de geschikte data en de visualisatie ervan. In dit onderzoek wordt gekeken naar de mogelijkheden voor het KNMI om aan de hand van ArcGIS sociale media data in kaart te brengen. ArcGIS is een platform voor het ontwikkelen en managen van oplossingen voor geografische problemen. De software wordt uitgegeven door Esri en kent een breed scala aan mogelijkheden en toepassingen (Esri, 2013). De keuze voor het ArcGIS Framework ligt voor de hand, omdat het KNMI de software al gebruikt voor andere toepassingen bij bijvoorbeeld klimaat en seismologie.

Het in kaart brengen van sociale media is een 'hot topic' en er verschijnen op internet dan ook steeds meer websites waarop tweets, Flickr-foto's of andere sociale media geografisch worden weergegeven. Een mooi voorbeeld is *The One Million Tweet Map* (2013), hierop worden de laatste miljoen tweets in *real-time* geografisch weergegeven. ArcGIS Online heeft zich de laatste jaren ook gericht op sociale media en biedt allerlei *templates*, *widgets* en toepassingen voor sociale media (Baranyi, 2011). Er wordt daarnaast steeds meer software door Esri op de markt gebracht om *real-time* data in kaart te brengen en sociale media als data te gebruiken. Zo heeft Esri in juli 2013 de GeoEvent Processor extensie voor ArcGIS for Server uitgebracht en biedt daarmee mogelijkheden voor dit onderzoek (Esri, 2013).

In het kader van de Nationale GI Minor aan de Vrije Universiteit van Amsterdam heeft dit onderzoek plaatsgevonden tussen 28 oktober 2013 en 31 januari 2014. Dit onderzoek is uitgevoerd aan de hand van de volgende probleemstelling:

'Wat zijn de mogelijkheden voor het KNMI om binnen het ArcGIS framework Twitter data betreffende het weer in kaart te brengen?'

De probleemstelling zal worden onderzocht aan de hand van een aantal deelvragen:

- Welke *real-time data streaming* functionaliteiten bieden deze mogelijkheden?
- Wat voor opties bieden de mogelijkheden wat betreft archivering van sociale media data?
- Welke mogelijkheden biedt de software voor visualisatie van gegevens ten behoeve van operationele toepassing in de weerkamer?
- Hoe makkelijk is de gepresenteerde informatie aan te passen aan de wensen van het KNMI?

Dit onderzoek zal zich richten op het medium Twitter, dit vanwege het grote aantal gebruikers en de redelijk grote toegankelijkheid tot de data. Daarnaast bieden ArcGIS-producten veel invoegtoepassingen voor Twitter. Door Rijkswaterstaat is al eerder onderzoek gedaan naar sociale media en hieruit kwam naar voren dat Twitter een goede optie is voor de voorziening van geografische data tijdens incidenten en evenementen (Rijkswaterstaat, 2012). In later vervolgonderzoek door het KNMI is ingegaan op de hevige sneeuwval van 3 februari 2012 en een analyse uitgevoerd op Twitterdata (M. Molendijk, persoonlijke communicatie, 8 november 2013). Naast deze interne onderzoeken van het Ministerie van Infrastructuur en Milieu, is er in de afgelopen jaren ook door verscheidende wetenschappers onderzoek gedaan naar Twitter als een databron en de eerste bevindingen zijn positief (Cheng e.a., 2010; Kwak e.a., 2010; Stefanidis e.a., 2013).

In het eerste hoofdstuk Analyse zal in worden gegaan op de verschillende mogelijkheden om *real-time* data in kaart te brengen. Daarnaast zal er ook dieper worden ingegaan op Twitter en de mogelijkheden en beperkingen hiervan. In het tweede hoofdstuk zal de GeoEvent Processor verder worden onderzocht en de functionaliteiten hiervan worden getest. Dit wordt gedaan aan de hand van twee *use cases*: 'Monitoring & Bewaking' en 'Analyse'. In het volgende hoofdstuk, de discussie, wordt de geschiktheid van de software ter discussie gesteld. Is de software geschikt voor het KNMI en de doeleinden? In de conclusie zal antwoord worden gegeven op de hoofdvraag en aanbevelingen worden gedaan.

Hoofdstuk 1: Analyse

In dit hoofdstuk worden de verschillende software mogelijkheden gepresenteerd en onderzocht op hun effectiviteit voor dit onderzoek. Aan de volgende criteria is de software getoetst: de beschikbaarheid en actualiteit van de software; functionaliteiten software met betrekking tot *real-time* in kaart brengen van data; toegankelijkheid software. Daarnaast zal er worden gekeken naar Twitter als voornaamste databron en de mogelijkheden en beperkingen hiervan.

1.1 Potentieel geschikte software

Het in kaart brengen van sociale media data is een zeer actueel onderwerp en er verschijnen binnen het ArcGIS framework dan ook steeds meer mogelijkheden en functionaliteiten. Uit een eerste inventarisatie van het KNMI is gebleken dat Esri verschillende potentieel geschikte software en mogelijkheden uit heeft gebracht de laatste jaren: ArcGIS Explorer, ArcGIS for Desktop (Tracking Server plug-in), ArcGIS Online en de ArcGIS GeoEvent Processor. Meer duidelijkheid over de structuur van de Esri software wordt gegeven in het volgende hoofdstuk: functionaliteiten. Onderzoek heeft uitgewezen dat echter niet alle mogelijkheden meer even functioneel zijn of worden ondersteund door Esri.

ArcGIS Explorer

De ArcGIS Explorer Desktop is een gratis GIS *viewer* waarmee GIS informatie makkelijk kan worden geanalyseerd, gevisualiseerd en gedeeld. ArcGIS Explorer werkt samen met ArcGIS Online en is makkelijk te gebruiken door iemand met weinig expertise. De ArcGIS Explorer is dus geen software speciaal gemaakt voor het analyseren en verwerken van *real-time* (sociale media) data, maar het biedt invoegtoepassingen of tools hiervoor. Zo is er een *Social Media Widget* en een *Tweet Mapping Template* (Baranyi, 2011). Deze tools zijn echter inmiddels redelijk achterhaald en bieden weinig aanpasmogelijkheden voor de gebruiker. ArcGIS Explorer biedt daarom ook niet genoeg functionaliteiten voor het KNMI.

ArcGIS for Desktop (Tracking Server plug-in)

Een plug-in voor ArcGIS for Desktop, de Tracking Server biedt op het eerste gezicht mogelijkheden voor het in kaart brengen van *real-time* data. Deze server verzamelt en distribueert *real-time* data naar het Web of naar de Desktop. Nader onderzoek wijst echter uit dat Tracking Server gedateerd is en niet meer wordt ondersteund door Esri. De interface van de Tracking Server werkt niet meer naar behoren, het is te ingewikkeld om met de *plug-in* bijvoorbeeld een onderzoek naar sociale media data uit te voeren. De Geo-Event Processor vervangt met het uitbrengen van ArcGIS Desktop 10.2 de Tracking Server.

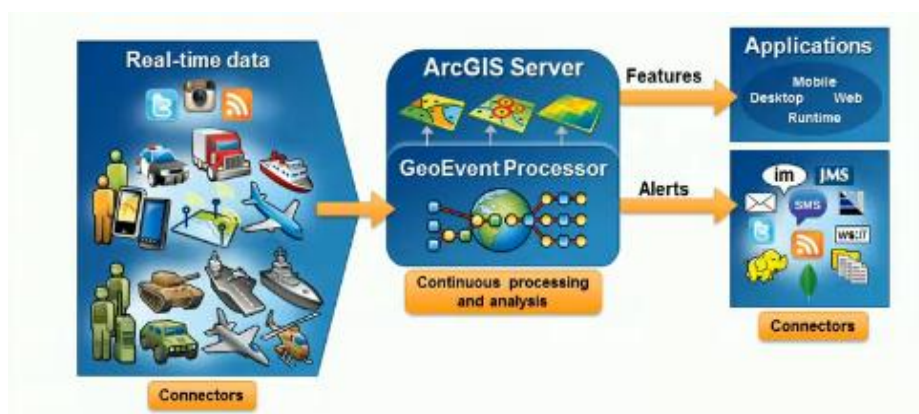
ArcGIS Online

ArcGIS Online biedt ArcGIS gebruikers een variëteit aan *web services* die zijn geïntegreerd in alle ArcGIS producten van Esri. ArcGIS Online bestaat uit kaart services, software ontwikkelingshulpmiddelen en het delen van kaarten en andere contents via het web. ArcGIS Online's grootste speerpunt is de compatibiliteit met andere software en de integratie hiervan. ArcGIS Online biedt uit zichzelf weinig mogelijkheden, er is wel een geschikte *template* aanwezig. Deze *template* geeft de mogelijkheid om tweets in kaarten te brengen aan de hand van een zoekfunctie. Deze *template* is vergelijkbaar met de mogelijkheden die ArcGIS Explorer biedt. ArcGIS

Online is op zichzelf staand dus te beperkt, maar kan wel goed worden gebruikt in combinatie met andere software. Zo kan ArcGIS Online worden gebruikt voor visualisatie van output van de GeoEvent Processor.

ArcGIS GeoEvent Processor

In juli 2013 heeft Esri haar nieuwste software op het gebied van *real-time GIS* gelanceerd als een extensie van ArcGIS for Server. De GeoEvent Processor (zie figuur 1.1) maakt het mogelijk om bijna elke vorm van *streaming data* te verbinden met een output, zoals *alerts* naar personeel, tweets of mails. Zo is het mogelijk om meldingen te krijgen wanneer bepaalde condities zich voordoen. De GeoEvent Processor bevat *connectors* voor veel sensoren, zoals GPS systemen en sociale media providers en andere *connectors* zouden naar wens kunnen worden toegevoegd. Het is mogelijk om de *real-time* data te filteren om zo alleen de belangrijkste informatie te verkrijgen voor een project. Op het eerste gezicht biedt de GeoEvent Processor veel mogelijkheden voor het KNMI en voor dit onderzoek. Sociale media *real-time* data wordt ondersteund, ook specifiek voor Twitter. Daarnaast zijn er mogelijkheden voor constante visualisatie voor gegevens en lijken de aanpas mogelijkheden groot.



