

Jeroen Fernandez-Alonso
 archivaris, Amsab-ISG
Kim Robensyn coördinator
 collectie-ontsluiting, Amsab-ISG

Meerstemmigheid, sociale media en ethiek in het archief - casus Twitter

Archieven profileren zich graag als ‘het geheugen van de samenleving’, maar wiens geheugen? Een archief is geen neutrale opslagplaats van historische documenten. Achter elk gearchiveerd document liggen beslissingen van erkenning en miskenning. Niet elke groep vindt haar geschiedenis terug in een archief. Sommige groepen hebben nooit de kans gekregen om hun verhaal op te tekenen, of het verhaal werd eenzijdig belicht of werd niet belangrijk genoeg gevonden om opgenomen te worden in een archief. Digitalisering kan een rol spelen in het creëren van een ‘rechtvaardiger’ archief met meerstemmigheid als uitgangspunt.¹ Sociale media zijn het communicatiemiddel van de 21e eeuw geworden waarbij iedereen – mits internetverbinding – kan deelnemen aan het publieke debat zonder redactionele tussenkomst. Sociale media geven op die manier een stem aan onder-gerepresenteerde groepen. Sociale bewegingen zoals Me Too en Black Lives Matter (BLM) hebben veel bereikt, grotendeels bottom-up en gedecentraliseerd, doordat eenieder via de hashtag de mogelijkheid heeft om persoonlijke ervaringen te verwoorden en een plaats te geven in het publieke debat.

Net daarom hecht Amsab-ISG belang aan deze bronnen. En daarom ook moeten we

methodes ontwikkelen om die verborgen geschiedenissen, deze vormen van activisme die zich voor een stuk online afspelen, in al hun diversiteit te capteren, te waarderen en te verzamelen. Onze aandacht gaat dan ook steeds meer uit naar wat zich online manifesteert, op websites en sociale media. We werken momenteel een heuse digitale strategie uit waarbij we ons in eerste instantie richten op websites, Facebookpagina’s, Twitteraccounts enzovoort van onze ‘bestaande’ archiefvormers en van nieuwe organisaties en initiatieven (soms volatiel van aard) die ontstaan rond actuele thema’s. Dit doen we zowel op regelmatige basis als naar aanleiding van belangrijke gebeurtenissen, zoals de coronacrisis of verkiezingen.

Daarnaast capteren we ook hashtags (#), los van onze archiefvormers, die relevant zijn voor onze missie. Deze spelen nog meer in op de actualiteit, vaak spontaan, decentraal en bottom-up maar daarom niet minder relevant, integendeel. De bovengenoemde #BLM is maar een van de voorbeelden. Dit houdt in dat wij, als archiefinstelling, zeer nauw de actualiteit moeten opvolgen en voeling moeten krijgen met wat zich nu afspeelt in de maatschappij. Het beeld van de verstofte archivaris die in het verleden leeft, mag bij deze dus overboord.

In dit artikel leggen we de focus op sociale media. Dat het verzamelen en valoriseren hiervan heel wat extra uitdagingen met zich meebrengt, is evident. We belichten hieronder een aantal technische en ethische uitdagingen.

Technische uitdagingen ↪ Er zijn twee verschillende strategieën om sociale media te archiveren. De eerste strategie werkt volgens dezelfde principes als die van het website-archiveren, ook wel ‘crawlen’ genoemd. Eigen aan deze methode is dat de oorspronkelijke look-and-feel wordt bewaard: de gebruiker kan door de data browsen zoals hij vroeger zou gedaan hebben op de site. Het probleem van deze methode is dat sociale media veel interactiever zijn dan websites: de berichten die iemand op zijn Facebookpagina ziet en de volgorde waarin ze verschijnen, zijn anders dan bij zijn contacten. Ook worden niet alle berichten onmiddellijk geladen. Handmatig scrollen is nodig om meer berichten te laden, en dit kan nagenoeg eindeloos doorgaan. Het is vooral met dit scrollen dat de archiveringssoftware het moeilijk heeft.