Figuur 1.1: GeoEvent Processor
Bron: Training Esri (2013)

Er kan worden gesteld dat de met de *release* van de GeoEvent Processor in juli 2013 andere mogelijkheden (ArcGIS Explorer, ArcGIS Tracking Server) niet meer relevant lijken voor dit onderzoek. Esri ondersteunt de andere software en mogelijkheden niet meer en richt zijn pijlen op de GeoEvent Processor voor het in kaart brengen van *real-time* data. De GeoEvent Processor biedt volgens Esri de vooraanstaande technologie op het gebied van *real-time GIS* (Esri, 2013).

Het is niet zinvol om dit onderzoek te richten op gedateerde mogelijkheden en die zullen dan ook niet verder onderzocht worden in dit onderzoek, het onderzoek zal zich voornamelijk richten op de GeoEvent Processor en de (bijkomende) toepassingen binnen het ArcGIS Framework. De GeoEvent Processor werkt namelijk nauw samen met ArcGIS Online, ook toepassingen met ArcGIS for Desktop zijn denkbaar.

1.2 Twitter

Naast een analyse van de geschikte software is ook een analyse van Twitter op zijn plaats, aangezien Twitter dient als de databron voor dit onderzoek. Twitter is een *microblogging medium*, opgericht in 2006 en uitgegroeid tot een van de grootste sociale media sites ter wereld (Stefanidis e.a., 2013). In januari 2013 werd de website zelfs uitgeroepen tot de 10^e populairste site ter wereld, net na Facebook en LinkedIn (Alexa, 2013). Het medium kent 500 miljoen geregistreerde gebruikers die gezamenlijk 400 miljoen tweets per dag posten. Tweets zijn korte, 140 karakters lange berichten, die gebruikers versturen naar hun volgers (Morstatter e.a., 2013). Twitter stelt daarnaast de gebruiker in staat om de tweets te *georeferencen*. Dit houdt in dat er een geografische locatie aan de tweet wordt toegevoegd door middel van latitude en longitude. Op deze manier is Twitter een enorm grote bron van geografische data en daarom zeer interessant voor sociale wetenschappers (Cheng e.a., 2010).

Waarom dan de grote interesse voor Twitter en niet voor andere sociale media waar ook gebruik gemaakt wordt van een geografische locatie zoals Facebook? Sociale media data is over het algemeen zeer lastig te verkrijgen, de meeste sites hebben een restrictie op hun data. Twitter's beleid staat echter loodrecht op dat van de meeste andere sociale media sites. De *Twitter Streaming API* stelt iedereen in staat om 1% van alle Twitter data in *real-time* te bemachtigen. Aan de hand van verschillende parameters (zoektermen, geografische grenzen of gebruikers) kan deze data worden binnengehaald door de gebruiker. Naast deze beperkte 1% *Streaming API* is er een *Firehose* die 100% van alle tweets kan bieden aan gebruikers. Deze *Firehose* heeft echter nogal wat restricties en de kosten zijn enorm hoog (Morstatter e.a., 2013). Waar Twitter de *Streaming API* aanbiedt aan het grote publiek is het juist beperkt met het aanbieden van de *Firehose*. Er is maar een aantal bedrijven die toegang hebben tot deze enorme bulk aan data (API Voice, 2012). Niet alleen de beschikbaarheid is een reden voor het gebruik van Twitter voor dit onderzoek, ook de omvang van de data speelt een belangrijke rol. Het medium Flickr geeft bijvoorbeeld ook toegang tot haar data, maar dit is een veel kleiner sociale media platform en biedt dan ook minder data. Twitter is een zeer groot en populair medium, waardoor meer data beschikbaar is voor onderzoek. Ook uit onderzoek van Rijkswaterstaat is gebleken dat Twitter (met name *georeferenced* tweets) geschikt is voor het onderzoeken van events, omdat trends en de algemene stemming tijdens zo'n event goed kunnen worden onderscheiden (Rijkswaterstaat, 2012).



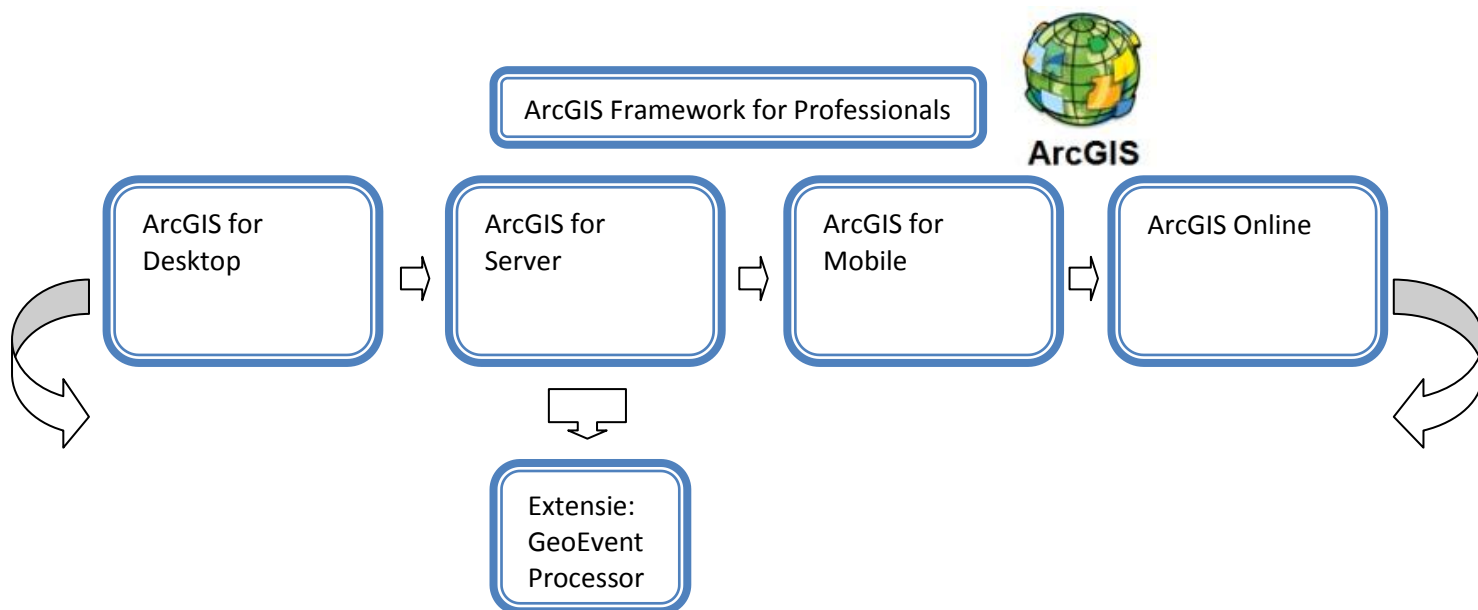
Voor dit onderzoek wordt er gebruik gemaakt van de *Streaming API* en hoewel het maar om 1% van alle tweets gaat, wijst onderzoek uit dat er vanuit kan worden gegaan dat dit voldoende is voor een onderzoek naar trends. De data kan worden *geharvest* en er kan een kleine analyse op worden uitgevoerd. Bij een onderzoek en een analyse in de breedte biedt deze *Streaming API* niet genoeg data en is deze dan ook ongeschikt. De betrouwbaarheid hangt af van het gebruik en de type analyse die wordt uitgevoerd en in het geval van een trend onderzoek pakt dit goed uit. Juist de *georeferenced* tweets zijn betrouwbaar en uit onderzoek van Morstatter e.a. (2013) blijkt dat Twitter hierbij meer dan 1% van haar tweets aanbiedt wanneer deze parameter wordt ingezet. Hierbij moet wel de kanttekening worden geplaatst dat maar een beperkt deel (ongeveer 1%) van alle tweets *georeferenced* is (Morstatter e.a., 2013). Een groot deel van alle tweets (99,99%) is dus niet beschikbaar voor het onderzoek, omdat hierin de *georeferenced* tweets van belang zijn. Door het enorme volume aan tweets blijft er echter nog steeds een substantieel aantal tweets over. Het is dus daarom belangrijk om een groot medium als Twitter te gebruiken, om voldoende data te kunnen *harvesten* voor onderzoek.

Hoofdstuk 2: Functionaliteiten

Zoals in het vorige hoofdstuk is toegelicht, zal dit onderzoek zich richten op de functionaliteiten en mogelijkheden van de GeoEvent Processor en het ArcGIS Framework. In dit hoofdstuk zullen de functionaliteiten beoordeeld worden en zal er een inventarisatie van de mogelijkheden voor het KNMI plaatsvinden. Dit wordt gedaan aan de hand van twee *use cases*: Monitoring & Bewaking en Analyse. Eerst zal er een uitleg worden gegeven over de werking van de GeoEvent Processor en de plek die deze inneemt binnen het ArcGIS framework. Daarna zal worden ingegaan op de functionaliteiten aan de hand van de *use cases*.

2.1 GeoEvent Processor binnen ArcGIS Framework

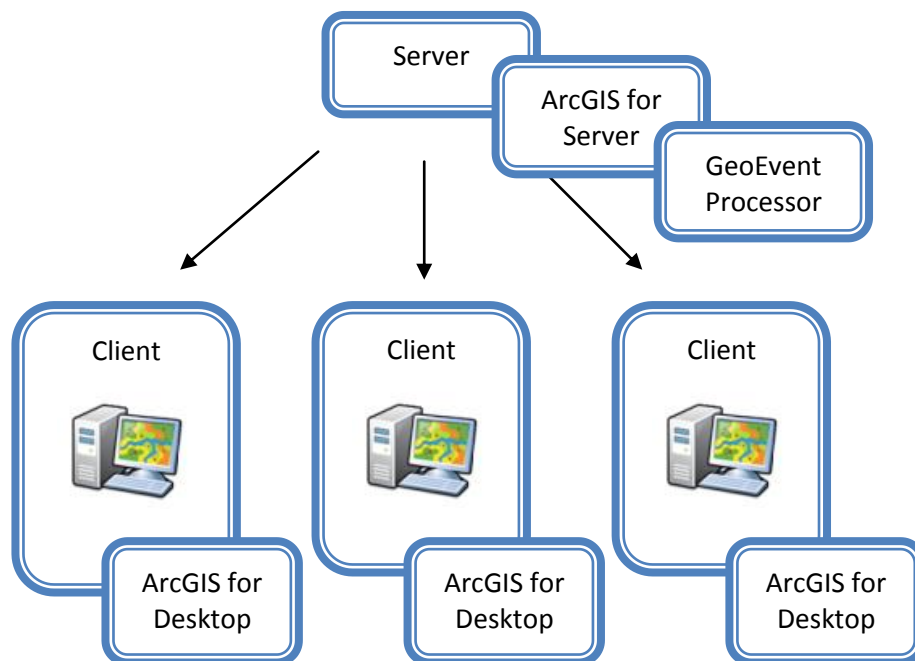
Esri ontwikkelt geografische informatie systemen (GIS) die functioneren als een integraal deel van vele organisaties. De GIS software die Esri heeft ontwikkeld, is erg uitgebreid en biedt enorm veel verschillende en uiteenlopende functionaliteiten. Esri gebruikt de naam ArcGIS voor alle GIS software producten, die functioneren op desktop, server, online en op mobiele platformen. Deze platformen zijn geïntegreerd en er is koppeling tussen alle onderdelen van het *framework* (zie figuur 2.1). De twee belangrijkste componenten van ArcGIS voor dit onderzoek zijn ArcGIS for Desktop en ArcGIS for Server, dit zijn 'Professionele GIS Systemen' (Esri, 2013). Er zijn andere toepassingen van Esri die buiten dit professionele systeem vallen, zoals bijvoorbeeld de ArcGIS Explorer (zie hoofdstuk Analyse).



Figuur 2.1: ArcGIS Framework

De GeoEvent Processor is een extensie van ArcGIS for Server, een server management tool voor het creëren en managen voor GIS Web Services, applicaties en data. ArcGIS for Server wordt met name gebruikt binnen bedrijven en vereist een *service-oriented architecture* binnen dit bedrijf. ArcGIS for Server wordt geïnstalleerd op een server en idealiter kunnen werknemers vanaf clients gebruik maken van de mogelijkheden van de server, zonder zelf kennis hierover te hebben. De GeoEvent Processor wordt ook geïnstalleerd op de server en kan worden ingericht op deze server (zie figuur

2.2). De gebruiker, de client, heeft zelf redelijk wat mogelijkheden met de GeoEvent Processor, maar is afhankelijk van de instellingen op de server.



Figuur 2.2: Software rondom GeoEvent Processor

Voor verdere visualisatie kan het nodig zijn om ArcGIS for Desktop te gebruiken, deze software kan worden geïnstalleerd op een cliënt en biedt mogelijkheden aan de gebruiker voor data modellering, analyse en visualisatie. ArcGIS for Desktop wordt geïnstalleerd op een desktop PC en is de meest gebruikte toepassing binnen het ArcGIS Platform.

Voor al deze 3 producten (ArcGIS for Server, ArcGIS for Desktop en de GeoEvent Processor) zijn licenties vereist van Esri. Uit dit onderzoek bleek dat het niet eenvoudig is om te werken met de GeoEvent Processor en dat verdere ArcGIS en ICT kennis nodig is voor het functioneren van de software. Vooral het inrichten van de server en het aanpassen naar de gewenste instellingen voor het KNMI bleken een uitdaging.

2.2 Werking GeoEvent Processor

De GeoEvent Processor werkt met veel verschillende tools en systemen, in dit hoofdstuk wordt getracht een eenvoudige uitleg te geven over de basisprincipes van de software. In de basis werkt de GeoEvent Processor met input, output en services.

Input

Inputs laden data in de GeoEvent Processor, elke input is zo geconfigureerd om data te ontvangen van een specifieke databron. Inputs kunnen *real-time* zijn en constant input leveren, een voorbeeld is het laden van tweets. Met behulp van verschillende parameters kan de input gefilterd worden op gebruiker, woorden of locatie. De parameter *track* is voornamelijk van belang, hiermee kan gezocht worden op woorden in een tweet. Zo levert bijvoorbeeld de parameter 'KNMI' alle tweets op waarin de lettercombinatie KNMI (al dan niet met hoofdletters) zit. Dit kunnen dan bijvoorbeeld ook tweets

zijn waarbij de combinatie voorkomt in een andere taal of gebruikersnaam. Zoals bijvoorbeeld bij de volgende tweet uit Indonesië:

*Bknmi bisaa jdi itumi mmg @Andi_Fitr4: Haaha bisa jadi, bisa jadi .. :p
@mayamayo2: @NaryaSR anaknya tante ija toh @Andi_Fitr4
from @NaryaSR at 11/29/2013 14:06, Indonesia*

De parameter locatie zou dit in theorie kunnen voorkomen, zodat alleen tweets binnen Nederland binnenkomen als input. De locatie parameter werkt echter niet in de processor (*bug*) en daardoor kan er niet op locatie worden gefilterd. Dit is een duidelijk gebrek aan de GeoEvent Processor, maar wellicht zal dit probleem worden opgelost in volgende versies van de software.

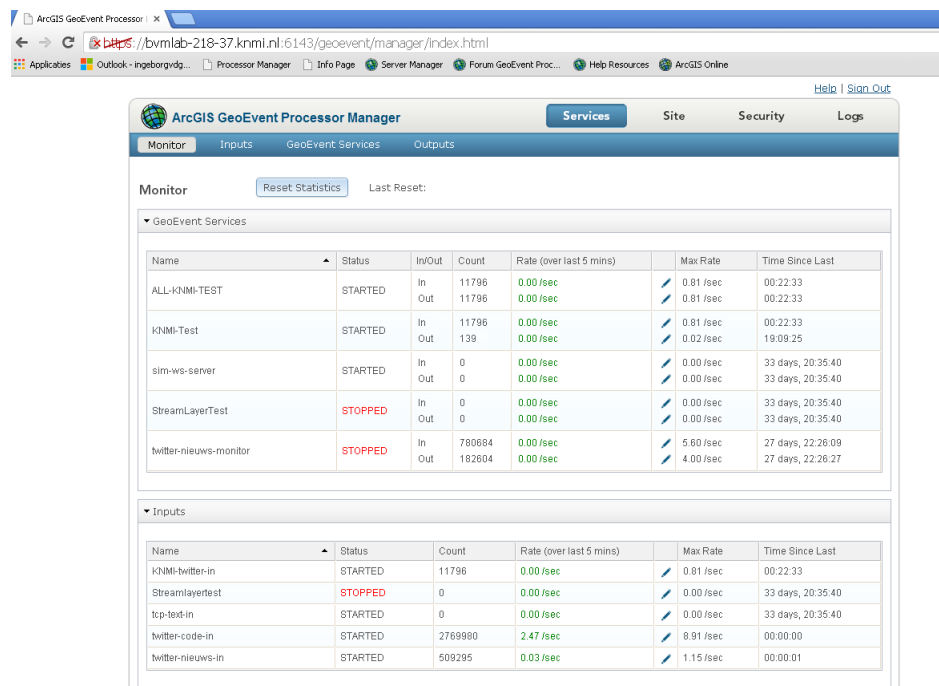
Output

De output zorgt ervoor dat de gestreamde data vanuit de GeoEvent Processor naar een specifieke bestemming wordt gezonden. Dit kan bijvoorbeeld een *Comma Separated Values* (CSV) bestand zijn, een *TCP-socket* (hier een netwerk verbinding tussen het Internet en het data verwerkende bestand) of een *feature layer* in een *geodatabase*. Alle tweets die zojuist ontvangen zijn in een input met het woord KNMI erin, kunnen bijvoorbeeld opgeslagen worden in een CSV bestand. Ook kunnen ze worden opgeslagen op de server, om daarna direct geopend te worden in ArcGIS for Desktop of in ArcGIS Online.