Een tweede archiveringsstrategie heet ‘API harvesting’. Een API is een applicatie die de gebruiker toelaat om meer te doen met het programma dan een normale gebruiker kan. De socialemediabedrijven willen een monopolie op hun data. Toch laten ze in beperkte mate toe dat derden hun data analyseren. Via die API’s behouden ze wel het overzicht – ze weten wie van de API gebruikmaakt – en ze kunnen regels opleggen voor het gebruik van data door derden. Wij maken gebruik van hun API’s om de data te ‘scrapen’ of verzamelen. Hiermee bewaren we helaas niet hun oorspronkelijke look-and-feel: we krijgen enkel een dataset met daarin alle tekstuele metadata (de inhoud van het bericht, *likes*, *shares*, datum, URL enzovoort). Als we ook alle foto’s en video’s willen, moeten we die

nog apart verzamelen. Deze methode laat toe om enorme hoeveelheden data te archiveren, maar het menselijke overzicht dreigt hiermee te verdwijnen: alle analyses moeten via de computer gebeuren. Dat is voor een onderzoeker geen probleem: computeranalyses zijn eigen aan het onderzoek naar sociale media. Maar als archivaris stelt het gebruik van API’s wel gevaren aangezien we volkomen afhankelijk zijn van de welwillendheid van de socialemediabedrijven. Het bedrijf Twitter maakt bijvoorbeeld gebruik van ‘rate limiting’, wat het archiveringsproces van hashtags enorm vertraagt doordat de API telkens vijftien minuten wacht alvorens een bepaalde hoeveelheid data te downloaden. Als privébedrijf kan het op elk moment beslissen om die regels nog verder te verstrengen, of ons zelfs de volledige toegang tot die publieke data te weigeren.

Sociale media zijn gratis, maar ‘gratis’ op het internet betekent dat de gebruiker het product is. De bedrijven willen een monopolie houden op de data van hun gebruikers en maken het daarom zeer moeilijk voor derden om die te archiveren. Vooral Facebook houdt de hand op de knip. Dit heeft deels te maken met het Cambridge Analytica-schandaal waarbij de persoonlijke data van miljoenen gebruikers werden geharvest om doelgericht gebruikt te worden voor de presidentscampagne van Trump. Twitter laat dan weer meer toe, maar er zijn limieten op de grootte van de datasets die toegankelijk gemaakt mogen worden. Daarnaast wil Twitter dat, zodra een gebruiker een tweet verwijdert, wij die in principe ook zouden moeten verwijderen uit onze gearchiveerde dataset. Het is niet enkel onmogelijk om te weten welke tweets uit de enorme datasets niet langer online staan, het is ook onwenselijk: dat iets offline is gehaald, is vaak enorm interessant voor onderzoekers. Gelukkig lijkt Twitter uitzonderingen te maken als de tweets bewaard

worden in het openbaar belang. Maar dat heeft meer te maken met de Amerikaanse politiek – Trump verwijderd nogal vaak zijn eigen tweets – dan met een bekommernis over erfgoedinstellingen.

Ethische uitdagingen ↪ Om de ethische kant van het archiveren van sociale media te illustreren, zoomen we in op Twitter. Voor Twitter gebruiken we namelijk de twee bovengenoemde strategieën. De Twitterpagina's van onze archiefvormers willen we zo bewaren dat de onderzoeker erin kan browsen zoals toen het bericht online stond. Dit is een tijdsintensieve archiveringsmethode. Een voordeel is wel dat we de kwaliteit van de data kunnen garanderen. Voor hashtags is diezelfde methode onmogelijk. Hiervoor moeten we gebruik maken van API harvesting. Via het opensourceprogramma Twarc kunnen we zeer gericht hashtags archiveren. Het interessante van Twarc is dat dit enkele jaren geleden ontwikkeld is door activisten naar aanleiding van Black Lives Matter. Ze noemen zichzelf de Documenting the Now-community en gaan zeer ethisch te

werk. Een netwerkanalyse maakt het verschil duidelijk tussen top-downinitiatieven en sociale bottom-upbewegingen en de daarmee samenhangende ethische overwegingen.

- **Fig 1:** In de afbeelding hieronder is een kleine dataset van #ikbensolidair, genomen op 16 april 2020, visueel voorgesteld. Vluchtelingenwerk Vlaanderen verspreidde deze hashtag als onderdeel van een campagne die lokale overheden wilde overtuigen solidair te zijn met vluchtelingen. Het is duidelijk te zien dat deze campagne centraal gestuurd werd. Er zijn weinig derden die de hashtag gebruikten los van de officiële berichten. De grote parse cluster omvat de tweets van het officiële account van Vluchtelingenwerk Vlaanderen en de retweets en *replies* daarop. De andere grote clusters staan in relatie tot Vluchtelingenwerk Vlaanderen. Dat is te zien aan de vele verbindingslijnen tussen de clusters. Zo is de lichtblauwe cluster bijvoorbeeld opgesteld rond de tweets van de Liga voor Mensenrechten, die zelf een retweet zijn van die van Vluchtelingenwerk Vlaanderen.