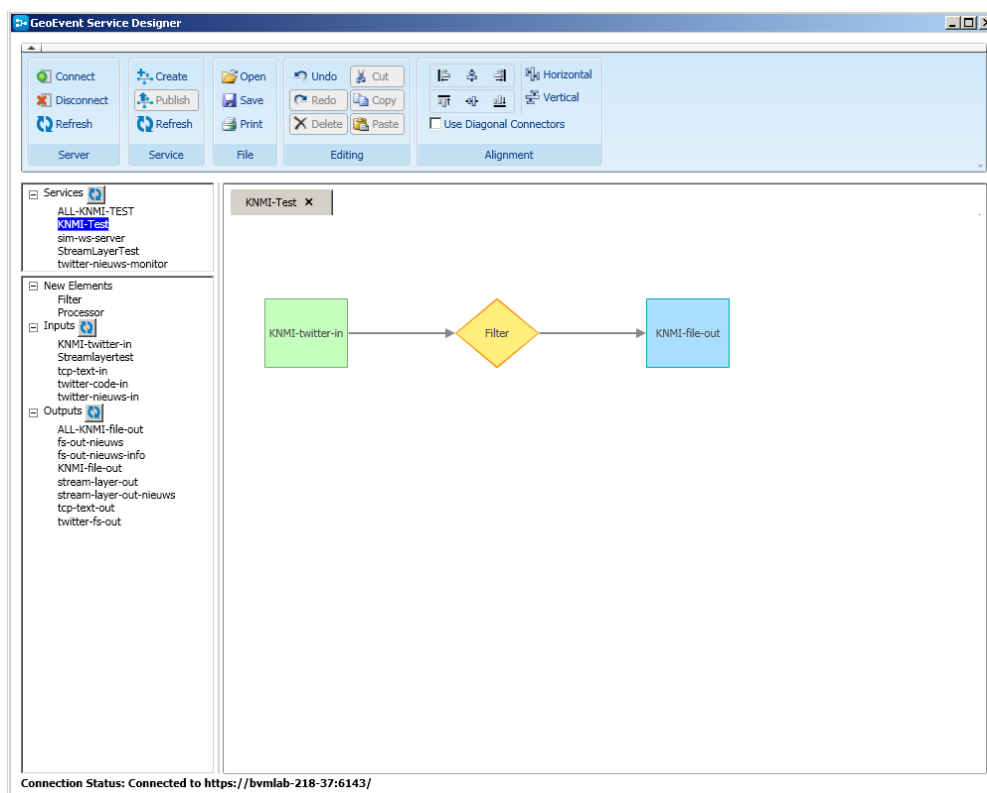
Services

Om de input en de output met elkaar te verbinden moet er een service worden gemaakt. Een GeoEvent service verbindt een of meerdere inputs met een of meerdere outputs. Tussen die verbinding kan er ook processing plaatsvinden. Een voorbeeld van processing (filtering) relevant voor dit onderzoek is het instellen van een *georeferenced* filter. Dit filter zorgt ervoor dat alleen de tweets met een *geolocation* (latitude/longitude) worden doorgestuurd naar de input. Zo komen bijvoorbeeld alle tweets met het woord KNMI binnen, de GeoEvent Service filtert dan alleen de *georeferenced* tweets eruit en in de output (een CSV bestand) verschijnen dan alleen deze tweets (zie figuur XXX).

De inputs en outputs worden gecreëerd in de GeoEvent **Processor Manager** (zie figuur 2.3), dit is een webpagina, toegankelijk via de server, waar de inputs, outputs en andere variabelen kunnen worden gecreëerd en beheerd. Deze inputs en outputs kunnen met elkaar verbonden worden en op deze manier services creëren in de GeoEvent **Service Designer** (zie figuur 2.4). Dit is een losse software tool die geïnstalleerd staat op de client. Er zijn dus verschillende tools (een software tool en een webpagina) waarmee gewerkt moet worden om een service te creëren, dit maakt de interface nogal omslachtig en onoverzichtelijk.



Figuur 2.3: GeoEvent Processor Manager. Een webpagina waar inputs en outputs aangemaakt kunnen worden en alle processen gemonitord kunnen worden.



Figuur 2.4: GeoEvent Service Manager. Hier kunnen de inputs en outputs aan elkaar verbonden worden tot services.

2.3 Use Cases

Om de mogelijkheden van de GeoEvent Processor binnen het ArcGIS Framework uit te testen, wordt er gebruik gemaakt van twee *use cases*, waarbij verschillende functionaliteiten worden getest. Hierbij wordt met name gekeken naar de bruikbaarheid van de GeoEvent Processor, maar ook van een aantal softwaremogelijkheden die hier verder aan gekoppeld kunnen worden, bijvoorbeeld ArcGIS for Desktop. De GeoEvent Processor maakt wellicht sommige toepassingen mogelijk, maar maakt deel uit van het framework waardoor deze in samenhang met andere software uit het framework moet worden beschouwd.

Use Case 1: Monitoring en Bewaking

Bij deze case worden de functionaliteiten getoetst voor het monitoren en bewaken van Twitter data over het weer. Het KNMI streeft ernaar om de waarschuwingen voor extreem weer aan te laten sluiten aan de gevolgen van dit extreme weer. Het is daarom goed inzicht te krijgen in de gevolgen van het weer, dit inzicht kan worden verbeterd door de Twitter data inzichtelijk te krijgen in een *real-time monitor*. Dit is vooral van toepassing op dagen met extreem weer (uitgifte code oranje of code rood). Deze case is met name relevant voor de meteorologen in de weerkamer en de informatie moet in *real-time* beschikbaar zijn. Op deze manier kunnen de meteorologen bijhouden wat er gebeurt, waar dit gebeurt en wanneer.

Use Case 2: Analyse

De tweede *use case* is een toepassing die aansluit bij onderzoekers, die om diverse redenen Twitter data willen analyseren. Zo kan het KNMI meer fundamentele kennis vergaren over sociale media en het weer en bijvoorbeeld feedback inwinnen van de gebruikers over de communicatie van het KNMI tijdens een dag met extreem weer. De redenen voor analyse met behulp van het ArcGIS Framework zijn met name gerelateerd aan de beschikbaarheid van de *georeferenced* tweets. Wat zijn de mogelijkheden voor het KNMI en wat kan het KNMI hiermee analyseren aan de hand van dit framework? Kunnen de *events* worden geanalyseerd met behulp van deze data, door bijvoorbeeld het combineren met andere (meteorologische) geografische data?

Vergelijking functionaliteit use cases

Uit de *use cases* blijkt dat in beide gevallen andere functionaliteiten gewenst zijn. Bij de eerste case is er vooral behoefte aan *real-time data streaming* en ook *real-time* visualisatie van deze informatie. Bij de tweede case zou er gebruik gemaakt kunnen worden van de *real-time streaming* optie om deze informatie op te slaan, om hier later verdere analyse op uit te voeren. Bij de tweede case zou er ook gebruik gemaakt kunnen worden van een set tweets die gekocht zijn bij een *databroker*, zoals bijvoorbeeld GNIP, de grootste databron voor sociale media data (GNIP, 2014). Dit is gedaan bij een eerder onderzoek van het KNMI naar onderzoek naar Twitter rondom de sneeuwval van 3 februari 2012. Op basis van de *use cases* is een lijst samengesteld met gewenste functionaliteiten (zie tabel 2.1); deze worden in de volgende paragrafen (2.4 – 2.6) beschreven.

Functionaliteiten

2.4 Koppeling Twitter

Volledigheid Dataset

Filtering o.b.v. metadata

Archivering tweets

2.5 Visualisatie

Presentatie tweets op Kaart

Automatische Update

Selectie Tijd

Selecteren en presenteren foto's

Zichtbaar maken data tweet

Presentatie 'Symbology'

Tweets verwijderen uit presentatie-set

In/uitzoomen kaart

Clustering

Presentatie metadata

Combineren verschillende kaartlagen

Combineren radarbeelden

Afdrukken kaart met tweets

2.6 Overige functionaliteiten

Dedicated Software op Client Nodig?

Handmatige verwerking nodig & drempel gebruiker

Kosten

Tabel 2.1: Overzicht functionaliteiten Use Cases. In de tabel zijn de gewenste functionaliteiten weergegeven voor beide use cases. Deze functionaliteiten zijn onderverdeeld in 3 categorieën, die overeenkomen met de paragrafen.

2.4 Koppeling Twitter Functionaliteiten (input)

Allereerst zal er gekeken worden naar de functionaliteiten wat betreft de input: de koppeling met Twitter. Zoals beschreven in paragraaf 2.2 werkt de GeoEvent Processor met inputs, outputs en services. Er zal nu worden ingegaan op de input en dan met name de input van Twitter.

Volledigheid dataset

De GeoEvent Processor werkt met de Twitter API, waarbij 1% van alle tweets wordt vrijgegeven en ontvangen als input in de GeoEvent Processor. Deze 1% is bruikbaar wanneer er gewerkt wordt met de eerste *use case*: Monitoring & Bewaking, maar niet optimaal. Immers, om een betrouwbaar beeld te krijgen van (*georeferenced*) Twitter berichten, zijn zoveel mogelijk relevante tweets gewenst. Een klein deel van alle tweets wordt weergegeven, maar elke tweet is er een en mogelijk relevant voor de meteorologen op de weerkamer. De onvolledigheid van de Twitter dataset is in sommige onderzoek onvoldoende bij de tweede *use case*: Analyse. Bijvoorbeeld bij een analyse over het bereik van de tweets van het KNMI. De wens bij de *use case* Analyse is een zo volledig mogelijke dataset, dit is echter niet mogelijk met de GeoEvent Processor (en achterliggende Twitter *data policy*).

Filtering o.b.v. metadata

Filtering in de GeoEvent Processor is mogelijk, zoals al geconstateerd eerder in dit hoofdstuk. De filtering vindt voor een deel plaats in de input (ontvangen tweets), dit is bijvoorbeeld filtering op basis van *keywords* (*track*). Daarnaast kan er processing plaatsvinden in de service, waarbij alleen de *georeferenced* tweets eruit gefilterd worden. Voor zowel de eerste als de tweede *use case* kan filtering met behulp van de GeoEvent Processor nuttig zijn. De mogelijkheden tot filtering bij de GeoEvent Processor zijn echter beperkt, niet alle gewenste filter functies zijn beschikbaar. Voor beide *use cases* zou het nuttig kunnen zijn om te filteren op bijvoorbeeld *hashtags*, *keywords*, geografische locatie, taal en profiel. Met de GeoEvent Processor kan echter alleen gefilterd worden op *keywords* (via input) en *georeferenced* tweets (met een filter in de service). Er kan wel gefilterd worden op een specifieke gebruiker (d.m.v. input), maar er kan niet worden gefilterd op basis van eigenschappen van profielen. Het filteren op *hashtags* lijkt daarnaast niet mogelijk, er worden in ieder geval geen specifieke tools voor aangeboden. Hetzelfde geldt voor filteren op taal, dit lijkt met de GeoEvent Processor niet mogelijk. Er kan wel worden gefilterd op *georeferenced* tweets, maar er kan niet gefilterd worden op alleen *georeferenced* tweets binnen bijvoorbeeld Nederland.

Bij de tweede *use case* kan er ook gebruik gemaakt worden van een dataset afkomstig van een *databroker*. In dat geval kunnen er *queries* worden uitgevoerd op de database waarin de tweets worden bewaard. Hierbij wordt er geen gebruik gemaakt van de GeoEvent Processor of andere software binnen het ArcGIS Framework.

Archivering tweets

Het archiveren van tweets kan vooral nuttig zijn voor analyse na afloop van een event (bijvoorbeeld na uitgifte van een weeralarm), de tweede *use case*. Archiveren is mogelijk met de GeoEvent Processor, op verschillende manieren. Tweets kunnen worden opgeslagen in een CSV bestand, of ze kunnen direct in een database op de server worden geplaatst. Bij deze laatste optie kunnen de *georeferenced* tweets gelijk worden geopend in ArcGIS for Desktop of met ArcGIS Online. Beide opties werken echter nog niet helemaal soepel. Bij het opslaan in een CSV bestand worden de gegevens niet heel netjes en gestructureerd opgeslagen, er worden bijvoorbeeld geen kolomnamen weergegeven en regelinden in tweets worden niet goed verwerkt. Daarnaast zijn er veel lege

kolommen en velden, soms worden de tweets incompleet opgeslagen. Het is daarom niet mogelijk om direct met het bestand verder te werken, eerst moet er uitgebreide handmatige opschoning plaatsvinden. Ook aan het opslaan in de *geodatabase* zitten haken en ogen, zo worden bijvoorbeeld niet alle velden hier opgeslagen en is er geen duidelijke tabel met alle attributen. De koppeling met ArcGIS for Desktop verloopt daardoor nog niet vanzelfsprekend, er is nog steeds een aantal stappen nodig om van de GeoEvent Processor de koppeling te kunnen maken met ArcGIS for Desktop.

Er is echter nog een aandachtspunt bij het archiveren van tweets: Twitter heeft beperkingen gesteld op het opslaan van tweets. Het is onduidelijk wat nu precies wel en niet mag met de GeoEvent Processor met betrekking tot Twitter. De *Rules of the Road* van Twitter zijn niet heel helder en overzichtelijk en hieruit zijn geen duidelijke regels af te leiden. Esri heeft met haar eerste *tutorial* voor Twitter en de GeoEvent Processor opties en instructies gegeven voor het opslaan van tweets, maar deze zijn in de tweede *tutorial* verwijderd. Het is dus onduidelijk of de opties gegeven in de eerste *tutorial* wel overeenkomen met de *Rules of the Road* van Twitter en of deze opties nog wel gebruikt mogen worden.

2.5 Visualisatie Functionaliteiten (output)

De tweets die (in *real-time*) worden binnen gehaald met de GeoEvent Processor kunnen op verschillende manieren worden gevisualiseerd: *real-time* visualisatie met de GeoEvent Processor Stream Layer en visualisatie achteraf met ArcGIS for Desktop. Deze eerste optie is vooral geschikt voor de eerste *use case*, de tweede voor analyse. In deze paragraaf worden de visualisatie opties besproken en wordt daarna ingegaan op de functionaliteiten van de verschillende visualisatie opties voor de twee *use cases*.

1. Stream Layer

Voor de eerste *use case* is het van belang om de tweets die in *real-time* binnen komen, ook in *real-time* te kunnen visualiseren. De GeoEvent Processor biedt hiervoor zelf geen tools, alle visualisatie wordt gedaan aan de hand van 'externe' software. Dit is software die Esri aanbiedt buiten de GeoEvent Processor om. De meest geschikte tool hiervoor is de Stream Layer, aangeboden door Esri in een *tutorial data pack*. Deze Stream Layer is een kaart op het Internet, die kan worden weergegeven in de *browsers* Chrome en Firefox. Er wordt gebruik gemaakt van een *websocket*, dit is een communicatie verbinding tussen (in dit geval) *web browsers* en *web servers*. Een *websocket* maakt het mogelijk om *live content* te faciliteren. Het wordt bijvoorbeeld ook gebruikt bij *real-time* gaming. Het HTML-bestand configureert een *web map* met de controle om te verbinden met een *websocket* en het ontvangen van de tweets. Deze tweets worden vervolgens weergegeven op de *web map*. De standaard uitvoering van deze Stream Layer is simpel (zie figuur 2.5), hierbij worden alleen stippen weergegeven van de locatie van de tweets.

Het HTML-bestand kan worden aangepast naar de wensen van de gebruiker, hiervoor is echter wel kennis van JavaScript nodig. Er is hier sprake van maatwerk en is niet gemakkelijk door een gebruiker aan te passen.



Figuur 2.5: Stream Layer. Een voorbeeld van de Stream Layer zoals hij standaard wordt geleverd door Esri. De blauwe stippen zijn georeferenced tweets die zojuist verstuurd zijn door gebruikers.

2. ArcGIS for Desktop

Het visualiseren van *georeferenced* tweets hoeft bij de tweede *use case* niet in *real-time* en daarom is de Stream Layer niet noodzakelijk. De meest geschikte optie voor het visualiseren van de data is ArcGIS for Desktop. Deze desktop toepassing van Esri is een uitgebreid pakket aan software, waaronder ArcMap en ArcCatalog. *Georeferenced* tweets die worden opgeslagen in een CSV bestand, of in een *geodatabase* op de server kunnen eenvoudig gevisualiseerd worden met ArcMap. Voornamelijk de optie waarbij de tweets worden opgeslagen in de *geodatabase* is gemakkelijk, omdat hierbij zo min mogelijk handelingen benodigd zijn. In ArcMap worden de tweets direct weergegeven, bij een CSV bestand moet er eerst een koppeling worden gemaakt met de data. Er is echter wel kennis nodig van ArcGIS for Desktop om de *georeferenced* tweets te kunnen visualiseren en bijvoorbeeld een overzichtelijke kaart te maken van deze tweets. Daarnaast is de koppeling tussen ArcGIS for Desktop en de GeoEvent Processor omslachtig. De koppeling is namelijk niet direct en er zijn veel handelingen vereist.

Visualisatie functionaliteiten use cases

In de vorige paragraaf (2.4) zijn de functionaliteiten met betrekking tot koppeling Twitter besproken, in deze paragraaf zal er verder worden ingegaan op de visualisatie functionaliteiten per *case* aan de

hand van de twee hierboven beschreven software opties: Stream Layer en ArcGIS for Desktop. Elke functionaliteit wordt besproken aan de hand van de twee *use cases* om zo tot een conclusie te kunnen komen over de bruikbaarheid van de software. Er is overleg geweest met Esri over de

functionaliteiten van de GeoEvent Processor. Hierin heeft Esri uitleg gegeven bij sommige functionaliteiten, of deze wel of niet beschikbaar zijn en hoe deze dan functioneren. De uitkomsten van dit overleg zijn terug te vinden in de bespreking van de functionaliteiten hieronder.

Presentatie tweets op kaart

Bij beide use cases kunnen de tweets worden gevisualiseerd op een kaart, met zowel de Stream Layer als ArcGIS for Desktop. Bij de standaard Stream Layer die wordt geleverd is het alleen mogelijk om een eenvoudige visualisatie weer te geven, met stippen op de locatie van de tweet. De visualisatie bij de tweede use case kan uitgebreider, aan de hand van de opties van ArcGIS for Desktop. Bij de tweede optie kan er ook gebruik gemaakt worden van ArcGIS Online. De data uit de *geodatabase* kan worden geopend in ArcGIS Online en een simpele visualisatie wordt hier uitgevoerd. Deze optie vereist minder ArcGIS kennis, maar biedt ook minder mogelijkheden voor onderzoek en analyse.

Automatische update

Het automatisch updaten van tweets houdt in dat de kaart constant ververscht wordt waarbij de nieuwste tweets worden toegevoegd aan de kaart. Deze update vindt automatisch plaats, niet handmatig. Automatische update van tweets op de kaart is niet van toepassing bij de tweede use case Analyse, omdat de analyse achteraf plaatsvindt in plaats van in *real-time*. Alle data is al verzameld en wordt in een keer gevisualiseerd, een verversing van het beeld is niet nodig. In het geval van monitoring en bewaking (*use case 2*), is het wel belangrijk dat er automatische update plaatsvindt. Hierbij wil de gebruiker namelijk de informatie in *real-time* kunnen bijhouden, dus is automatische update noodzakelijk. Het in *real-time* updaten van tweets werkt goed met de GeoEvent Processor, de Stream Layer wordt constant ververscht en de tweets komen in *real-time* binnen. Er zit een kleine vertraging in, maar deze is te verwaarlozen.

Selectie tijd

Voor de eerste *use case* kan het nuttig zijn om een selectie in de tijd te kunnen maken, zodat de meteorologen bijvoorbeeld alleen de tweets van het laatste uur kunnen opvragen. Momenteel biedt de standaard Steam Layer alleen de laatste tweets aan, deze worden na verloop van tijd (onbekend hoe lang deze periode is) weer verwijderd. In de standaard functionaliteit die wordt aangeboden kan deze periode niet worden veranderd, maar door een Java Script specialist kan de Stream Layer worden aangepast. Het is echter niet mogelijk om bijvoorbeeld 's middags de tweets van 09.00 tot 11.00 op te vragen, om dus terug te gaan in de tijd. Het actuele tijdstip is het tijdstip dat wordt weergegeven in de Stream Layer.

Ook voor de tweede case kan er een selectie in de tijd worden gemaakt, wanneer er sprake is van een dataset met tweets. Het verschil hierbij is dat dit niet wordt gedaan via de StreamLayer, maar met ArcGIS for Desktop. Daarnaast moet de Stream Layer worden aangepast om de selectie in tijd te kunnen gebruiken, bij ArcGIS for Desktop is het een standaardfunctionaliteit. Er kunnen queries worden uitgevoerd op de database en deze kunnen worden gepresenteerd in ArcGIS for Desktop.