Fig 1

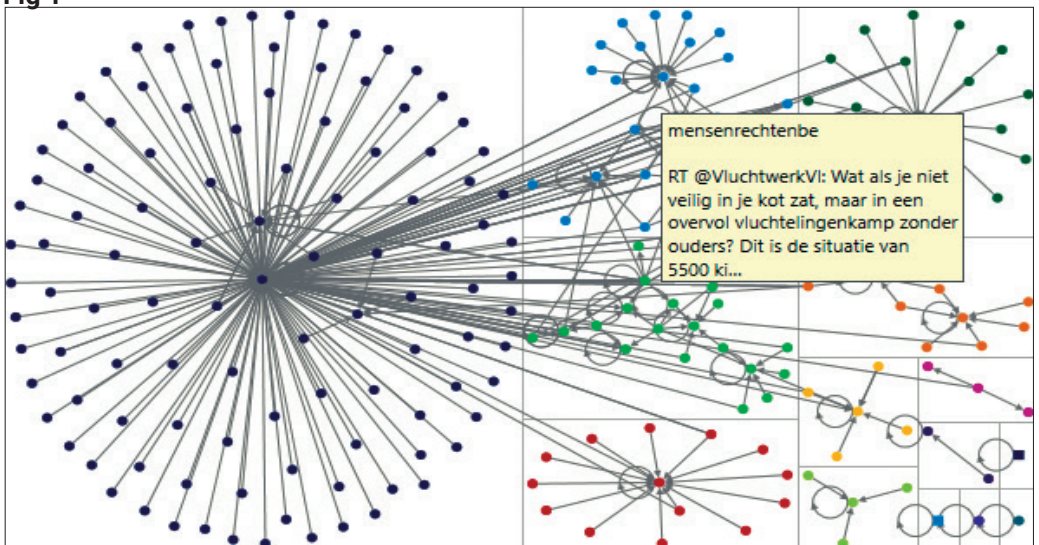
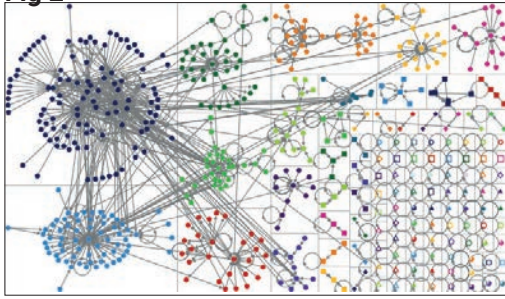


Fig 2



- **Fig 2:** Hierboven zien we een dataset van een hashtag die ook gecentraliseerd verspreid werd, maar waarbij de hashtag meer werd opgenomen door derden. #beternacorona was een gezamenlijke oproep van elf mediaorganisaties en denktanks (Aktief, Apache, DeWereldMorgen, De Gids op Maatschappelijk Gebied, Furia, Kifkif, Lava, Denktank Minerva, MO*, Oikos en Sampol) voor een dringend debat over heropbouw na de pandemie. Ook hier zijn de grote clusters te zien van elk van deze elf organisaties met hun volgers die retweeten en antwoorden. De verspreiding van de hashtag gebeurt nog steeds zeer gecoördineerd en gecentraliseerd, maar we zien toch ook dat heel wat derden los van de grote organisaties hierover hun stem laten horen (zie de kleine clusters rechtsonder).

- **Fig 3:** Op de volgende pagina ten slotte een dataset van een volkomen gedecentraliseerde sociale beweging. Naar aanleiding van de Black Lives Matter-protesten in België hebben we onder meer alle tweets gearchiveerd over Leopold II. De visualisatie toont alle tweets met trefwoord 'Leopold II' van de periode 18-25 juni van dit jaar. Er zijn enkele grote clusters te zien. Maar als we die vergelijken met de bovenstaande visualisaties, dan zien we dat de grote clusters onderling veel minder verbindingslijnen hebben. Meestal zijn het ook niet de organisaties die centraal staan in de grote clusters, maar individuen (vaak activisten) die viraal zijn gegaan of al een groot aantal volgers hadden. Daarnaast

zien we enorm veel kleine clusters, zoveel dat het Nodexl-programma ze niet meer kon visualiseren. Onder hen zijn ongetwijfeld enkele organisaties, maar de meeste zijn gewone burgers die los van elkaar hun mening over de kwestie wilden geven. Het gaat ook verder dan louter 'gepraat'. Vergelijk het aantal tweets die er volkomen ongepland zijn gekomen met de bovenstaande reacties op de gecentraliseerde en geplande acties rond #ikbensolidair en #beternacorona. Het gaat hier (deels) om andere groepen die online enorm actief zijn en veel minder verbonden zijn met de traditionele progressieve organisaties. Eigen aan dit soort datasets zijn de bottom-upinitiatieven die we zien ontstaan. Deze dataset bevat bijvoorbeeld allereerst nieuwe initiatieven in verschillende vormen, zoals petitie die gedeeld worden of informatie over nieuwe acties en nieuwe organisaties (denk bijvoorbeeld aan Belgian Network for Black Lives en Belgian Youth Against Racism).

In Amerikaanse activistische kringen heerst vaak de vrees dat dit soort datasets tegen hen kan worden gebruikt. Naar aanleiding van de Black Lives Matter-protesten in de VS riep het Internet Archive op om al het beeldmateriaal van de betogingen te archiveren op hun website. Hierop werd heftig gereageerd, onder meer door de Documenting the Now-community², die de archiveringstool Twarc heeft ontwikkeld. De vrees leeft er dat politiediensten gearchiveerd beeldmateriaal van betogingen in beslag zullen nemen om de deelnemers te kunnen identificeren. Denk bijvoorbeeld aan livestreams waarbij standbeelden werden neergehaald. Die data worden wel online gestreamd, maar worden dikwijls ook zeer snel offline gehaald, ofwel door de persoon die filmde ofwel door Twitter, die dit soort content niet altijd toelaat. Dit soort gevoelige data staat niet meer online, maar is dikwijls wel tijdig gearchiveerd. Hierbij is het nodig om een afweging te maken tussen cultuurhistorische en persoonlijke belangen

en belangen van de gemeenschap. In de VS hecht de Documenting the Now-community veel belang aan de persoonlijke belangen van de activisten. Zij verwijderen bijvoorbeeld gearchiveerde berichten die ‘problematisch’ zijn. In België zal deze belangenafweging wellicht minder aanwezig zijn. Omwille van onze privacy- en auteursrechtenwetgeving moeten wij deze datasets sowieso intra muros houden.

Een andere ethische vraag heeft te maken met de manier waarop archieven omgaan met actuele gebeurtenissen. Tijdens de coronacrisis riepen lokale overheidsarchieven, private archiefinstellingen en universiteiten wereldwijd op om alle aspecten van het door COVID-19 ontwrichte dagelijkse leven te documenteren door het opsturen van persoonlijke verhalen, getuigenissen, foto’s, filmpjes, dagboeken ... Maar hoe kunnen we zo’n gebeurtenis archiveren als de mensen die er het zwaarst onder lijden vaak degenen zijn die ook in goede tijden geen plaats krijgen in de archieven? Het archiveren van sociale media biedt deels een oplossing om het verhaal van ondergerepresenteerde groepen vast te leggen. Op sociale media worden discussies

gevoerd die we nergens anders terugvinden. Maar hoe beslissen we welk materiaal rond welke gebeurtenis belangrijk genoeg is om te archiveren? Dat is een hele uitdaging. Bij papieren archief kunnen we passiever afwachten tot het materiaal – hopelijk – ooit vanzelf zijn weg zal vinden naar het archief. Bij onlinebronnen daarentegen moeten we snel kunnen beslissen. In extreme gevallen kunnen we in onze verzamelwoede mensen in gevaar brengen als hun uitingen op sociale media of het beeldmateriaal in verkeerde handen vallen.

Het archiveren van sociale media biedt dus enorme mogelijkheden om een stem te geven aan ondergerepresenteerde gemeenschappen, maar we moeten dit voorzichtig en respectvol aanpakken. Dat is niet louter een kwestie van softwaretools en privacywetgeving, ook ethische richtlijnen en contacten met deze groepen zijn van fundamenteel belang.

- 1 Charles JEURGENS, *Geschiedenischaamte? Het witte archief onder vuur*, Ketelaarlezing 2019.
- 2 <https://twitter.com/documentnow/status/1267072586233470983>

Fig 3