Selecteren en presenteren foto's

Foto's op Twitter staan op een externe webpagina, waarnaar een link wordt geplaatst in de tweet zelf. Tweets bevatten dus geen foto's, maar links naar deze foto's. Bij een dataset afkomstig

van een *databroker* (*use case 2*) staan deze links netjes gesorteerd in een aparte kolom. Door middel van een query kunnen de tweets worden opgevraagd met een verwijzing naar een foto. In ArcGIS for Desktop kan dan aan de hand van de *identifier tool* de link naar de foto worden opgevraagd. De foto's kunnen echter niet in ArcGIS for Desktop verschijnen, alleen de links naar de foto's.

Bij de tweets die zelf worden binnengehaald aan de hand van de GeoEvent Processor (*use case 1*), zijn deze links naar foto's niet netjes gesorteerd in kolommen en vergt het meer moeite en handmatig werk om de links uit de tweets te kunnen destilleren. Volgens Esri is het daarna wel mogelijk om de foto's op de Stream Layer te plaatsen aan de hand van aanpassingen in de Stream Layer door een JavaScript specialist, dit is geen standaardfunctionaliteit.

Zichtbaar maken data tweet

Naast het visualiseren van de locatie van de tweet, kan het ook nuttig zijn om meer informatie over deze tweet te weten te komen. Bijvoorbeeld de tekst van de tweet, de tijd, de gebruiker en andere extra informatie. Voor de *use case* Analyse is het mogelijk om in ArcGIS for Desktop de informatie van de tweet in een *pop-up* venster weer te geven, dit aan de hand van de *Identifier Tool*.

Het zichtbaar maken van de (meta)data van de tweet is als functionaliteit niet beschikbaar in de standaard Stream Layer, de tweets zijn niet *clickable*. Dit zou echter wel mogelijk moeten zijn door aanpassingen in de Stream Layer. Er zouden dan *pop-ups* en *mouse-events* kunnen worden toegevoegd aan de layer, waardoor deze tweets wel *clickable* worden. Hiervoor geldt echter weer, er is een expert nodig om dit aan te passen.

Presentatie 'symbolology'

Er kan daarnaast behoefte zijn aan een aanpassing van de symbolen van de tweets, zodat bijvoorbeeld oudere tweets kleinere symbolen krijgen toegekend dan nieuwere symbolen. Dit is voor de tweede *use case* gemakkelijk te bewerkstelligen aan de hand van ArcGIS for Desktop. Er zijn veel opties om de *symbolology* aan te passen, bijvoorbeeld een kleurschakering op basis van de leeftijd van de tweet. Voor de eerste *case* is kennis nodig van JavaScript, omdat de Stream Layer moet worden aangepast om deze functionaliteit beschikbaar te stellen.

Tweets verwijderen uit presentatie-set

Het kan nuttig zijn om bepaalde tweets te verwijderen uit de presentatie set, bijvoorbeeld tweets die helemaal geen betrekking hebben op het onderwerp van het onderzoek, maar door de keuze van *keywords* toch in de selectie beland zijn. Wanneer er bijvoorbeeld gezocht wordt op het *keyword* 'glad' (bijvoorbeeld sneeuw of ijzel), dit woord betekent in het Nederlands iets anders dan in het Engels. Het kan dan bijvoorbeeld zo zijn dat er tweets in de dataset worden opgeslagen die betrekking hebben op de Engelse vertaling van het woord. Dit is bij de tweede *use case* gemakkelijker, de tweets kunnen handmatig uit het bestand worden gehaald en op deze manier niet worden meegenomen in de analyse. Wanneer er echter sprake is van een zeer grote database, kost het opschonen van een bestand veel tijd en moeite. Tweets kunnen daarnaast wel verwijderd worden uit de Stream Layer (zoals duidelijk is geworden bij de selectie op tijd), maar het is niet mogelijk om deze zelf handmatig uit de Stream Layer te verwijderen. De functionaliteit is dus niet beschikbaar bij de eerste *use case*, waardoor er kans bestaat dat er veel 'rommel tweets' in de Stream Layer terecht komen.

In/Uitzoomen kaart en clustering

Bij beide use cases is het nodig om te kunnen in- en uitzoomen op de kaart. Met de Stream Layer is het mogelijk om in en uit te zoomen op de kaart, zodat meteorologen kunnen inzoomen op een bepaald gebied waarvoor bijvoorbeeld een weeralarm is uitgegeven. In en uit zoomen is ook mogelijk met ArcGIS for Desktop bij de tweede use case.

Er zou echter ook behoefte kunnen zijn aan clustering, wanneer de kaart wordt uitgezoomd, en declustering wanneer er wordt ingezoomd. Clustering en declustering kan nuttig zijn om de kaart duidelijker en overzichtelijker te maken, doordat bijvoorbeeld bij uitzoomen het aantal tweets per regio wordt weergegeven in plaats van alle tweets los. Dit is niet een standaardfunctionaliteit van de Stream Layer, het is echter wel mogelijk door de Stream Layer aan te passen. Dit is echter maatwerk, dat moet worden gedaan door een specialist. Het is niet mogelijk om dit te kunnen uitvoeren bij de tweede use case in ArcGIS Desktop.

Presentatie metadata

Het is noodzakelijk om de metadata te kunnen weergeven op de kaart, zodat het duidelijk is waar de gebruiker naar kijkt. Dit zou kunnen gaan om bijvoorbeeld de tijd en de *keywords* waarop wordt gezocht. Zonder deze metadata kunnen de stippen namelijk van alles zijn, er moet een duidelijke legenda bij. Dit is gemakkelijk bij de tweede *use case*, die kan handmatig worden toegevoegd in ArcGIS for Desktop. Bij Monitoring & Bewaking is dit een ander verhaal, de Stream Layer kent geen legenda of presentatie van metadata. Het is wellicht mogelijk om dit aan te passen in de Stream Layer, maar hiervoor zijn geen duidelijk instructies gegeven door Esri. Het is daarnaast mogelijk om in de HTML code van de Stream Layer bijvoorbeeld handmatig een titel in te voeren. Dit verandert echter niet automatisch mee wanneer er sprake is van een andere zoekfunctie. Er kan dus worden gesteld dat de functionaliteit om de metadata te presenteren in een legenda in de standaarduitvoering van de Stream Layer niet mogelijk is.

Combineren verschillende kaartlagen en radarbeelden

Bij beide *use cases* kan het nuttig zijn om de *basemap* van de kaart aan te passen en te veranderen. Dit kan met ArcGIS for Desktop zeer gemakkelijk, de optie is daarnaast ook beschikbaar voor de Stream Layer. Dit is echter in dit onderzoek niet getest, omdat hier extra kennis voor nodig is op het gebied van JavaScript.

Het KNMI zou echter ook graag de Twitter data willen combineren met data van het primaire proces van de weerkamer. Dat bijvoorbeeld radar beelden en de Twitter data gecombineerd kunnen worden in een kaart, zodat alle informatie in 1 oogopslag zichtbaar is. Vanuit Esri is geen duidelijkheid gegeven over de mogelijkheden van het combineren van verschillende datalagen.

Afdrukken kaart met tweets

Het afdrukken van een kaart met tweets is mogelijk: de webpagina van de Stream Layer kan gewoon afgedrukt worden, hierbij zijn echter de aanpasmogelijkheden niet heel groot. ArcGIS for Desktop biedt veel opties de opmaak van de kaart, die geschikt zijn in het geval van de tweede *use case*. De afgedrukte kaart kan aan de eisen van een goede kaart voldoen, zoals een legenda, noord-pijl, schaal, titel, bronvermelding e.d. Dit is bij de Stream Layer lastiger, omdat de kaart wordt uitgeprint zoals die zichtbaar is in de browser.

2.6 Overige functionaliteiten

Dedicated Software op Client nodig?

Bij de eerste use case is het voor de gebruikers (meteorologen) niet nodig om de GeoEvent Processor software te beheren of te beheersen. Gebruikers hebben voor de visualisatie alleen toegang tot de server nodig, de URL van de Stream Layer en een geschikte browser (Chrome of Firefox). Dit is gemakkelijk te bemachtigen en zal werken op elk computer systeem. In dit geval is er een beheerder nodig die de GeoEvent Processor instelt en onderhoudt en er voor zorgt dat de output in de Stream Layer terecht komt. Voor de tweede *use case* is het echter wel nodig om *dedicated* software te installeren op de client en/of server van de onderzoeker. De gebruiker heeft toegang nodig tot ArcGIS for Desktop, ArcGIS for Server en de GeoEvent Processor

Handmatige verwerking nodig & drempel gebruiker

Handmatige verwerking zorgt over het algemeen voor een hogere drempel voor de gebruiker. Het is prettig als deze drempel zo laag mogelijk ligt en de handmatige verwerking tot een minimum wordt beperkt. In de eerste use case is handmatige verwerking niet nodig, de meteoroloog hoeft alleen de URL te openen en kan daarmee de tweets in de gaten houden. Het *harvesten* van de tweets en de bewerking van de Stream Layer gebeurt vooraf door een beheerder. In de tweede use case is de handmatige verwerking juist wel nodig. ArcGIS for Desktop is een arbeidsintensief programma, wat nogal wat voorkennis verwacht. Hierdoor is de drempel voor de gebruiker een stuk hoger dan bij de eerste use case. Er is een basis ArcGIS kennis nodig voor de geografische analyse, deze is niet vereist bij de *real-time* monitoring en bewaking. In het geval van Monitoring & Bewaking is er echter wel veel kennis nodig bij de beheerder van de GeoEvent Processor.

Kosten

Voor beide use cases zijn er licenties nodig voor de software van Esri. Voor de eerste use case is dit in ieder geval ArcGIS for Server en GeoEvent Processor, die moeten worden geïnstalleerd op de server. Eventueel zou daarbij ArcGIS for Desktop geïnstalleerd kunnen worden, hiervoor is ook een licentie nodig. Voor de tweede use case zijn alledrie de licenties noodzakelijk.

Hoofdstuk 3: Discussie

De GeoEvent Processor is in juli 2013 geïntroduceerd door Esri en het is in de loop van dit onderzoek duidelijk geworden dat het om de eerste versie van de software gaat. Er zitten nogal wat ‘kinderziektes’ in, niet alles werkt al naar behoren. Daarnaast vereisen de installatie, het beheer en onderhoud van de GeoEvent Processor nogal wat ICT, JavaScript en ArcGIS kennis. In dit hoofdstuk wordt er gekeken naar de geschiktheid van de software voor het KNMI, op basis van de functionaliteiten van de eerder besproken *use cases*. Wat werkt wel goed, wat werkt er minder goed?

3.1 Software

Zoals gezegd, vereist de installatie nogal wat kennis van niet alleen ArcGIS, maar ook ICT kennis en kennis van JavaScript zijn nodig om de GeoEvent Processor te kunnen gebruiken voor beide *use cases* die het KNMI voor ogen heeft. Bij de installatie van ArcGIS for Server werden de eerste problemen al geconstateerd: ArcGIS for Server werkt niet op de standaard PC’s van het KNMI. Deze zijn namelijk Windows XP 32 bits, 64 bits is het minimale wat de software nodig heeft (dit probleem zal echter in april 2014 verholpen worden, wanneer het KNMI over gaat op nieuwe Windows systemen). ArcGIS for Server en de GeoEvent Processor zijn daarom voor dit onderzoek geïnstalleerd op een Linux Workstation.

Daarnaast moet er een *geodatabase* worden aangemaakt en worden geïnstalleerd op de server. Dit vereist zowel ICT kennis als vergaande kennis van ArcGIS, wat de installatie compliceert. Deze database is essentieel voor het functioneren van de GeoEvent Processor, maar de installatie kost tijd en moeite. Met deze database is het koppelen van ArcGIS for Desktop en de GeoEvent Processor mogelijk, maar dit is wel een indirecte koppeling. De GeoEvent Processor is niet zo gekoppeld aan ArcGIS for Desktop, dat er direct gevisualiseerd kan worden, er moeten eerst stappen aan vooraf gaan.

Er is veel software nodig voor het laten draaien van de GeoEvent Processor (ArcGIS for Server, GeoEvent Processor en eventueel ArcGIS for Desktop). Er moet dus redelijk veel geïnstalleerd worden en worden aangepast, daarnaast is de GeoEvent Processor niet een op zichzelf staand programma. Hier moet rekening mee worden gehouden wanneer men besluit gebruik te maken van de GeoEvent Processor: ook ArcGIS for Server moet worden aangeschaft. Daarnaast dient de gebruiker er rekening mee te houden dat het beheer en onderhoud moet worden gedaan door iemand die zowel ICT als ArcGIS kennis beheerst. De interface van de software is daarnaast nogal omslachtig en gevoelig voor foutjes. Er moet worden gewerkt met verschillende software, maar hierdoor moeten er ook verschillende schermen en webpagina’s open staan. Vooral de tool GeoEvent Service Designer is omslachtig. De inputs en outputs moeten worden aangemaakt op de GeoEvent Processor Manager pagina, maar deze moeten worden verbonden via de GeoEvent Service Designer.

3.2 Tekortkomingen

Tijdens het onderzoek zijn er een aantal tekortkomingen van de GeoEvent Processor aan het licht gekomen, deze bemoeilijken het gebruik van de GeoEvent Processor. Een eerste tekortkoming die is ontdekt is dat niet alle onderdelen van de GeoEvent Processor werken in alle beschikbare browsers. Sommige onderdelen werken alleen in Chrome en Firefox, sommige alleen in Internet Explorer. Esri raadt aan om Chrome of Firefox te gebruiken als browser, maar niet alle functies van de GeoEvent

Processor Manager werken in deze browsers. Daardoor moet er soms geschakeld worden tussen verschillende browsers om alles werkend te krijgen.



In de input *Receive Tweets* zit de optie *Location*, deze werkt echter niet. Aan de hand van deze optie kan er gefilterd worden op locatie, zodat bijvoorbeeld alleen de tweets binnen Nederland worden binnengehaald. Wanneer er een locatievenster wordt ingevoerd aan de hand van coördinaten, stopt de GeoEvent Processor er zelfs mee. Doordat deze functie niet werkt, is het niet mogelijk om alleen de tweets van een bepaald gebied binnen te krijgen. Hierdoor is de input van tweets veel groter dan gewenst en moet er filtering plaatsvinden achteraf, in plaats van direct bij de input.

Als laatste is er tijdens het onderzoek aan het licht gekomen, dat veel functies niet eenduidig gemakkelijk werken of een aantal keer moeten worden aan/uit gezet voordat ze werken. De GeoEvent Processor loopt nog niet zo soepel als gewenst en de server en de software lopen soms vast om onduidelijke redenen. Dit maakt het werken met de GeoEvent Processor niet gemakkelijk of toegankelijk, er is redelijk wat geduld vereist om met de software te werken.

Functionaliteiten	Use Case 1	Use Case 2
	Mogelijk?	Mogelijk?
2.4 Koppeling Twitter ¹		
Volledigheid Dataset	oo	oo
Filtering o.b.v. metadata	oo	ooo ³
Archivering tweets	oo	oo
2.5 Visualisatie ²		
Presentatie tweets op Kaart	ooo	ooo
Automatische Update	ooo*	
Selectie Tijd	oo*	ooo
Selecteren en presenteren foto's	ooo*	oo
Zichtbaar maken data tweet	ooo*	ooo
Presentatie 'Symbology'	ooo*	ooo
Tweets verwijderen uit presentatie-set	o	ooo
In/uitzoomen kaart	ooo	ooo
Clustering	ooo*	o
Presentatie metadata	o	ooo
Combineren verschillende kaartlagen	ooo*	ooo
Combineren radarbeelden	o	ooo
Afdrukken kaart met tweets	oo	ooo

Tabel 2.2: Functionaliteiten Koppeling Twitter & Visualisatie

- o Niet mogelijk
- oo Gedeeltelijk mogelijk
- oo* Gedeeltelijk mogelijk na aanpassingen expert
- ooo Mogelijk
- ooo* Geen standaard mogelijkheid, aanpassingen expert vereist

¹ Tool beide use cases: GeoEvent Processor

² Tool use case 1: Stream Layer

Tool use case 2: ArcGIS for Desktop

³ Indien gebruik gemaakt wordt van database databroker

2.6 Overige functionaliteiten	Use Case 1	Use Case 2
Dedicated Software op Client Nodig?	Niet nodig voor gebruiker, alleen beheerder	Software nodig voor gebruiker
Handmatige verwerking nodig & drempel gebruiker	Geen handmatige verwerking, lage drempel	Veel handmatige verwerking, hogere drempel
Kosten	Licenties vereist	Licenties vereist

Tabel 2.3: Overige Functionaliteiten

3.3 Bruikbaarheid GeoEvent Processor voor KNMI use cases

In het vorige hoofdstuk zijn alle functionaliteiten voor beide *use cases* behandeld en hieruit is gebleken dat de GeoEvent Processor niet op alle functionaliteit-onderdelen geschikt is. Voor een overzicht van de resultaten zie tabel 2.2 en tabel 2.3. Voor de tweede *use case* is voornamelijk het *harvesten* en archiveren van tweets bruikbaar voor het KNMI. Op basis van de aanwezige functionaliteit binnen de GeoEvent Processor kunnen er bijvoorbeeld tweets worden opgeslagen op een dag met slecht weer en later worden geanalyseerd aan de hand van ArcGIS for Desktop. In tegenstelling tot de verwachtingen die zijn opgewekt door Esri, werkt de GeoEvent Processor niet goed samen met ArcGIS for Desktop. Visualisatie opties met deze software zijn alleen indirect mogelijk. De GeoEvent Processor werkt in dit geval slechts als doorgeefluik, maar biedt verder geen aanvullende mogelijkheden. De vraag is of dit niet gemakkelijker kan aan de hand van een script, gebouwd door een ICT specialist op het KNMI. Aan de hand van een script zouden ook de tweets kunnen worden *geharvest* in samenwerking met de Twitter API. Het voordeel van de GeoEvent Processor is echter wel dat de data gelijk opgeslagen kan worden in een *geodatabase* en geopend kan worden in ArcGIS for Desktop. De GeoEvent Processor biedt echter maar 1% van alle tweets aan, in overeenkomst met de Twitter API. Mocht er behoefte zijn aan de volle 100%, biedt de GeoEvent Processor of een script niet de oplossing, maar moet er contact worden opgenomen met een databroker.

De bruikbaarheid voor de eerste use case ligt anders, hiervoor lijkt de GeoEvent Processor op het eerste gezicht veel mogelijkheden te bieden en dan voornamelijk bij het *real-time* binnenhalen van Twitter data. Dit *real-time harvesten* kan echter ook met een script, dus het is de vraag of de GeoEvent Processor voordelen biedt ten opzichte van zo'n script. Dit zou bijvoorbeeld kunnen door de koppeling met visualiseren. De moeilijkheid zit echter in het visualiseren van deze *real-time* informatie. Verschillende visualisatie opties zijn getest in de loop van dit onderzoek, elke optie bracht andere complicaties met zich mee. Het visualiseren aan de hand van een *geodatabase* en ArcGIS Online bood mogelijkheden, maar hiervoor zijn handmatige acties vereist en niet in *real-time*. Visualiseren met het Esri programma Operations Dashboard werkte niet, vanwege de software van de KNMI computers. Dit programma werkt alleen op Windows 7 en hoger en op het KNMI wordt nog gebruik gemaakt van Windows XP. Daarnaast is deze functie nogal omslachtig en vereist handmatig werk.

De laatste mogelijkheid is het visualiseren aan de hand van een Stream Layer, aangeboden door Esri in een *tutorial* pakket. Deze Stream Layer werkt in *real-time* en wordt automatisch geüpdate, de standaard functies van deze Stream Layer zijn echter erg beperkt. Zo worden tweets automatisch na verloop van tijd verwijderd en wordt er geen metadata gepresenteerd. De locatie van de tweet wordt weergegeven in de Stream Layer, maar geeft geen extra informatie weer zoals de



tijd, de gebruiker en de tekst van de tweet. Een aantal van deze functionaliteiten kunnen wel worden toegevoegd aan de Stream Layer, maar dit moet gebeuren door een specialist op het gebied van JavaScript. De bruikbaarheid van de Stream Layer zou daardoor kunnen worden vergroot, maar dan moet er eerst maatwerk worden uitgevoerd door een expert.

Conclusie & Aanbevelingen

Het KNMI is geïnteresseerd in het kaart brengen van data afkomstig van Twitter, zowel in *real-time* als door middel van analyse achteraf. De data kan zorgen voor nieuwe inzichten betreffende extreem weer *events* in het verleden, of kan juist worden gebruikt in *real-time* in de weerkamer tijdens een dag met extreem weer. Dit onderzoek heeft zich gericht op de mogelijkheden voor het in kaart brengen van deze Twitter data met betrekking tot het ArcGIS framework. Dit is gedaan aan de hand van de volgende hoofdvraag:

‘Wat zijn de mogelijkheden voor het KNMI om binnen het ArcGIS framework Twitter data betreffende het weer in kaart te brengen?’

In dit onderzoek zijn verschillende mogelijkheden binnen het ArcGIS framework onder de loep genomen en een toepassing binnen dit *framework* is als beste naar voren gekomen: de GeoEvent Processor. Het onderzoek heeft zich dan ook met name gericht op de functionaliteiten van deze software, in samenwerking met andere software van uitgever Esri. De functionaliteiten zijn getest aan de hand van twee *use cases* die relevant zijn voor het KNMI: Monitoring & Bewaking en Analyse. Er is gebleken dat de GeoEvent Processor niet van veel toegevoegde waarde is voor de tweede *use case*: Analyse. Het voordeel dat de GeoEvent Processor hier kan bieden is voornamelijk het harvesten van data, maar dit zou ook kunnen worden gedaan aan de hand van scripts.

De voordelen van de GeoEvent Processor zijn groter bij de eerste *use case*: Monitoring & Bewaking. Het in *real-time* binnenhalen van Twitter data werkt naar behoren, alleen het visualiseren van deze informatie is een knelpunt. Er zijn zeker mogelijkheden met de Stream Layer, maar deze moet dan worden aangepast door een expert of het gebied van JavaScript. Daarnaast werkt de GeoEvent Processor nog niet erg gemakkelijk, de installatie is een lastige klus en zitten er fouten in de werking van de software. De interface is nogal omslachtig, waardoor het programma al met al niet erg toegankelijk is. Er is veel kennis op het gebied van ArcGIS, ICT en JavaScript gewenst, om de GeoEvent Processor goed te laten werken als toepassing in de weerkamer van het KNMI.

Terugkomend op de hoofdvraag kan er gesteld worden dat er wel mogelijkheden zijn binnen het ArcGIS Framework voor het in kaart brengen van Twitter data, maar dat de mogelijkheden nogal wat haken en ogen hebben. De GeoEvent Processor is de beste mogelijkheid voor *real-time* visualisatie van Twitter gegevens, maar deze werkt nog niet optimaal. Voor analyse achteraf is ArcGIS for Desktop geschikt, maar daarbij kan wellicht beter gebruik gemaakt worden van een script voor het *harvesten* van de Twitter data. Ook zou er gebruik gemaakt kunnen worden van een *databroker*, zodat de analyse kan worden uitgevoerd op een volledige dataset, in plaats van op de nu publiek beschikbare 1% van alle tweets.

Mocht het KNMI verder willen met de GeoEvent Processor voor monitoring en bewaking in de weerkamer, moet er worden geïnvesteerd in het aanpassen van de Stream Layer. De standaardfunctionaliteit van de Stream Layer is nu namelijk niet voldoende voor gebruik. Als dit kan worden aangepast door een expert of het gebied van JavaScript, kan deze Stream Layer zeker van toegevoegde waarde zijn voor gebruik in de weerkamer. De vraag is echter of het KNMI niet beter zelf kan investeren in het ontwikkelen van een goed script en een visualisatie mogelijkheid, zodat de software van Esri niet nodig is voor het in kaart brengen van *real-time* Twitter data.

Als laatste kan de vraag gesteld worden of de GeoEvent Processor op de langere termijn wel aan de wensen van het KNMI zal voldoen, indien verbeteringen worden toegepast op de software.



Belangrijke verbeterpunten voor het KNMI zouden zijn: gemakkelijkere en toegankelijker software, het combineren van Twitter data met radarbeelden en betere archiveringsopties. De verbeteringen moeten hiernaast voornamelijk in de visualisatiemogelijkheden worden gezocht, zodat het KNMI gemakkelijk in *real-time* en na afloop *events* met extreem weer kan monitoren en analyseren. Hierbij kan gedacht worden aan nauwere samenwerking tussen ArcGIS for Desktop en ArcGIS Online en de GeoEvent Processor. In de nabije toekomst ziet het er niet naar uit dat de software zal voldoen aan alle eisen van het KNMI *use cases*, hiervoor zitten er teveel haken en ogen aan de huidige software.

Bronnen

Alexa (2013), www.alexacom. Geraadpleegd 29 november 2013.

API Voice (2012). <http://apivoice.com/2012/07/12/the-twitter-firehose/>. Geraadpleegd 29 november 2013.

Baranyi, Jeff (2011), <http://blogs.esri.com/esri/arcgis/2011/08/04/tools-for-social-media/>. Geraadpleegd 31 oktober 2013.

Cheng, Z, J. Caverlee & K. Lee (2010), You Are Where You Tweet: A Content-Based Approach to Geo-locating Twitter Users. International Conference on Information and Knowledge Management, Proceedings, pp. 759-768.

ESRI (2013), <http://esri.com>. Geraadpleegd 28 oktober 2013.

ESRI Training (2013), http://training.esri.com/Courses/ts_GeoEventSer/player.cfm. Geraadpleegd: 29 oktober 2013.

GNIP (2014), <http://gnip.com/>. Geraadpleegd 17 januari 2014.

Kwak, H, C. Lee, H. Park & S. Moon (2010), What is Twitter, a Social Network of a News Media? Proceedings of the 19th International Conference on World Wide Web, WWW '10, pp, 591-600.

Morstatter, F., J. Pfeffer, H. Liu, & K.M. Carley (2013). Is the sample good enough? Comparing data from Twitter's Streaming API with Twitter's Firehose. Proceedings of ICWSM 2013.

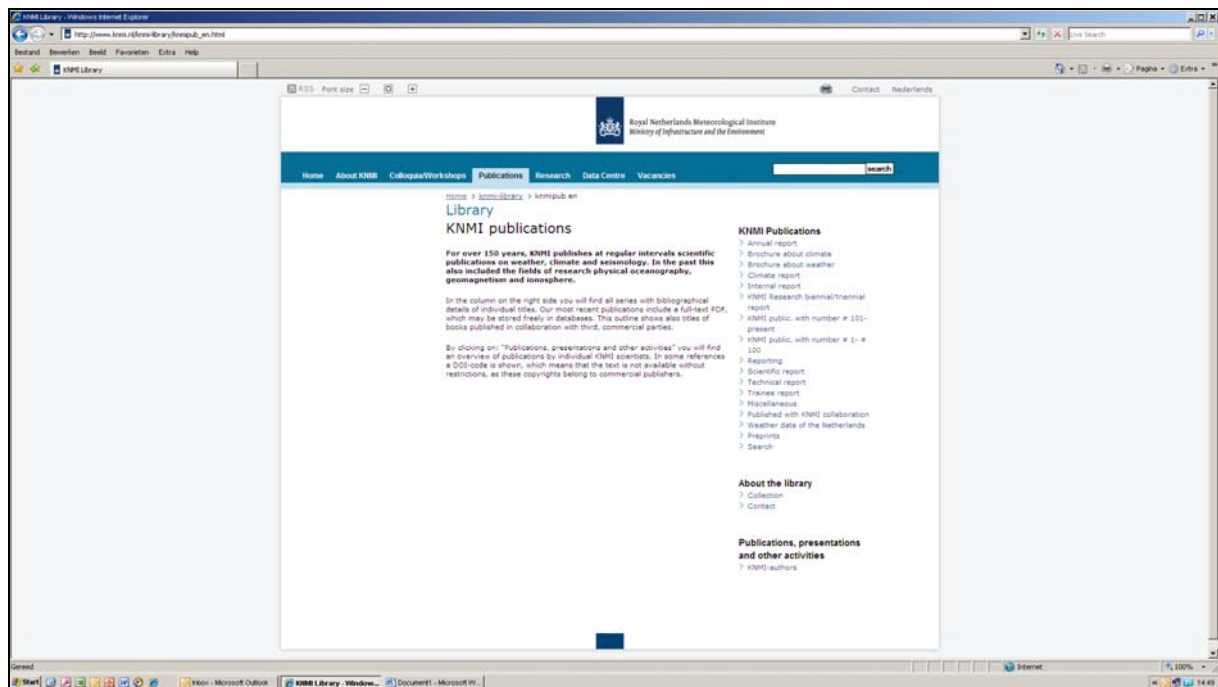
OneMillionTweetMap (2013), <http://onemilliontweetmap.com/>. Geraadpleegd: 29 oktober 2013.

Rijkswaterstaat (2012), Social Media Data Analysis for Crisis Management in the Netherlands. Den Haag: Rijkswaterstaat.

Stefanidis, A., A. Crooks & J. Radzikowski (2013), Harvesting ambient geospatial information from social media feeds. GeoJournal 78 (2), pp. 319–338.

A complete list of all KNMI-publications (1854 – present) can be found on our website

www.knmi.nl/knmi-library/knmipub_en.html



The most recent reports are available as a PDF on this site.

